2
Technicalities and Generalizations

2.1 Introduction

This chapter is largely technical in nature. Its aim in part is to consider in more detail some of the theoretical points raised in Chapter 1, and in part to put these in a setting that allows a more detailed and up-to-date discussion of them in later chapters. A second aim is to introduce some further techniques not discussed in Chapter 1. Some rather straightforward generalizations of the theory are also made. Finally, the statement of the Fundamental Theorem of Natural Selection for one gene locus will be given and proved.

Population genetics models often make a number of simplifying assumptions, for example that random mating obtains, that fitnesses are fixed constants, that the population size is effectively infinite, and so on. In this chapter we consider what happens when some of these assumptions are relaxed or even dropped altogether. It is difficult enough to consider the effect of relaxing two or three of these assumptions simultaneously and quite impossible to consider the effect of relaxing them all. In the various sections of this chapter we therefore consider one or other generalization of the theory brought about by relaxing one or other of these assumptions, without attempting to assess the effect of simultaneous relaxation of two or more assumptions. Such an assessment must, at the moment, be largely nonquantitative.

2.2 Random Union of Gametes

In elementary textbooks the way in which the frequencies of the various genotypes in a daughter generation are derived from those in the parent generation is by means of a two-way table. All the various possible matings are listed, their frequencies and the relative frequencies with which they produce various offspring genotypes are noted, and thus the frequencies of the daughter generation genotypes are calculated. This procedure was outlined in Chapter 1 for the case of non-random-mating populations. It is far more efficient, however, for random-mating populations, to proceed in a different way. Restricting attention to autosomal loci, we observe that each individual transmits, for each locus, one gene to each of his/her offspring: The union of two such genes, one from each parent, defines at that locus the genotype of the offspring individual. Random mating of parents is equivalent to random union of genes. Thus, for example, using the notation of Section 1.2, since the frequency of A_1 in the parent generation is $X+Y$, the frequency of A_1A_1 in the daughter generation, being the probability that two genes drawn at random from the parent generation are both A_1, is $(X+Y)^2$. This argument, and parallel arguments for the other genotypes, together give equations (1.1)–(1.3) immediately. Only minor extensions of the argument are needed for more complex cases such as sex-linked loci, multiple alleles, dioecious populations, and so on, and we use this form of argument below in developing the properties of these more complex models.

It was stated in Section 1.6 that explicit models should be set up before any mathematical analysis is attempted, so it is necessary to state more explicitly the model assumed in the above argument. It has been assumed that the population is monoecious, of effectively infinite size and that any daughter-generation individual is formed by the mating of two randomly chosen individuals of the parent generation. It is also assumed that there are no geographical effects, no mating success differentials, and so on. Perhaps most important, it is also assumed that distinct generations can be recognized, so that matings occur only between individuals of the same generation, and that these individuals do not participate in further mating once the daughter generation is formed. These assumptions imply that there is no population age structure. Later, models with assumptions that are more general than, and also rather different from, these will be introduced.

2.3 Dioecious Populations

In this section we drop the assumption that the population is monoecious and suppose instead that it is dioecious, that is admits two sexes. The other

assumptions of the previous section are maintained. We focus initially on the autosomal case, deferring the analysis of the sex-linked case to later.

Suppose first there is no selection, and that in a given generation the genotypic frequencies are as given in (2.1) below:

$$\begin{array}{lccc} & A_1A_1 & A_1A_2 & A_2A_2 \\ \text{males:} & X_M & 2Y_M & Z_M \\ \text{females:} & X_F & 2Y_F & Z_F \end{array} \tag{2.1}$$

The argument of the random union of gametes, suitably modified to the dioecious case, shows that the frequency of A_1A_1 individuals among both males and females of the daughter generation is $(X_M + Y_M)(X_F + Y_F)$, with parallel formulas for A_1A_2 and A_2A_2. This implies that after one further generation of random mating the frequencies in both sexes are in the Hardy–Weinberg form

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ x^2 & 2x(1-x) & (1-x)^2 \end{array} \tag{2.2}$$

where

$$x = \frac{1}{2}(X_M + X_F + Y_M + Y_F). \tag{2.3}$$

The frequencies of the three genotypes among males and among females now remain equal in all further generations. For this reason we often make the modeling simplification of ignoring the existence of two sexes, except of course in special cases, for example in discussing the sex ratio.

One case where the existence of two sexes has to be taken into account is that where genotype fitness values are different in males and females. Suppose then that viability selection exists, so that the relative fitnesses of the genotypes A_1A_1, A_1A_2 and A_2A_2 in males are w_{11}, w_{12}, and w_{22}, with corresponding values v_{11}, v_{12} and v_{22} in females. We consider genotypic frequencies immediately after the formation of the zygotes of any generation, and suppose that in a given generation the males produce A_1 gametes with frequency x and A_2 gametes with frequency $1 - x$. Let the corresponding frequencies for females be y and $1 - y$. Then at the time of conception of the zygotes in the daughter generation the genotypic frequencies are, in both sexes,

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ xy & x(1-y) + y(1-x) & (1-x)(1-y) \end{array}$$

By the age of maturity these frequencies will have been altered by differential viability to the relative values

$$\begin{array}{lccc} & A_1A_1 & A_1A_2 & A_2A_2 \\ \text{males:} & w_{11}xy & w_{12}\{x(1-y) + y(1-x)\} & w_{22}(1-x)(1-y) \\ \text{females:} & v_{11}xy & v_{12}\{x(1-y) + y(1-x)\} & v_{22}(1-x)(1-y) \end{array}$$

The frequencies x' and y' of A_1 gametes produced by males and females of the daughter generation are thus

$$x' = \frac{w_{11}xy + \frac{1}{2}w_{12}\{x(1-y) + y(1-x)\}}{w_{11}xy + w_{12}\{x(1-y) + y(1-x)\} + w_{22}(1-x)(1-y)}\,, \tag{2.4a}$$

$$y' = \frac{v_{11}xy + \frac{1}{2}v_{12}\{x(1-y) + y(1-x)\}}{v_{11}xy + v_{12}\{x(1-y) + y(1-x)\} + v_{22}(1-x)(1-y)}\,. \tag{2.4b}$$

These recurrence relations cannot in general be solved explicitly. It is nevertheless possible to arrive at certain important properties concerning their equilibrium points. It is clear that if selection favors the same allele in both males and females there will be no internal equilibrium, so the two cases of real interest are, first, that where different genes are favored in the two sexes, and second, that where overdominance is involved. Our analysis of these two cases follows that of Kidwell et al. (1977).

Suppose first there is no dominance in fitness for each sex and that selection acts in opposite directions in the two sexes. We thus write the fitnesses in the form

	A_1A_1	A_1A_2	A_2A_2
males	1	$1 - \frac{1}{2}s_m$	$1 - s_m$
females	$1 - s_f$	$1 - \frac{1}{2}s_f$	1

where $s_m,\ s_f > 0$. Solution of the equilibrium equations $x = x'$, $y = y'$ gives, as the only possible equilibrium,

$$x = 1 - s_m^{-1} + \{(s_m s_f - s_m - s_f + 2)(2s_m s_f)^{-1}\}^{1/2},$$
$$y = 1 - s_f^{-1} + \{(s_m s_f - s_m - s_f + 2)(2s_m s_f)^{-1}\}^{1/2}.$$

This equilibrium will be admissible ($0 < x < 1$, $0 < y < 1$) only if

$$\frac{s_m}{1 + s_m} < s_f < \frac{s_m}{1 - s_m} \tag{2.5a},$$

or, equivalently, if

$$\frac{s_f}{1 + s_f} < s_m < \frac{s_f}{1 - s_f}\,. \tag{2.5b}$$

When these conditions apply the equilibrium can be shown to be stable. We conclude that especially if s_m and s_f are small, additive selection acting in opposite directions in the two sexes will maintain a stable equilibrium only if the selective differences in the two sexes are fairly close.

Suppose now that dominance is introduced, so that the fitness scheme becomes

	A_1A_1	A_1A_2	A_2A_2
males	1	$1 - h_m s_m$	$1 - s_m$
females	$1 - s_f$	$1 - h_f s_f$	1

An interesting special case occurs when $h_f + h_m = 1$. Here the conditions (2.5) that there exist a single stable internal equilibrium point continue to apply. When $h_f + h_m < 1$ there will be at most one equilibrium point, and the conditions on s_m and s_f for this to occur are rather less stringent than (2.5). Thus, speaking roughly, for smaller h_f and h_m values, a larger range of s_m and s_f values will lead to an equilibrium point. When $h_f + h_m > 1$ it is possible that more than one internal equilibrium point can arise, but the conditions for this are not given here.

When directional selection obtains for one sex and overdominance in the other, one suspects that a stable polymorphic equilibrium is possible provided the directional selection is not too strong. We quantify this statement in a moment when considering conditions for a stable polymorphic equilibrium to exist.

It is of considerable interest to ask how effective the existence of different selective schemes in the two sexes is in maintaining genetic variation compared to the corresponding effect when identical selective schemes obtain in the two sexes. We attack this question quantitatively by considering the conditions for the existence of an internal polymorphism. For practical purposes we may suppose that such a polymorphism exists when the two equilibria $\text{freq}(A_1) = 0$ in males and females, $\text{freq}(A_2) = 0$ in males and females, are both unstable. If we linearize the recurrence relations (2.4) around $x = y = 0$ and around $x = y = 1$, we find that the condition for an internal polymorphism is that both the inequalities

$$(w_{12}/w_{22}) + (v_{12}/v_{22}) > 2, \tag{2.6a}$$

$$(w_{12}/w_{11}) + (v_{12}/v_{11}) > 2 \tag{2.6b}$$

should hold. These requirements are the natural extensions to the corresponding monoecious population requirement that the heterozygote be more fit than both homozygotes.

When A_1 is at a selective advantage in males (so that $w_{11} > w_{12} > w_{22}$) but overdominance applies in females (so that $v_{12} > v_{11}$, v_{22}), condition (2.6a) holds automatically. However, condition (2.6b) will hold only if the overdominance in females is sufficiently strong compared to the directional selection in males. Thus (2.6b) quantifies our earlier discussion of this point.

How stringent are the conditions given in (2.6)? Suppose we normalize so that $w_{12} = v_{12} = 1$. The conditions (2.6) then reduce to the requirements that the harmonic means of v_{11} and w_{11}, and that of v_{22} and w_{22}, should both be less than unity. Since harmonic means are less than arithmetic means, this is a less stringent requirement than that the arithmetic means both be less than unity. In other words, the existence of different selective parameters in the two sexes provides a stronger mechanism for maintaining genetic polymorphism than taking average selective values over the two sexes would suggest.

The above analysis concerns autosomal loci, and clearly a special analysis is needed in the sex-linked case. Taking the males as the heterogametic sex, the frequencies of the various genotypes in the sex-linked case can be written

male		female		
A_1	A_2	A_1A_1	A_1A_2	A_2A_2
x	$1-x$	Y_{11}	$2Y_{12}$	Y_{22}

If there is no selection, the discussion outlined in the previous section shows that the frequencies in the following generation are

$$\begin{aligned} x' &= Y_{11} + Y_{12}, \\ Y'_{11} &= x(Y_{11} + Y_{12}), \\ 2Y'_{12} &= x(Y_{12} + Y_{22}) + (1-x)(Y_{11} + Y_{12}), \\ Y'_{22} &= (1-x)(Y_{12} + Y_{22}). \end{aligned}$$

In contrast to the autosomal case, one generation of random mating is not sufficient to yield equal frequencies of A_1 in the two sexes. Nor does one further generation of random mating produce female genotypic frequencies in Hardy–Weinberg form. On the other hand, since

$$x' - (Y'_{11} + Y'_{12}) = -\tfrac{1}{2}\{x - (Y_{11} + Y_{12})\},$$

the absolute value of the difference between male and female frequencies of A_1 is halved between successive generations. For practical purposes we may thus assume that after a short time, these frequencies are equal: If this is so, one further generation of random mating yields frequencies in the form

males		females		
A_1	A_2	A_1A_1	A_1A_2	A_2A_2
z	$(1-z)$	z^2	$2z(1-z)$	$(1-z)^2$

where

$$z = \tfrac{1}{3}x + \tfrac{2}{3}(Y_{11} + Y_{12}).$$

When selection operates the behavior is clearly more complex, as is shown by Sprott (1957), Bennett (1957) and Cannings (1967, 1968). We do not go into details here, and in this book we give little attention, perhaps less than is deserved, to sex-linked genes, under the assumption that properties of autosomal loci are normally mirrored, perhaps with minor alterations, in the sex-linked case.

While, in both autosomal and the sex–linked cases, the evolutionary behavior of two-sex systems is slightly more complex than in the monoecious case, the important Mendelian properties of conservation of genetic variation and the suitability of the Mendelian system for evolutionary processes continue to apply.

2.4 Multiple Alleles

We turn now to the case of multiple alleles, considering only random-mating populations. Suppose that at an autosomal locus A, alleles A_1, $A_2, \ldots, A_k$ can occur. We consider a model identical to that of Section 2.2 and assume there is no selection. If the frequency of A_i in any generation is x_i, the concept of the random union of gametes shows that in the next generation the frequency of A_iA_i will be x_i^2 and that of A_iA_j $(i \neq j)$ will be $2x_ix_j$. These frequencies are in generalized Hardy–Weinberg form and are maintained through future generations.

Suppose now that viability differentials exist and that the fitness of A_iA_j is w_{ij}. It is clear that if we continue to count individuals at the moment of conception of each generation, the genotypic frequencies are in Hardy–Weinberg form at that time. The gene frequencies will normally change from one generation to another, and the appropriate recurrence relations are

$$x_i' = x_i \sum_j w_{ij}x_j/\bar{w}, \tag{2.7}$$

$$= x_i w_i/\bar{w}, \tag{2.8}$$

the sum (as with all sums in this section) being over $1, 2, \ldots, k$. In this equation w_i, the "marginal fitness of the allele A_i", is defined as

$$w_i = \sum_j w_{ij}x_j. \tag{2.9}$$

In equations (2.7) and (2.8) the quantity $\bar{w}$, the mean fitness of the population, is defined by

$$\bar{w} = \sum_i x_i w_i = \sum_i \sum_j w_{ij}x_i x_j. \tag{2.10}$$

In view of the statement of the mean fitness increase theorem in Section 1.4, and the condition given there for the existence of a stable internal equilibrium point under the action of selection only, it is natural to ask whether the mean fitness increases from one generation to another in the multiple allele case, and to seek the conditions on the w_{ij} that ensure a stable internal equilibrium point (that is each $x_i > 0$) of gene frequencies.

The most efficient proof that mean fitness increases in the multiple allele case was given by Kingman (1961a) and is reproduced in detail here. The daughter generation mean fitness $\bar{w}'$ is defined by $\bar{w}' = \sum\sum w_{ij}x_i'x_j'$, and we are required to prove that with this definition, $\bar{w}' - \bar{w} \geq 0$. Using (2.7),

we obtain

$$\begin{aligned}\bar{w}' &= \bar{w}^{-2}\Big(\sum_i \sum_j w_{ij}(x_i w_i)(x_j w_j)\Big) \\ &= \bar{w}^{-2}\Big(\sum_i \sum_j \sum_m w_{ij} w_{im} x_i x_j x_m w_j\Big).\end{aligned}$$

By interchanging the roles of j and m we also have

$$\bar{w}' = \bar{w}^{-2}\Big(\sum_i \sum_j \sum_m w_{ij} w_{im} x_i x_j x_m w_m\Big).$$

Thus by averaging, we find

$$\begin{aligned}\bar{w}' &= \tfrac{1}{2}\bar{w}^{-2}\Big(\sum_i \sum_j \sum_m w_{ij} w_{im} x_i x_j x_m (w_j + w_m)\Big) \\ &\geq \bar{w}^{-2}\Big(\sum_i \sum_j \sum_m w_{ij} w_{im} (w_j w_m)^{1/2} x_i x_j x_m\Big) \qquad (2.11) \\ &= \bar{w}^{-2} \sum_i x_i \Big(\sum_j x_j w_{ij} (w_j)^{1/2}\Big)^2 \\ &\geq \bar{w}^{-2}\Big(\sum_i x_i \sum_j x_j w_{ij} (w_j)^{1/2}\Big)^2 \qquad (2.12) \\ &= \bar{w}^{-2}\Big(\sum_j x_j (w_j)^{1/2} \sum_i x_i w_{ij}\Big)^2 \\ &= \bar{w}^{-2}\Big(\sum_j x_j (w_j)^{3/2}\Big)^2 \\ &\geq \bar{w}^{-2}\Big(\Big\{\sum (x_j w_j)\Big\}^{3/2}\Big)^2 \qquad (2.13) \\ &= \bar{w}^{-2}\Big(\sum x_j w_j\Big)^3 \\ &= \bar{w}.\end{aligned}$$

In this sequence of steps the inequality (2.11) is justified by the inequality $\frac{1}{2}(a+b) \geq (ab)^{1/2}$ for positive quantities a and b, and the inequalities (2.12) and (2.13) are justified by the convexity property $\sum x_i a_i^n \geq \big(\sum x_i a_i\big)^n$ for nonnegative a_i and $n \geq 1$. If we assume each $x_i > 0$, this proof also shows that $\bar{w}' = \bar{w}$ if and only if $w_1 = w_2 = \cdots = w_k$, and when this is so,

$$w_i = \bar{w}, \quad i = 1, 2, \ldots, k. \qquad (2.14)$$

This equation and (2.8) together imply that $x_i' = x_i$, so that the system is at an equilibrium point. We thus conclude that in the evolutionary system (2.7), the mean fitness always increases except when the system has reached an equilibrium point, where of course it remains unchanged. This conclusion also applies when some of the x_i are zero, although here of course (2.14) is true only for those values of i for which x_i is positive at the equilibrium point.

In view of the discussion in Chapter 1, it is natural to ask whether the change in mean fitness can be approximated by σ_A^2, the additive genetic variance in fitness. The natural generalization of the procedure that led to (1.16) is to define σ_A^2 as the maximum sum of squares removed by $\alpha_1, \ldots, \alpha_k$ in the expression S, defined by

$$S = \sum\sum x_i x_j (w_{ij} - \bar{w} - \alpha_i - \alpha_j)^2. \tag{2.15}$$

It is found that the values of the α_i that lead to the minimizing of S are

$$\alpha_i = w_i - \bar{w}, \quad i = 1, 2, \ldots, k. \tag{2.16}$$

From this it follows, after some algebra, that

$$\sigma_A^2 = 2\sum_i x_i (w_i - \bar{w})^2. \tag{2.17}$$

When $k = 2$ this reduces to the value given by (1.42).

We now wish to compare the expression in (2.17) with the mean fitness change $\bar{w}' - \bar{w}$, which we write as

$$\bar{w}' - \bar{w} = \bar{w}^{-2}\left[\sum\sum w_{ij} x_i x_j w_i w_j - \bar{w}^3\right].$$

If $w_{ij} = \bar{w} + \delta_{ij}$, $w_i = \bar{w} + \delta_i$, where the δ_{ij} are assumed small, this becomes, on ignoring terms of order δ_{ij}^3,

$$\begin{aligned} \bar{w}' - \bar{w} &\approx \sum\sum \{\delta_{ij}\delta_i + \delta_{ij}\delta_j + \delta_i\delta_j\} x_i x_j \\ &= 2\sum_i x_i \delta_i^2 + \sum_i x_i \delta_i \sum_j x_j \delta_j \\ &= 2\sum_i x_i \delta_i^2. \end{aligned} \tag{2.18}$$

This is identical to (2.17), and we conclude that for small fitness differentials the increase in mean fitness is very closely approximated by the additive genetic variance in fitness. Thus, under the assumptions made, in particular that of small fitness differentials, the MFIT holds for an arbitrary number of alleles at the locus. When fitness differentials are not small a rather different conclusion is found (Seneta (1973)).

Suppose that each x_i is positive. Then (2.17) shows that σ_A^2 is zero if and only if $w_1 = w_2 = \ldots = w_k = \bar{w}$. If some of the x_i are zero, the additive genetic variance σ_A^2 is zero if (2.14) applies for those values of i for which x_i is positive. In both cases the discussion above shows that σ_A^2 is zero if and only if the system is at an equilibrium point. We see later that in multilocus systems the identification just reached for one locus, namely

$$\sigma_A^2 = 0 \Leftrightarrow \text{population in equilibrium} \tag{2.19}$$

no longer holds, although a restricted version of this conclusion can be found.

We consider now the evolution of a metrical character, not necessarily fitness, under the evolutionary system (2.7). Consider some character which for A_iA_j individuals takes the measurement m_{ij}. The mean value $\bar{m}$ of this character is given by $\bar{m} = \sum\sum x_i x_j m_{ij}$, and we wish to compute the change in this mean after one generation. To a first order of approximation,

$$\begin{aligned}\Delta\bar{m} &= 2\sum\sum(\Delta x_i)x_j m_{ij} \\ &= 2\sum(\Delta x_i)m_i \\ &= 2\sum(\Delta x_i)(m_i - \bar{m}) \\ &\approx 2\sum x_i(w_i - \bar{w})(m_i - \bar{m}), \end{aligned} \tag{2.20}$$

where we have defined m_i, the marginal measurement for the allele A_i, by

$$m_i = \sum x_j m_{ij}. \tag{2.21}$$

A verbal description of this conclusion is that the change in the character is twice the covariance between marginal allelic values of the character itself and fitness. For further details, see Robertson (1966, 1968). When the character is fitness itself this conclusion reduces to that obtained in (2.18).

We turn now to the condition under which a stable equilibrium of gene frequencies exists. We first assume that each x_i is positive at the equilibrium. The equilibrium conditions (2.14) can be written

$$\begin{aligned} &w_i - w_1 = 0, \quad i = 2, 3, \ldots, k, \\ &x_1 + x_2 + \cdots + x_k = 1, \end{aligned} \tag{2.22}$$

and this is just a system of k linear equations in k unknowns. It thus possesses no solution, one solution or an infinity of solutions. The first and third cases arise only for special values of the w_{ij}, such as, for example, when all fitnesses are equal. In practice it is most interest to ignore these cases and suppose there is a unique solution of (2.22). Unfortunately this solution might be inadmissible, that is the condition $0 < x_i < 1$, $i = 1, \ldots, k$, might not be met, and even if the equilibrium is admissible it need not be stable. Fortunately the stability criteria have been obtained (Kingman, (1961b)). A unique admissible solution to (2.22) will be stable if and only if the matrix $W = \{w_{ij}\}$ has exactly one positive eigenvalue and at least one negative eigenvalue. In this case the system moves, for any initial frequency point for which each x_i is positive, to this equilibrium. If the equilibrium (2.22) is not admissible or is unstable, the system (2.7) evolves in such a way that one or more alleles become eliminated. The behavior then becomes considerably more complicated, and in practice perhaps the best procedure is to note that the system always moves so that $\bar{w}$ is maximized, so that finding the maximum value of $\bar{w}$ subject to the constraints $0 \leq x_i \leq 1$, $\sum x_i = 1$, via the Kuhn–Tucker theory for quadratic programming, will

provide the stable equilibrium point. A result of Kingman (1961b) relevant to this is that if W has j positive eigenvalues, then at most $k-j+1$ alleles will exist with positive frequencies at this equilibrium.

As the simplest possible example of this theory we consider the case where all homozygotes have fitness $1-s$ $(0<s<1)$, and all heterozygotes have fitness 1. Clearly there is an admissible equilibrium point at $x_i = k^{-1}$. This will be stable if the matrix

$$W = \begin{pmatrix} 1-s & 1 & 1 & \cdots & 1 \\ 1 & 1-s & 1 & \cdots & 1 \\ 1 & 1 & 1-s & \cdots & 1 \\ \vdots & \vdots & \vdots & & \\ 1 & 1 & 1 & \cdots & 1-s \end{pmatrix}$$

has exactly one positive eigenvalue and at least one negative eigenvalue. But standard theory shows that the eigenvalues of this matrix are $k-s$, $-s, \ldots, -s$, and thus the stability conditions are indeed met.

We turn finally to the correlation between relatives in the k-allele system, and take as an example the correlation between father and son. Suppose the father has genotype A_iA_i (and thus measurement m_{ij}). The son will be A_iA_j (and have measurement m_{ij}) with probability x_j, and since the frequency of A_iA_i fathers is x_i^2 this will make a contribution to the covariance of

$$x_i^2 \sum x_j m_{ii} m_{ij} = x_i^2 m_i m_{ii}. \tag{2.23}$$

If the father is A_iA_j (frequency $2x_ix_j$) the son will be A_iA_i (probability $\frac{1}{2}x_i$) or A_jA_j (probability $\frac{1}{2}x_j$), A_iA_j (probability $\frac{1}{2}(x_i+x_j)$), A_iA_ℓ (probability $\frac{1}{2}x_\ell$) or A_jA_ℓ (probability $\frac{1}{2}x_\ell$). The contribution to the covariance corresponding to this case is

$$\begin{aligned} 2x_ix_jm_{ij}\left[\tfrac{1}{2}(x_1m_{i1}+\cdots+x_km_{ik})+\tfrac{1}{2}(x_1m_{j1}+\cdots+x_km_{jk})\right] \\ = \; x_ix_jm_{ij}(m_i+m_j). \end{aligned} \tag{2.24}$$

Adding (2.23) over all i and (2.24) over all i,j $(i<j)$ we arrive at the covariance

$$\sum_i x_i^2 m_i m_{ii} + \sum_{i<j}\sum x_ix_jm_{ij}(m_i+m_j) - \bar{m}^2 = \sum_i x_i(m_i-\bar{m})^2.$$

This is just half the expression (2.17) (if we replace w_{ij} by the more general m_{ij}), and in this way we recover expression (1.10) for the correlation in the measurement between father and son, where now both variance terms have the more general k-allele interpretation. Identical conclusions apply for other relationships, and we conclude that the correlation formulas found in Chapter 1 are not affected by the number of alleles at the locus in question.

2.5 Frequency-Dependent Selection

In all of the above constant fitness values for each genotype have been assumed. It is likely in reality that many fitness values are not constant but depend on the number of individuals in the population, on the frequencies of the various alleles, or on both. In this short section we consider briefly some aspects of frequency-dependent selection. We assume the model of Section 2.2 with two alleles at the locus considered.

Using the fitness scheme (1.25a) we arrived at the equation

$$\Delta x = x(1-x)\{w_{11}x + w_{12}(1-2x) - w_{22}(1-x)\}/\bar{w},$$

and this equation continues to hold if the w_{ij} are functions of the allele frequency x. Clearly there are equilibria when $x = 0$, $x = 1$, or when

$$w_{11}x + w_{12}(1-2x) - w_{22}(1-x) = 0. \tag{2.25}$$

If the functions w_{ij} are sufficiently complex functions of x, (2.25) can have a number of solutions, several of which can be stable. There is little point in considering special cases. Further, $\bar{w}$ need not be maximized at an equilibrium point of the system. (2.25) and the equation $d\bar{w}/dx = 0$ show that mean fitness will not be maximized at an equilibrium if, at that equilibrium,

$$x^2 dw_{11}/dx + 2x(1-x)dw_{12}/dx + (1-x)^2 dw_{22}/dx \neq 0.$$

Thus evolution can cause a steady decrease in mean fitness. In a classical example due to Wright (1948) it is supposed that the fitnesses of A_1A_1, A_1A_2, and A_2A_2 individuals are $1-s+t(1-x)$, 1, and $1+s-t(1-x)$ where $s, t > 0$. If $s < t$ there is a point of stable equilibrium where $x = x^* = 1 - st^{-1}$, whereas the mean fitness is maximized at $\frac{1}{2}\left(\frac{1}{2} + x^*\right)$, halfway between x^* and $\frac{1}{2}$. Clearly, for suitable initial frequencies of A_1, the mean fitness can steadily decrease during the course of evolution.

2.6 Fertility Selection

Until now we have assumed that selection operates through viability differentials. This assumption was made for mathematical convenience, and we now suppose that further selective differences between genotypes arise through differential fertility as well as through viability differences. The analysis now becomes more complex, since fertility relates to mating combinations rather than single genotypes. Our discussion assumes the natural generalizations of the model of Section 2.2 and closely follows the work of Bodmer (1965) and Kempthorne and Pollak (1970). We follow the natural generalization of (1.25a) and suppose that the viability of an A_iA_j genotype is w_{ij} $(i, j = 1, \ldots, k)$ (assumed the same in both sexes) and that the fertility of an $A_iA_j \times A_mA_n$ mating is f_{ijmn}. (We adopt some standard ordering convention such that A_iA_j is the male and A_mA_n the female.) It

is clear that male and female genotypic frequencies will be equal: Let X_{ij} be the frequency of A_iA_j just before the conception of a new generation. Those matings leading to A_iA_i offspring must be of the form $A_iA_j \times A_iA_m$ for some j and m. Consideration of the genotypic products of such matings shows that the frequency of A_iA_i at the birth of the next generation will be proportional to

$$Z_{ii} = f_{iiii}X_{ii}^2 + \tfrac{1}{2}\sum_{m\neq i} f_{iiim}X_{ii}X_{im} + \tfrac{1}{2}\sum_{j\neq i} f_{ijii}X_{ij}X_{ii} + \tfrac{1}{4}\sum_{j\neq i}\sum_{m\neq i} f_{ijim}X_{ij}X_{im}. \tag{2.26}$$

These A_iA_i individuals are now subject to viability selection between birth and the age of maturity, and it follows that the frequency X'_{ii} of A_iA_i just before the birth of the next following generation is given by

$$\mu X'_{ii} = w_{ii}Z_{ii}, \quad i = 1, 2, \ldots, k, \tag{2.27a}$$

where μ is a normalizing constant to be discussed later. Similar considerations for A_iA_j individuals yield

$$\mu X'_{ij} = w_{ij}Z_{ij}, \quad i, j = 1, 2, \ldots, k, i \neq j, \tag{2.27b}$$

where

$$Z_{ij} = (f_{iijj} + f_{jjii})X_{ii}X_{jj} + \tfrac{1}{2}\sum_{m\neq j} f_{iijm}X_{ii}X_{jm} + \sum_{m\neq i} f_{imjj}X_{im}X_{jj} + \tfrac{1}{4}\sum_{m\neq i}\sum_{n\neq j} f_{imjn}X_{im}X_{jn}.$$

The constant μ in (2.27a) and (2.27b) is now chosen so that $\sum\sum X'_{ij} = 1$.

These recurrence relations are far too complex to solve in general, and we make no attempt to do so. Questions concerning the existence and stability of equilibrium points of the system (2.27) have been discussed by Hadeler and Liberman (1975), but we do not pursue them here. Some simplification is possible if it is supposed that the fertilities f_{ijmn} are of the multiplicative form

$$f_{ijmn} = a_{ij}b_{mn}, \quad (a_{ij} = a_{ji}, b_{mn} = b_{nm}). \tag{2.28}$$

Introducing the new variables

$$x_i = (a_{ii}X_{ii} + \tfrac{1}{2}\sum_{j\neq i} a_{ij}X_{ij}) / \sum_{j\leq i}\sum a_{ij}X_{ij},$$
$$y_i = (b_{ii}X_{ii} + \tfrac{1}{2}\sum_{j\neq i} b_{ij}X_{ij}) / \sum_{j\leq i}\sum b_{ij}X_{ij}, \tag{2.29}$$

the recurrence relations (2.27) become, for the multiplicative case,

$$\mu^* X'_{ii} = w_{ii}x_iy_i,$$
$$\mu^* X'_{ij} = w_{ij}(x_iy_j + x_jy_i), \quad i \neq j, \tag{2.30}$$

where μ^* is a new normalizing constant ensuring that the sum of genotypic frequencies is unity. Use of (2.29) and (2.30) shows that

$$x_i' = \big(a_{ii}w_{ii}x_iy_i + \tfrac{1}{2}\sum_{j\neq i} a_{ij}w_{ij}(x_iy_j + x_jy_i)\big)/\sum\sum a_{ij}w_{ij}x_iy_j,$$
$$y_i' = \big(b_{ii}w_{ii}x_iy_i + \tfrac{1}{2}\sum_{j\neq i} b_{ij}w_{ij}(x_iy_i + x_jy_i)\big)/\sum\sum b_{ij}w_{ij}x_iy_j. \quad (2.31)$$

These recurrence relations are identical in form to those in (2.4), and thus the latter system, once appropriate changes in fitnesses have been made to include the viability parameters, continue to apply. Some specific examples are given by Bodmer (1965). One question of particular interest is whether the mean fitness of the system increases with time. Unfortunately it is not at all evident that a natural definition for mean fitness exists in the fertility selection case. Using (2.30) and the analogy with previous recurrence systems, it would be reasonable to define mean fitness as

$$\sum_i w_{ii}x_iy_i + \sum_{i<j}\sum w_{ij}(x_iy_j + x_jy_i). \quad (2.32)$$

With this definition, it is possible for mean fitness to decrease with time. Thus (Kempthorne and Pollak (1970)) if $k = 2$, $w_{11} = w_{12} = 1$, $w_{22} = 0.5$, $a_{11} = a_{12} = 1$, $a_{22} = 2$, $b_{11} = 0.25$, $b_{12} = b_{22} = 1$, $X_{11} = X_{22} = 0$, $X_{12} = 1$, then $x_i = y_i = 0.5$, and the mean fitness, as defined by (2.32), is 0.875. From (2.31), $x_1' = x_2' = \frac{1}{2}$, $y_1' = 5/11$, $y_2' = 6/11$ and using these values in (2.32) the daughter generation mean fitness is $19/22 \approx 0.864$. It is clear that this decrease is caused essentially because the genotype with highest fecundity has lowest viability.

Suppose now that in (2.28), it is assumed that $a_{ij} = b_{ij}$. Then immediately $x_i = y_i$ and that the birth of the new generation genotypic frequencies are in Hardy–Weinberg form. Further the recurrence relations (2.31) are of the form (2.7), and therefore the conclusions deriving from that system, including in particular the result that the mean fitness, defined now as $\sum\sum a_{ij}w_{ij}x_ix_j$, cannot decrease, continue to hold. The change in mean fitness again is approximately equal to the additive genetic variance when the latter is suitably defined so as to include both viability and fertility parameters.

Despite this, it is possible that (2.32) is not a natural definition of the mean fitness of the infant population. The classical definition is that the fitness of any genotype is proportional to half the number of offspring individuals (of whatever genotype) from individuals of the genotype in question, counting being performed at the same stage of the life cycle. We now attempt to find an algebraic definition of mean infant fitness along these lines.

Consider infants of genotype A_iA_j: These survive to adulthood with probability w_{ij}. An A_iA_j individual mating with an A_mA_n individual has

$a_{ij}a_{mn}$ offspring and crediting half of these to the A_iA_j individual and averaging over all A_mA_n, the A_iA_j individuals are credited with a proportionate amount

$$\tfrac{1}{2}w_{ij}a_{ij}\sum_m\sum_n x_m x_n w_{mn}a_{mn}/\bar{w} = w_{ij}a_{ij}\bar{m}/2\bar{w}$$

of offspring, where $m_{ij} = a_{ij}w_{ij}$ and $\bar{m} = \sum\sum x_i x_j a_{ij} w_{ij}$. The mean fitness of the infant population may then reasonably be defined as the weighted average of these quantities, or

$$\sum\sum x_i x_j a_{ij}\bar{m}/2\bar{w} = (\bar{m})^2/2\bar{w}. \tag{2.33}$$

In a parallel fashion the mean fitness of the adult population may be defined: Details are given by Kempthorne and Pollak (1970). Curiously neither the infant mean fitness, defined by (2.33), nor the adult mean fitness, must necessarily increase with time, decreases again possibly occurring when those genotypes with high fertility have low viability. We do not pursue this matter further and simply note the great complexity in general of fertility selection models. During most of the rest of this book selection will be taken to mean viability selection. This is no more than a reflection of the fact that, because the mathematics of viability fitness models is easier than that of fertility fitness models, more is known about viability selection models.

2.7 Continuous-Time Models

In all of this book so far it has been assumed that populations reproduce at discrete time points. There are certainly some real-world populations for which this is a reasonable assumption. On the other hand, it is sometimes more appropriate biologically, or simpler mathematically, to use continuous-time models in which births and deaths can take place at any instant. This normally leads to mathematical systems where changes in gene frequency are described by a differential equation or by differential equation systems. In this section we outline some of these mathematical models and discuss their properties, relying heavily on the definitive work of Nagylaki (1974c, 1976), Nagylaki and Crow (1974) and Kimura (1958).

Consider a locus "A" in a monoecious population and let this locus admit alleles $A_1, \ldots, A_k$. At a given time let the number of A_iA_j individuals be n_{ij}, where we adopt an ordering notation such that the A_i gene has derived from the male parent. Define n_i by $n_i = \frac{1}{2}\sum(n_{ij} + n_{ji})$: Then $2n_i$ is the number of A_i genes in the population. If $N = \sum n_i$ is the population size we may write

$$x_i = n_i/N, \quad X_{ij} = n_{ij}/N \tag{2.34}$$

as the frequencies of A_i and the (ordered) genotype A_iA_j, respectively. Consider a continuous-time deterministic process of population change in which, if terms of order $(\delta t)^2$ are ignored throughout, $NX_{ij}d_{ij}\delta t$ individuals of genotype A_iA_j die in the time interval $(t, t+\delta t)$. Let $M\delta t$ be the number of matings during this time interval, $X_{im,nj}$ be the fraction of these matings which are of the (ordered) type $A_iA_m \times A_nA_j$, and $\tilde{a}_{im,nj}$ the number of offspring from such a mating. We introduce the standardized parameter $a_{im,nj} = M\tilde{a}_{im,nj}/N$, so that $NX_{im,nj}a_{im,nj}\delta t$ is the number of offspring from all (ordered) $A_iA_m \times A_nA_j$ matings in the time interval $(t, t+\delta t)$. Defining $n_{ij}(t)$ as the number of A_iA_j individuals in the population at time t and noting that A_iA_j individuals can arise from various ordered matings in various frequencies, we get

$$n_{ij}(t+\delta t) = n_{ij}(t) + \delta t \left(\sum_{m,n} NX_{im,nj}a_{im,nj} - d_{ij}n_{ij}(t) \right).$$

Letting $\delta t \to 0$ in the usual way, we obtain

$$\dot{n}_{ij} = \sum_{m,n} NX_{im,nj}a_{im,nj} - d_{ij}n_{ij}, \tag{2.35}$$

where the time derivative, here and below, is denoted by a superior dot. This equation and the verbal description leading to it form the basis of the model we shall consider.

It is convenient to define a birth-rate for A_iA_m individuals. Noting that the number of offspring (of whatever genotype) to such individuals acting as first partner in an $A_iA_m \times A_nA_j$ mating during $(t, t+\delta t)$ is $N\sum_{n,j} X_{im,nj}a_{im,nj}\delta t$ and that the number of A_iA_m individuals available to act as parents is n_{im}, it is reasonable for us to define the birth-rate b_{im} for such individuals by the equation

$$n_{im}b_{im} = N\sum_{n,j} X_{im,nj}a_{im,nj}. \tag{2.36}$$

From this, the fecundity b_i, mortality d_i, and "Malthusian parameter" m_i of the allele A_i are defined by

$$x_ib_i = \sum_j X_{ij}b_{ij}, \quad x_id_i = \sum_j X_{ij}d_{ij}, \quad m_i = b_i - d_i. \tag{2.37}$$

The mean fecundity $\bar{b}$, mortality $\bar{d}$, and Malthusian parameter $\bar{m}$ are then given by

$$\bar{b} = \sum x_ib_i, \quad \bar{d} = \sum x_id_i, \quad \bar{m} = \bar{b} - \bar{d}. \tag{2.38}$$

Equations (2.35)–(2.38) jointly yield

$$\dot{N} = \bar{m}N, \tag{2.39}$$

$$\dot{X}_{ij} = \sum_{mn} X_{im,nj} a_{im,nj} - (d_{ij} + \bar{m}) X_{ij}, \tag{2.40}$$

and

$$\dot{x}_i = x_i(m_i - \bar{m}). \tag{2.41}$$

To make further progress it is necessary to make certain assumptions. We assume first that random mating obtains, so that

$$X_{im,nj} = X_{im} X_{nj} \tag{2.42}$$

and that $a_{in,nj}$ can be expressed in the additive form

$$a_{im,nj} = \beta_{im} + \beta_{nj} \tag{2.43}$$

for some set of parameters $\{\beta_{ij}\}$. Equation (2.43) is the natural analogue for continuous-time models of an equation like (2.28) for discrete-time models. Equations (2.37)–(2.43) then lead to

$$a_{im,nj} = \bar{b} + (b_{im} - \bar{b}) + (b_{nj} - \bar{b})$$

so that

$$\dot{X}_{ij} = x_i x_j (b_i + b_j - \bar{b}) - (d_{ij} + \bar{m}) X_{ij}. \tag{2.44}$$

Perhaps the most important question to ask is whether Hardy–Weinberg frequencies hold in this model. Defining $Q_{ij} = X_{ij} - x_i x_j$ as a measure of departure from Hardy–Weinberg, (2.41) and (2.44) yield

$$\dot{Q}_{ij} = x_i x_j (d_i + d_j - d_{ij} - \bar{d}) - (d_{ij} + \bar{m}) Q_{ij}. \tag{2.45}$$

Suppose that $d_i + d_j - d_{ij} - \bar{d} \neq 0$. Then even if Hardy–Weinberg frequencies obtain initially, (2.45) shows that they do not persist and do not hold at an equilibrium of the system (2.35). One particular consequence of this is that the rate of change of mean fitness is not necessarily approximately equal to the additive genetic variance in fitness. It is of some interest to determine the relationship between the two quantities, and we now do this in the simple special case where the quantities $a_{im,nj}$ and d_{ij} (which are functions of the $X_{im,nj}$ and of time) are adjusted so that the Malthusian parameter m_{ij} $(= b_{ij} - d_{ij})$ of the genotype $A_i A_j$ is constant in time.

To find the additive genetic variance we minimize the quantity S, defined by

$$S = \sum \sum X_{ij} (m_{ij} - \bar{m} - \alpha_i - \alpha_j)^2. \tag{2.46}$$

If Hardy–Weinberg frequencies do obtain, so that $X_{ij} = x_i x_j$, this would be done following the lines of the analysis in Section 2.4. To measure the effect of departure from Hardy–Weinberg frequencies we introduce the parameters θ_{ij}, defined by

$$X_{ij} = x_i x_j \theta_{ij}. \tag{2.47}$$

Clearly $\theta_{ij} \equiv 1$ implies that Hardy–Weinberg frequencies obtain. If we insert (2.47) into (2.46), we find that the minimization equations yield

$$x_i\alpha_i + \sum_j x_i x_j \theta_{ij}\alpha_j = \sum_j x_i x_j \theta_{ij}(m_{ij} - \bar{m}) \tag{2.48}$$

or

$$\alpha_i + \sum_j x_j \theta_{ij}\alpha_j = \sum_j x_j \theta_{ij} a_{ij} = a_i, \tag{2.49}$$

where we define

$$a_{ij} = m_{ij} - \bar{m}, \quad a_i = x_i^{-1} \sum_j X_{ij} a_{ij}. \tag{2.50}$$

Further, the additive genetic variance, being the sum of squares removed by this procedure, is

$$\sigma_A^2 = 2\sum_i x_i a_i \alpha_i, \tag{2.51}$$

where a_i is defined explicitly by (2.50) and α_i implicitly by (2.49). In view of (2.41) this may also be written

$$\sigma_A^2 = 2\sum_i \dot{x}_i \alpha_i. \tag{2.52}$$

We turn now to the rate of change of the mean fitness $\bar{m}$. By definition

$$\bar{m} = \sum\sum m_{ij} X_{ij}$$

and since under our assumptions the m_{ij} are constant,

$$\begin{aligned}
\dot{\bar{m}} &= \sum\sum m_{ij}\dot{X}_{ij} \\
&= \sum\sum a_{ij}\dot{X}_{ij} \\
&= \sum\sum a_{ij}(\dot{x}_i x_j \theta_{ij} + x_i \dot{x}_j \theta_{ij} + x_i x_j \dot{\theta}_{ij}) \\
&= 2\sum\sum a_{ij}\dot{x}_i x_j \theta_{ij} + \sum\sum a_{ij} x_i x_j \dot{\theta}_{ij} \qquad (2.53) \\
&= 2\sum_i \dot{x}_i \sum_j a_{ij} x_j \theta_{ij} + \sum\sum a_{ij} x_i x_j \dot{\theta}_{ij} \\
&= 2\sum_i \dot{x}_i \big(\alpha_i + \sum_j x_j \theta_{ij}\alpha_j\big) + \sum\sum a_{ij} x_i x_j \dot{\theta}_{ij} \\
&= \sigma_A^2 + 2\sum\sum \dot{x}_i x_{ij} \theta_j + \sum\sum a_{ij} x_i x_j \dot{\theta}_{ij}. \qquad (2.54)
\end{aligned}$$

We wish to simplify the final two terms in (2.54). Now

$$x_j = \sum_i X_{ij} = \sum_i x_i x_j \theta_{ij}$$

so that

$$\sum_i x_i \theta_{ij} \equiv 1.$$

Differentiating with respect to t,

$$\sum_i \dot{x}_i \theta_{ij} + \sum_i x_i \dot{\theta}_{ij} \equiv 0 \quad \text{for each } j.$$

Thus the second term in (2.54) can be written

$$-2 \sum\sum x_i x_j \alpha_j \dot{\theta}_{ij} = -\sum\sum x_i x_j (\alpha_i + \alpha_j) \dot{\theta}_{ij}.$$

The final two terms in (2.54) thus become

$$\sum\sum (a_{ij} - \alpha_i - \alpha_j) x_i x_j \dot{\theta}_{ij} = \sum\sum \delta_{ij} X_{ij} (\dot{\theta}_{ij}/\theta_{ij}),$$

where $\delta_{ij} = a_{ij} - a_i - a_j$ is a measure of nonadditivity in the Malthusian parameters m_{ij}. We conclude that

$$\dot{\bar{m}} = \sigma_A^2 + \sum\sum X_{ij} \delta_{ij} d(\log \theta_{ij})/dt. \tag{2.55}$$

Thus the rate of increase of mean fitness is equal to the additive genetic variance in general only if Hardy–Weinberg frequencies hold (which, as we have seen in our model at least, they do not) or if the Malthusian parameter is additive ($m_{ij} = \alpha_i + \alpha_j$). A more general and more important conclusion, with m_{ij} no longer kept constant, is given by Kimura (1958).

How important then are departures from Hardy–Weinberg frequencies? In our model (2.45) shows that departures will be negligible after some time has passed if $d_i + d_j - d_{ij} - \bar{d} = 0$. But there is another circumstance under which departures will also be negligible. Suppose that the deviations $b_{ij} - \bar{b}$ and $d_{ij} - \bar{d}$ are all of order s, where s is a small parameter. Then Nagylaki (1976) has shown that the deviation Q_{ij} defined above changes in time (according to (2.45)) in such a way that after a small time period t_1 (an explicit formula for which is given by Nagylaki), Q_{ij} differs from zero only by a term of order s, even though at that time the gene frequencies themselves may be far from their equilibrium values. After time $2t$, the rate of change of Q_{ij} is of order s^2. When this occurs a state of "quasi-Hardy–Weinberg" (QHW) is said to obtain. In this case departures from Hardy–Weinberg frequencies may be trivial, and as a consequence the mean fitness increase theorem should hold to an excellent approximation. More exactly, under the assumptions we have made, the term σ_A^2 in (2.55) is of order s^2, and when QHW obtains the final term is of order s^3. Thus the first term on the right-hand side will dominate the second, leading, as noted, to the essential accuracy of the theorem. The only exception to this rule occurs when the various frequencies are close to their respective equilibrium points: Since $\sigma_A^2 = 0$ at equilibrium, it is possible that near equilibrium σ_A^2 is smaller than the final term in (2.55). This is probably

of minor importance, and during the period of substantial change in gene frequencies the MFIT is effectively true.

2.8 Non-Random-Mating Populations

In this section and the next we consider properties of the discrete-time models considered above, focussing attention on the case where random mating is no longer assumed. In this section we consider calculations associated with the one-locus version of the mean fitness increase theorem (MFIT) and in the next on calculations associated with the Fundamental Theorem of Natural Selection (FTNS). In both sections we use a notation that generalizes readily to the multilocus extensions considered in later chapters.

Suppose that fitness depends on the genotype at one locus only, at which occur alleles A_1, A_2, ..., A_k. Any form of mating is allowed, random or otherwise. We denote the frequency of the (ordered) genotype A_uA_v at the time of conception of any generation of individuals by X_{uv} $(= X_{vu})$, so that the frequency x_u of the allele A_u is given by $x_u = \sum_v X_{uv}$.

We assume that the genotype A_uA_v has (viability) fitness w_{uv}. The mean fitness $\bar{w}$ of the population is then given by

$$\text{mean fitness } = \bar{w} = \sum_u \sum_v w_{uv} X_{uv}. \tag{2.56}$$

The additive genetic variance in fitness is found by the non-random-mating generalization of the procedure that led to the "random-mating" expression (2.17). That is, it is found by minimizing the function S, now defined more generally than in (2.15) as

$$S = \sum \sum X_{uv}(w_{uv} - \bar{w} - \alpha_u - \alpha_v)^2, \tag{2.57}$$

subject to the constraint

$$\sum_u x_u \alpha_u = 0. \tag{2.58}$$

The values of α_1, α_2, ..., α_k found through this minimizing procedure, that is the *average effects* of the alleles A_1, A_2, ... , A_k, are the implicit solutions of the equations

$$x_u \alpha_u + \sum_v X_{uv} \alpha_v = x_u a_u, \quad u = 1, 2, \ldots, k, \tag{2.59}$$

where a_u, the *average excess* of the allele A_u, is given by

$$a_u = x_u^{-1} \sum_v X_{uv}(w_{uv} - \bar{w}). \tag{2.60}$$

Equation (2.59) shows that, under random mating, the average effect α_u of A_u and the average excess a_u of A_u are equal, since under random mating the second term on the left-hand side of (2.59) is 0. When mating is not random, α_u and a_u are, in general, different from each other.

Standard regression theory shows that the sum of squares removed by fitting the α_j values in (2.57), that is the additive genetic variance σ_A^2, is given by

$$\sigma_A^2 = 2\sum_u x_u a_u \alpha_u. \tag{2.61}$$

With the definition of a_u given in (2.60), the change Δx_u in the frequency of A_u between consecutive generations is

$$\Delta x_u = x_u a_u / \bar{w}, \tag{2.62}$$

so that an alternative expression for the additive genetic variance is

$$\sigma_A^2 = 2\bar{w}\sum_u \alpha_u \Delta x_u. \tag{2.63}$$

Similarly an alternative set of formulas implicitly defining the quantities $\{\alpha_u\}$ is

$$x_u\alpha_u + \sum_v X_{uv}\alpha_v = \bar{w}\Delta x_u, \quad u = 1, 2, \ldots, k. \tag{2.64}$$

If we define D as a diagonal matrix whose uth term is x_u, P as a matrix whose (u, v)th term is X_{uv}, $\boldsymbol{\Delta}$ as a vector of the Δx_u values and $\boldsymbol{\alpha}$ as a vector of the α_u values, this equation can be written in matrix and vector form as

$$(D + P)\boldsymbol{\alpha} = \bar{w}\boldsymbol{\Delta}. \tag{2.65}$$

When this matrix form is used, the extension of the definition of the α_u to the multilocus case in Chapter 7 will be almost immediate.

An explicit solution of the equations in (2.59) for the α_u values is not in general possible. However in the two-allele case an explicit solution of is straightforward. For this case we get

$$\alpha_u = a_u x_1 x_2 / \{X_{11}X_{12} + 2X_{11}X_{22} + X_{12}X_{22}\}, \quad u = 1, 2. \tag{2.66}$$

Under random mating $X_{uv} = x_u x_v$, and this equation confirms that in this case a_u and α_u are equal. The equation also shows that under non-random mating, a_u and α_u have the same sign and are zero or nonzero together. In the two-allele case Fisher often described $\alpha_2 - \alpha_1$ as the average effect of replacing A_1 by A_2, but in the k allele case, to which we now return, the definition of α_u simply as the average effect of A_u is rather more flexible.

We now consider the change in mean fitness from one generation to another. We write

$$w_{uv} = \bar{w} + \alpha_u + \alpha_v + \epsilon_{uv},$$

and with this definition, (2.58) implies that

$$\sum_u \sum_v X_{uv}\epsilon_{uv} = 0. \tag{2.67}$$

The frequency of A_u at the birth of any given generation is $\sum_v X_{uv}$, and in the next generation at birth it will be $\sum_j X_{uv}w_{uv}/\bar{w}$. Thus the change in mean fitness between consecutive generations becomes

$$\begin{aligned}
\Delta\bar{w} &= \sum\sum X'_{uv}w_{uv} - \bar{w} \\
&= \sum\sum X'_{iuv}(\alpha_u + \alpha_v + \epsilon_{uv}) \\
&= 2\sum_u \alpha_u x'_u + \sum\sum(X_{uv} + \Delta X_{uv})\epsilon_{uv} \qquad (2.68) \\
&= 2\sum_u \alpha_u(\Delta x_u) + \sum\sum(\Delta X_{uv})\epsilon_{uv} \quad \text{(from (2.59)) and (2.67)} \\
&= \sigma_A^2/\bar{w} + \sum\sum(\Delta X_{uv})\epsilon_{uv}, \quad \text{(from (2.61) and (2.62))}.
\end{aligned}$$

If the second term on the right-hand side of this expression is small, the conclusion of the mean fitness increase theorem approximately applies.

2.9 The Fundamental Theorem of Natural Selection

We now turn to the Fundamental Theorem of Natural Selection (FTNS), considering first the discrete-time version, and later the continuous-time version, of this theorem.

Equation (2.58) shows that $\sum_u \sum_v X_{uv}(\alpha_u + \alpha_v) = 0$, and from this the mean fitness $\bar{w}$ may be written in the form

$$\bar{w} = \sum_u \sum_v X_{uv}(\bar{w} + \alpha_u + \alpha_v). \tag{2.69}$$

In the FTNS, Fisher considered the change in mean fitness from one generation to another *only* through changes in the frequencies X_{uv} in the expression (2.69), with the quantities $\bar{w}, \alpha_u$ and α_v being kept constant. This is called the "partial change" in mean fitness, and we denote it by $\Delta_p(\bar{w})$. If X'_{uv} is the frequency of the (ordered) genotype A_uA_v in the

daughter generation, this partial change $\Delta_p(\bar{w})$ is

$$\Delta_p(\bar{w}) = \sum_u \sum_v (X'_{uv} - X_{uv})(\bar{w} + \alpha_u + \alpha_v) \tag{2.70}$$

$$= \sum_u \sum_v (X'_{uv} - X_{uv})(\alpha_u + \alpha_v)$$

$$= 2 \sum_u \alpha_u \sum_j (X'_{uv} - X_{uv})$$

$$= 2 \sum_u \alpha_u \Delta x_u \tag{2.71}$$

$$= \sigma_A^2 / \bar{w}. \tag{2.72}$$

The final step in this sequence comes from (2.63).

We call the interpretation of the FTNS in the above form the "Price" interpretation, since it was first given by Price (1972). This interpretation follows the spirit of the wording in Fisher (1930, 1958).

Thus the partial change in mean fitness is exactly equal to $\sigma_A^2/\bar{w}$, and this is the one-locus statement of the FTNS. Thus, as asserted by Fisher (1930, 1958), the FTNS is an exact result, implying no approximations, and it applies to non-random-mating as well as random-mating populations, since no assumption about the mating scheme is made in the analysis. We extend the FTNS as an exact result in Chapter 7 to the case where fitness depends on an arbitrary number of loci, up to and including all those in the entire genome, under any form of mating, random or otherwise.

An alternative way of writing the FTNS in this interpretation is

$$\Delta_p(\bar{w}) = \sum_u \sum_v (\Delta X_{uv})(w_{uv})_\alpha = \sigma_A^2 / \bar{w}. \tag{2.73}$$

Here $(w_{uv})_\alpha = \bar{w} + \alpha_u + \alpha_v$ may be thought of as the best estimate of the fitness of the genotype $A_u A_v$ as predicted from the alleles in that genotype. In this form the Price interpretation bears an interesting similarity to a second interpretation to the FTNS, one which is closer in spirit to the wording in Fisher (1941), and which was developed by Lessard (1997). Lessard's interpretation uses a concept of partial change different from, although mathematically equivalent to, that in the Price interpretation. In the Lessard interpretation the actual fitness w_{uv} of the genotype $A_u A_v$ is retained, but the change in genotype frequency is replaced by a "alleles derived" value. More explicitly, the statement of the theorem under this interpretation is that

$$\Delta_p(\bar{w}) = \sum_u \sum_v (\Delta X_{uv})_\alpha \; w_{uv} = \sigma_A^2 / \bar{w}, \tag{2.74}$$

where $(\Delta X_{uv})_\alpha$ is defined by

$$(\Delta X_{uv})_\alpha = \frac{(X_{uv})(\alpha_u + \alpha_v)}{\bar{w}}. \tag{2.75}$$

$(\Delta X_{uv})_\alpha$ is not the actual change in the frequency of the genotype A_uA_v from one generation to another, but is thought of the change as predicted from the alleles A_u and A_v in that genotype. The similarity of the forms of the middle terms in (2.73) and (2.74), and the identity of the right-hand sides, together indicate the mathematical identity of the two concepts of partial change. The difference between the two concepts is in the interpretation: In the first interpretation the genes in a genotype may be thought of as assessing the genotype fitness, while in the second they may be thought of as assessing the change in the frequency of that genotype.

The background to Lessard's interpretation of the FTNS is as follows. Fisher (1941) discussed in some detail the circumstances under which the equation

$$\frac{\Delta X_{uu}}{X_{uu}} + \frac{\Delta X_{vv}}{X_{vv}} = 2\frac{\Delta X_{uv}}{X_{uv}} \tag{2.76}$$

will hold for all u and v. If these equations do hold for all u and v, then $\Delta X_{uv}/X_{uv}$ can be expressed in the form

$$\frac{\Delta X_{uv}}{X_{uv}} = \beta_u + \beta_v, \tag{2.77}$$

for some set of constants $\beta_1, \beta_2, \ldots, \beta_k$. From this,

$$X_{uv}(\beta_u + \beta_v) = \Delta X_{uv}. \tag{2.78}$$

Summation in this identity over all v gives

$$x_u\beta_u + \sum_v X_{uv}\beta_v = \Delta x_u \quad \text{for all } u. \tag{2.79}$$

Equation (2.64) then shows that we may take $\beta_v = \alpha_v/\bar{w}$ for all v, where α_v is the average effect of A_v. It follows from (2.77) that

$$\Delta X_{uv} = \frac{(X_{uv})(\alpha_u + \alpha_v)}{\bar{w}}. \tag{2.80}$$

Comparison of this equation with (2.75) shows that when all the equations of the form (2.76) hold, the actual change genotype frequency (2.80) is identical to the change as assessed by the alleles in the genotype. This implies that the *total* change in mean fitness is equal to the *partial* change defined in both equation (2.72) and equation (2.75).

However, equation (2.76) will hold only under very restrictive mating conditions. The random-mating case is perhaps the most important of these. Under random mating the equation $X_{uv}^2 = 4X_{uu}X_{vv}$ holds, so that $2\log X_{uv} = \log 4 + \log X_{uu} + \log X_{vv}$. From this, $2\Delta \log X_{uv} = \Delta \log X_{uu} + \Delta \log X_{vv}$. If small-order terms are ignored, so that $\Delta \log x$ can be replaced by $(\Delta x)/x$, equation (2.76) then follows. More generally the conclusion still follows, to this level of approximation, if $X_{uv}^2 = \lambda X_{uu}X_{vv}$ for any fixed constant λ. Again ignoring small-order terms, it follows that the restriction $X_{uv}^2 = \lambda X_{uu}X_{vv}$ is required for the total change in mean

fitness to be predictable from parental generation genotype frequencies and fitnesses. The point of the FTNS is that random mating is *not* required for the theorem to hold, so that (2.76) does not necessarily hold. Then the total change in mean fitness is not predictable unless the mating scheme is known. Despite this, the FTNS holds whatever the mating scheme might be, and whether it be known or unknown.

It is straightforward to give also a continuous-time version of the FTNS. This shows that the continuous-time partial rate of change in mean fitness, defined as

$$\sum_u \sum_v \left(\frac{d}{dt} X_{uv}\right)(\bar{w} + \alpha_u + \alpha_v) \tag{2.81}$$

is exactly equal to the additive genetic variance. We do not provide the details since the closely follow those in the discrete-time case.

What biological relevance does the FTNS have? There are two points to raise here. First, the restrictive assumptions made in the theorem should be noted. Matters such as geographical dispersion, the existence of two sexes, stochastic changes in gene frequency in finite populations, and so on are ignored. On the other hand fertility selection is handled by Lessard and Castilloux's (1995) extension of the theorem to that case. Second, Fisher viewed the partial change in mean fitness as that change brought about by natural selection. It is not clear how this interpretation can be sustained, and it is possible that the MFIT, even though it is restricted to random-mating populations and, as we show in the following section, might not hold when fitness depends on a two-locus and more generally a multilocus genotype, nevertheless gives a greater biological insight into the evolutionary process than does the FTNS. Associated with this view is the approach, initiated by Nagylaki (1974c), which delimits the circumstances under which the MFIT is approximately true.

2.10 Two Loci

So far in this chapter we have assumed that the fitness of any individual depends on his genetic constitution at a single locus. This is of course only an initial simplification: We have already noted in Chapter 1 that for some questions, for example, the evolution of recombination rate, a more complicated theory is required. We now introduce briefly the case where fitness depends on the genetic constitution at two loci, deferring a more complete treatment to Chapter 6. Although such a "two-locus" theory may often be little more realistic than "single-locus" theory, it does allow at least two advances to be made. First, some assessment can be made of the accuracy of approximating two-locus behavior and measurements by combining two single-locus results. Second, no assessment of the evolutionary importance of linkage between loci can be made without at least a two-locus analysis.

For convenience we assume viability selection only, random mating and discrete nonoverlapping generations. Consider two loci "A" and "B" at which occur alleles A_1, A_2 and B_1, B_2, respectively, and let the recombination fraction between the loci be R ($0 < R \leq 0.5$). (When $R = 0$ the two loci in effect become one locus, the theory of which has already been considered. This is why we impose the assumption $R > 0$.) It is convenient conceptually to suppose that these loci are on the same chromosome: The unlinked case ($R = 0.5$) may be treated by imagining the distance along the chromosome between the two loci to be so long that the recombination fraction between them is 0.5. We then use the words gamete and chromosome interchangeably in what follows.

It is possible to write down recurrence relations connecting the (ten) zygotic frequencies (of A_1B_1/A_1B_1, $A_1B_2/A_1B_1, \ldots, A_21B_2/A_2B_2$). These relations show that a simpler set of recurrence relations can be found for the frequencies of the four gametes A_1B_1, A_1B_2, A_2B_1 and A_2B_2, called here gametes 1, 2, 3, 4, respectively. This simplification arises through the concept of the random union of gametes and is parallel to treating gene frequencies rather than genotypic frequencies at a single locus.

We consider first the case where there is no selection. The gametes forming the zygotes of any generation may be thought of as being drawn randomly from a pool containing gametes of type 1–4 in certain proportions. These gametes will not necessarily be passed on to the next generation of gametes in the same proportions since, for example, there will be a decrease in the frequency of A_1B_1 gametes through recombination in A_1B_1/A_2B_2 individuals which might not be exactly counterbalanced by an increase through recombination in A_1B_2/A_2B_1 individuals. If the frequency of gamete i is denoted c_i ($i = 1, \ldots, 4$), these arguments and some straightforward calculations show that the frequencies c_i' in the next generation are given by

$$\begin{aligned} c_1' &= c_1 + R(c_2c_3 - c_1c_4), \\ c_2' &= c_2 - R(c_2c_3 - c_1c_4), \\ c_3' &= c_3 - R(c_2c_3 - c_1c_4), \\ c_4' &= c_4 + R(c_2c_3 - c_1c_4), \end{aligned} \tag{2.82}$$

or more economically as

$$c_i' = c_i + \eta_i R(c_2c_3 - c_1c_4), \tag{2.83}$$

where

$$\eta_1 = \eta_4 = 1, \quad \eta_2 = \eta_3 = -1. \tag{2.84}$$

Several conclusions can be drawn immediately from these equations. First, since $c_1' + c_2' = c_1 + c_2$ and $c_1' + c_3' = c_1 + c_3$, there is no change in the frequencies of A_1 and B_1 This confirms, fortunately, the one-locus analysis

of Chapter 1. Second, elementary algebra shows that

$$c_1'c_4' - c_2'c_3' = (1 - R)(c_1c_4 - c_2c_3), \tag{2.85}$$

so that since $R > 0$,

$$c_1(t)c_4(t) - c_2(t)c_3(t) \to 0 \quad \text{as } t \to \infty. \tag{2.86}$$

It follows that under the assumptions we have made, in particular that of no selection, we may reasonably assume that the equation

$$c_1c_4 - c_2c_3 = 0 \tag{2.87}$$

holds if the population has evolved for some time. It is important to establish what this equation means in genetical terms. Algebraic manipulation shows that (2.87) is equivalent to

$$\text{freq}(A_iB_j) = \text{freq}(A_i) \times \text{freq}(B_j) \tag{2.88}$$

for all possible pairs i, j. When (2.88), or equivalently (2.87), holds, the population is said to be in a state of *linkage equilibrium* with respect to these loci. The quantity $c_1c_4 - c_2c_3$, which we denote by D, is often called the "coefficient of linkage disequilibrium". As we see below, this can be a rather misleading expression for the quantity $c_1c_4 - c_2c_3$, which we would prefer to call the "coefficient of association". An alternative expression for D, sometimes more useful than $c_1c_4 - c_2c_3$, is

$$D = c_1 - \text{freq. } A_1 \times \text{freq. } B_1. \tag{2.89}$$

We turn now to the case where selective differences between genotypes exist. In the previous chapter we used a fitness display such as that in (1.92), which focusses attention on the genotypes at each of the two loci. For theoretical purposes, however, it is usually more convenient to adopt a notation focussed around the two gametes making up each individual. This is so since, as (2.82) shows, gametic frequencies are the most natural vehicle for studying evolutionary behavior in two-locus systems under random mating. We thus adopt the fitness scheme shown in (2.90) below:

	A_1B_1	A_1B_2	A_2B_1	A_2B_2
A_1B_1	w_{11}	w_{12}	w_{13}	w_{14}
A_1B_2	w_{21}	w_{22}	w_{23}	w_{24}
A_2B_1	w_{31}	w_{32}	w_{33}	w_{34}
A_2B_2	w_{41}	w_{42}	w_{43}	w_{44}

(2.90)

In the notation of this fitness scheme the fitness of zygotes made up of gametes i and j is written as w_{ij} (which we assume equal to w_{ji}). If coupling and repulsion double heterozygotes have the same fitness, then also $w_{23} = w_{14}$. We make this assumption throughout. If, for specific purposes, we wish to adopt a fitness display emphasizing single-locus genotypes, (2.90)

becomes

$$\begin{array}{lccc} & B_1B_1 & B_1B_2 & B_2B_2 \\ A_1A_1 & w_{11} & w_{12} & w_{22} \\ A_1A_2 & w_{13} & w_{14} & w_{24} \\ A_2A_2 & w_{33} & w_{34} & w_{44} \end{array} \tag{2.91}$$

The marginal fitness w_i of gamete i is defined by

$$w_i = \sum_j c_j w_{ij}, \tag{2.92}$$

and the mean fitness $\bar{w}$ of the population then becomes

$$\bar{w} = \sum\sum c_i c_j w_{ij} = \sum c_i w_i. \tag{2.93}$$

Consideration of all possible matings, their frequencies, and their genetic outputs, as well as the fitnesses of the various genotypes, shows that the gametic frequencies c_i' in the following generation are given by

$$c_i' = \bar{w}^{-1}\big(c_i w_i + \eta_i R w_{14}(c_2c_3 - c_1c_4)\big), \quad i = 1, 2, 3, 4. \tag{2.94}$$

Here η_i is defined in (2.84). If the w_{ij} are all equal, these recurrence relations reduce to (2.83). These important equations are due in this form to Lewontin and Kojima (1960), but they were essentially derived earlier, for a continuous-time model, by Kimura (1956b). Our present aim is to discuss some of the more immediate consequences of these equations.

First, the mean fitness, as defined in (2.93), is similar in form to the definition (2.10) with $k = 4$. It follows from the discussion in Section 2.4 that if we assume that mean fitness is maximized at a unique internal ($c_i > 0$) point, then at this point $w_i = \bar{w}$, where now w_i and $\bar{w}$ defined by (2.92) and (2.93). What is the connection between this maximization point and the equilibrium points of the system (2.94)? The equations $c_i' = c_i$ show that the system (2.94) is in equilibrium when

$$\bar{w} = w_i + c_i^{-1}\eta_i R w_{14}(c_2c_3 - c_1c_4), \quad i = 1 \ldots 4. \tag{2.95}$$

Unless linkage equilibrium holds at the equilibrium point, this point cannot be a point of maximum fitness. We show later that linkage equilibrium holds at equilibrium only in special cases, so that mean fitness can decrease in the system (2.94). The MFIT cannot then be true in general in two-locus selection systems. By contrast, we shall show in Section 7.4.5 that the FTNS does hold with a multilocus fitness scheme, and thus in particular with a two-locus fitness scheme.

We now demonstrate the possible decrease in mean fitness by a numerical example. Suppose, using the notation (2.91), that the fitness scheme is

$$\begin{array}{lccc} & B_1B_1 & B_1B_2 & B_2B_2 \\ A_1A_1 & 1.000 & 1.024 & 1.021 \\ A_1A_2 & 1.025 & 1.066 & 1.026 \\ A_2A_2 & 1.018 & 1.019 & 1.007 \end{array} \tag{2.96}$$

and let $R = \frac{1}{2}$, so that A and B loci unlinked. If initially

$$c_1 = 0.168, \quad c_2 = 0.362, \quad c_3 = 0.292, \quad c_4 = 0.178, \tag{2.97}$$

the population mean fitness is 1.033106. The mean fitness now *decreases* for about 14 generations and after that steadily increases, reaching a value of 1.031212 at the equilibrium point

$$c_1 = 0.24136, \quad c_2 = 0.28164, \quad c_3 = 0.22192, \quad c_4 = 0.25508. \tag{2.98}$$

The net effect of the evolution of the population from the starting point (2.97) to the equilibrium point (2.98) is to decrease mean fitness by 0.001894. At this equilibrium point the value of $D = c_1c_4 - c_2c_3$ is -0.000935.

Apart from the fact that mean fitness can decrease, the above analysis demonstrates two further points. The first is that the coefficient of linkage disequilibrium can be nonzero at an equilibrium point of the evolutionary system, even though the two loci upon which fitness depends are unlinked. This is why we prefer the term "coefficient of association" for the quantity $c_1c_4 - c_2c_3$, rather than the term "coefficient of linkage disequilibrium".

The second point to observe is that the location of the equilibrium point or points of (2.94) will depend on the recombination fraction R between the loci in those cases where linkage equilibrium does not obtain at equilibrium. Thus various values of R can be considered and the equilibrium mean fitnesses computed for each. When $R = 0$ the "equilibrium" equation (2.95) and the "maximization" equation $\bar{w} = w_i$ $(i = 1, \ldots, 4)$ agree, so that if each $c_i > 0$ at equilibrium, the value of R for which the greatest equilibrium mean fitness is achieved is for $R = 0$. This conclusion remains true if some of the c_i are zero at equilibrium but strangely, as we see later, it is not necessarily true that equilibrium mean fitness is a monotonically decreasing function of R. To the extent that equilibrium mean fitness is maximized for extremely tight linkage, the argument of Fisher given in Chapter 1 concerning the evolution of tight linkage between epistatic loci is justified. This argument can be made only when $D \neq 0$ at equilibrium for all R values: If $D = 0$ at equilibrium for all R the equilibrium mean fitness is independent of R.

The third topic we treat, at rather greater length, concerns the additive genetic variance in fitness. We are particularly interested in the relationship between this and the two marginal single-locus values, and we begin by

defining the latter. Using the fitness scheme (2.91), we may define the marginal fitnesses of the various single-locus genotypes as follows:

Genotype	Frequency	Marginal Fitness
A_1A_1	$(c_1+c_2)^2$	$(w_{11}c_1^2+2w_{12}c_1c_2+w_{22}c_2^2)/(c_1+c_2)^2=u_{11}$
A_1A_2	$2(c_1+c_2)\times$ (c_3+c_4)	$(w_{13}c_1c_3+w_{14}c_1c_4+w_{14}c_2c_3+w_{24}c_2c_4)/$ $(c_1+c_2)(c_3+c_4)=u_{12}$
A_2A_2	$(c_3+c_4)^2$	$(w_{33}c_3^2+2w_{34}c_3c_4+w_{44}c_4^2)/(c_3+c_4)^2=u_{22}$
B_1B_1	$(c_1+c_3)^2$	$(w_{11}c_1^2+2w_{13}c_1c_3+w_{33}c_3^2)/(c_1+c_3)^2=v_{11}$
B_1B_2	$2(c_1+c_3)\times$ (c_2+c_4)	$(w_{12}c_1c_2+w_{14}c_1c_4+w_{14}c_2c_3+w_{34}c_3c_4)/$ $(c_1+c_3)(c_2+c_4)=v_{12}$
B_2B_2	$(c_2+c_4)^2$	$(w_{22}c_2^2+2w_{24}c_2c_4+w_{44}c_4^2)/(c_2+c_4)^2=v_{22}$

(2.96)

From (1.42), the marginal additive genetic variance at the A locus may be defined as

$$2(c_1+c_2)(c_3+c_4)G_A^2, \tag{2.97}$$

where

$$G_A=u_{11}(c_1+c_2)+u_{12}(1-2c_1-2c_2)-u_{22}(c_3+c_4). \tag{2.98}$$

Similarly the marginal additive genetic variance at the B locus is

$$2(c_1+c_3)(c_2+c_4)G_B^2, \tag{2.99}$$

where

$$G_B=v_{11}(c_1+c_3)+v_{12}(1-2c_1-2c_3)-v_{22}(c_2+c_4). \tag{2.100}$$

We now find the two-locus additive genetic variance. To do this we assign additive parameters α_{11} and α_{12} to A_1 and A_2 and α_{21} and α_{22} to B_1 and B_2, and then minimize the expression

$$\begin{aligned}S=&c_1^2(w_{11}-\bar{w}-2\alpha_{11}-2\alpha_{21})^2+2c_1c_2(w_{12}-\bar{w}-2\alpha_{11}-\alpha_{21}-\alpha_{22})^2\\&+\cdots+c_4^2(w_{44}-\bar{w}-2\alpha_{12}-2\alpha_{22})^2\end{aligned}$$

with respect to the α_{ij}. Now that two loci are involved in the minimization is is appropriate to add constraints on the α_{ij}, since, for example, adding some constant to each α_{1x} and subtracting the same constant from each α_{2x} does not change the value of S. Such a change would, however, affect the definitions of marginal additive genetic variances. The natural constraints to impose are those which arise automatically in the one-locus case as given in (2.58). In the two-locus case these are

$$(c_1+c_2)\alpha_{11}+(c_3+c_4)\alpha_{12}=0,\quad (c_1+c_3)\alpha_{21}+(c_2+c_4)\alpha_{22}=0, \tag{2.101}$$

and the minimization is carried out subject to these constraints. Details of this procedure are given by Kojima and Kelleher (1961) and Kimura (1965)

and are not pursued here. It is found that the additive genetic variance can be written as

$$2\{(c_1+c_2)(c_3+c_4)H_A^2+2H_AH_BD+(c_1+c_3)(c_2+c_4)H_B^2\}, \quad (2.102)$$

where H_A and H_B are the solutions of the equations

$$\begin{aligned} H_A+\{(c_1+c_2)(c_3+c_4)\}^{-1}DH_B&=G_A,\\ H_B+\{(c_1+c_3)(c_2+c_4)\}^{-1}DH_A&=G_B, \end{aligned} \quad (2.103)$$

G_A and G_B being given by (2.98) and (2.100).

Several interesting conclusions follow from these equations. Perhaps the most important is that if $D=0$ (that is, linkage equilibrium between the two loci) then $H_A=G_A$, $H_B=G_B$, and the true two-locus additive genetic variance is the sum of the two single-locus marginal values. When $D\neq 0$ this is no longer true, and there is no simple relationship between this sum and the true two-locus additive genetic variance value. This is an important conclusion since it seems to be widely assumed in the classical literature (see for example Fisher (1918, p. 405), (1958, p. 37) and Wright (1969, p. 439)) that in a multilocus system the true additive genetic variance can be found by simply summing single-locus marginal values. Since we have shown above that changes in mean fitness can be negative in two-locus systems, and thus cannot be equal to any form of genetic variance, it follows that

$$\Delta\bar{w}, \quad \sigma_A^2(\text{two-locus}), \quad \sum\sigma_A^2(\text{single-locus marginals}) \quad (2.104)$$

have in general no clear and obvious connection with each other. This conclusion is generalized in Section 7.3.3.

These conclusions may also be associated with properties of changes in gene frequency. Equations (2.97), (2.99), and (2.102) show that

$$\sigma_A^2(\text{two-locus})-\sum\sigma_A^2(\text{single-locus marginals})=2D(G_AH_B+H_AG_B), \quad (2.105)$$

and if D is small this may be approximated by $-4DG_AG_B$. Since

$$\Delta(\text{frequency } A_1)=(c_1+c_2)(c_3+c_4)G_A/\bar{w},$$

with a corresponding expression for Δ(frequency B), it is found, if terms of order D^2 are ignored, that the left-hand side in (2.105) may be written

$$\frac{-4D\bar{w}^2\Delta(\text{frequency } A_1)\Delta(\text{frequency } B_1)}{(c_1+c_2)(c_3+c_4)(c_1+c_3)(c_2+c_4)}.$$

This gives an interesting relationship between the various additive genetic variances, the linkage disequilibrium, and the gene frequency changes in a two-locus system. If in a certain generation Δ(frequency A_1) $=0$, then to the order of accuracy we use the equation $G_A=0$ holds, and the total additive genetic variance is simply the marginal B locus value. However, this is true only as an approximation and, more precisely, whenever there

is linkage disequilibrium between A and B loci there is a small perturbation from the A locus to the total additive variance, even though gene frequencies are not changing at that locus.

We expect the additive genetic variance to be of importance in discussing the correlation between relatives. Before exploring this, we recall that gene frequencies alone are not sufficient to describe the evolution of two-locus systems, so that it is reasonable to argue that the additive genetic variance, which fundamentally involves gene frequencies, is not the appropriate component of variance for evolutionary considerations. We thus consider a variance defined by gamete frequencies which, since gamete frequencies do describe the evolutionary behavior, might be thought to be of greater evolutionary significance that the additive genetic variance.

The marginal fitnesses w_i of the four gametes have been defined in (2.92). The total chromosomal, or gametic, variance in fitness, denoted σ_G^2, may be defined by

$$\sigma_G^2 = 2\sum_{i=1}^{4}(w_i - \bar{w})^2 c_i, \tag{2.106}$$

the factor 2 being inserted because there are two gametes per zygote. Suppose now we attempt to fit the marginal gametic fitnesses by additive components depending on the genes on each gamete. This is done by minimizing

$$\begin{aligned} &c_1(w_1 - \bar{w} - \alpha_{11} - \alpha_{21})^2 + c_2(w_2 - \bar{w} - \alpha_{11} - \alpha_{22})^2 \\ &+c_3(w_3 - \bar{w} - \alpha_{12} - \alpha_{21})^2 + c_4(w_4 - \bar{w} - \alpha_{12} - \alpha_{22})^2 \end{aligned}$$

with respect to α_{11}, α_{12}, α_{21} and α_{22}, subject to the constraints in (2.101). The sum of squares so removed may be described as being due to the additive effects of genes within gametes, and for short may be called the additive gametic variance. It is found (see Kimura, (1965)) that this is identical to the additive genetic variance (2.102) and thus the latter, perhaps unexpectedly, is of use in evolutionary and other considerations. This conclusion is generalized in Section 7.3.3. The total gametic variance in (2.106) has three degrees of freedom, of which the additive component of it has two. The remaining degree of freedom is taken up by the epistatic gametic variance σ_{EG}^2, which is

$$\sigma_{EG}^2 = 2(w_1 - w_2 - w_3 + w_4)^2/(c_1^{-1} + c_2^{-1} + c_3^{-1} + c_4^{-1}). \tag{2.107}$$

This is zero if and only if an additive genetic fitness scheme exactly fits the marginal gametic fitnesses.

We turn now to the correlation between relatives, restricting attention to the case where (2.88) holds, that is that the two loci are in linkage equilibrium. This assumption was also made by Fisher (1918). We consider both linked and unlinked loci: Fisher's 1918 analysis is concerned only with

the unlinked case. Our treatment is based on Cockerham (1954, 1956) and Kempthorne (1954).

We first isolate various components of the total variance of the character measured. Suppose that the measurements for the various genotypes are

$$\begin{array}{cccc} & B_1B_1 & B_1B_2 & B_2B_2 \\ A_1A_1 & m_{11} & m_{12} & m_{13} \\ A_1A_2 & m_{21} & m_{22} & m_{23} \\ A_2A_2 & m_{31} & m_{32} & m_{33} \end{array} \tag{2.108}$$

We form these measurements into a single vector $\mathbf{m} = (m_{11}, m_{12}, \ldots, m_{33})'$. If the frequency of A_1 is x and of B_1 is y, then since linkage equilibrium is assumed, the frequency of $A_1A_1B_1B_1$ is x^2y^2, of $A_1A_1B_1B_2$ is $2x^2y(1-y)$ and so on. It is convenient to write these frequencies as the entries in a diagonal matrix F, so that

$$F = \begin{pmatrix} x^2y^2 & & & \\ & 2x^2y(1-y) & & 0 \\ & & \ddots & \\ & 0 & & (1-x)^2(1-y)^2 \end{pmatrix}. \tag{2.109}$$

Evidently the mean value $\bar{m}$ in the measurement is given by

$$\bar{m} = x^2y^2m_{11} + 2x^2y(1-y)m_{12} + \cdots + (1-x)^2(1-y)^2m_{33}. \tag{2.110}$$

Further, adopting the notation of (2.96), the marginal means of A_1A_1, A_1A_2 and A_2A_2 are

$$\begin{aligned} u_{11} &= y^2m_{11} + 2y(1-y)m_{12} + (1-y)^2m_{13}, \\ u_{12} &= y^2m_{21} + 2y(1-y)m_{22} + (1-y)^2m_{23}, \\ u_{22} &= y^2m_{31} + 2y(1-y)m_{32} + (1-y)^2m_{33}. \end{aligned} \tag{2.111}$$

Similarly the marginal means at the B locus are

$$\begin{aligned} v_{11} &= x^2m_{11} + 2x(1-x)m_{21} + (1-x)^2m_{31}, \\ v_{12} &= x^2m_{12} + 2x(1-x)m_{22} + (1-x)^2m_{32}, \\ v_{22} &= x^2m_{13} + 2x(1-x)m_{23} + (1-x)^2m_{33}. \end{aligned} \tag{2.112}$$

Finally the total variance σ^2 in the character measured is

$$\sigma^2 = x^2y^2m_{11}^2 + \cdots + (1-x)^2(1-y)^2m_{33}^2 - \bar{m}^2 = \mathbf{m}'F\mathbf{m} - \bar{m}^2. \tag{2.113}$$

This total variance has eight degrees of freedom, and our aim is to break it down into the sum of eight components, each having one degree of freedom and each being of genetical significance. These components will measure two additive variances, one at each of the two loci, two dominance variances, one at each of the two loci and the four interaction variances.

Suppose a matrix T exists such that $TFT' = I$ (or equivalently $(T')^{-1}F^{-1}T^{-1} = I$), where I is the unit 9×9 matrix, and define a vector

$\mathbf{z}$ by $\mathbf{z} = TF\mathbf{m}$. Then

$$\begin{aligned}\mathbf{m}'F\mathbf{m} &= \mathbf{z}'(T')^{-1}F^{-1}FF^{-1}T^{-1}\mathbf{z} \\ &= \mathbf{z}'\mathbf{z} \\ &= z_1^2 + z_2^2 + \cdots + z_9^2. \end{aligned} \tag{2.114}$$

If the last row in T can be chosen to be $(1, 1, \ldots, 1)$, then $z_9 = \bar{m}$ and

$$\sigma^2 = z_1^2 + z_2^2 + \cdots + z_8^2. \tag{2.115}$$

The equation $TFT' = I$ reduces to the requirement

$$x^2y^2t_{i1}t_{j1} + 2x^2y(1-y)t_{i2}t_{j2} + \cdots + (1-x)^2(1-y)^2t_{i9}t_{j9} = \delta_{ij}, \tag{2.116}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. The choice $t_{91} = t_{92} = \ldots = t_{99} = 1$ does satisfy (2.116) with $i = j = 9$. Thus σ^2 can indeed be broken down into the sum (2.115), where

$$z_i = x^2y^2t_{i1}m_{11} + 2x^2y(1-y)t_{i2}m_{12} + \cdots + (1-x)^2(1-y)^2t_{i9}m_{33}, \tag{2.117}$$

provided that the t_{ij} satisfy (2.116) and the further requirement

$$x^2y^2t_{i1} + 2x^2y(1-y)t_{i2} + \cdots + (1-x)^2(1-y)^2t_{i9} = 0, \quad i = 1\ldots, 8. \tag{2.118}$$

Apart from these purely mathematical requirements we wish to choose that t_{ij} so that the z_i have the genetical interpretations described above.

Suppose z_1^2 and z_2^2 are to represent the additive and dominance variance components of the character from the A locus. Recalling equations (1.9) and using the marginal fitness values (2.111), we would like to have

$$\begin{aligned} z_1^2 &= 2x(1-x)\{xu_{11} + (1-2x)u_{12} - (1-x)u_{22}\}^2, \\ z_2^2 &= x^2(1-x)^2\{2u_{12} - u_{11} - u_{22}\}^2. \end{aligned} \tag{2.119}$$

Such a representation is in fact possible if, in (2.117), we choose

$$\begin{aligned} t_{11} = t_{12} = t_{13} &= x^{-1}\{2x(1-x)\}^{1/2}, \\ t_{14} = t_{15} = t_{16} &= (1-2x)\{2x(1-x)\}^{-1/2}, \\ t_{17} = t_{18} = t_{19} &= -(1-x)^{-1}\{2x(1-x)\}^{1/2}, \end{aligned} \tag{2.120}$$

and

$$\begin{aligned} t_{21} = t_{22} = t_{23} &= -x^{-1}(1-x), \\ t_{24} = t_{25} = t_{26} &= 1, \\ t_{27} = t_{28} = t_{29} &= (1-x)^{-1}x. \end{aligned} \tag{2.121}$$

These choices do satisfy the requirements (2.116) and (2.118), and thus our desired representation (2.119) is allowable. A parallel procedure gives additive and dominance variance components at the B locus as

$$z_3^2 = 2y(1-y)\{yv_{11} + (1-2y)v_{12} - (1-y)v_{22}\}^2$$

and

$$z_4^2 = y^2(1-y)^2\{2v_{12} - v_{11} - v_{22}\}^2.$$

Once more, with the choice of the t_{ij} implicit in these definitions, the orthogonality conditions are met. If z_5^2 is to represent the additive-by-additive component of the total variance it would be natural to choose $t_{5i} = t_{1i} \times t_{3i}$, and the remaining three interactive components would naturally be chosen by similar multiplications. If this is done it is found that all the orthogonality conditions are met, and this also implies that the representation (2.115) is completed. We do not go into details here and note only that the various components can be expressed as

$$\begin{aligned}
&(\text{add} \times \text{add}):\ z_5^2 = 4xy(1-x)(1-y)\{xye_{11} + x(1-y)e_{12} + (1-x)ye_{21} \\
&\qquad\qquad + (1-x)(1-y)e_{22}\}^2, \qquad (2.122)\\
&(\text{add} \times \text{dom}):\ z_6^2 = 2x(1-x)y^2(1-y)^2\{x(e_{11} - e_{12}) + (1-x)(e_{21} - e_{22})\}^2,\\
&(\text{dom} \times \text{add}):\ z_7^2 = 2x^2(1-x)^2y(1-y)\{y(e_{11} - e_{21}) + (1-y)(e_{12} - e_{22})\}^2,\\
&(\text{dom} \times \text{dom}):\ z_8^2 = x^2y^2(1-x)^2(1-y)^2\{e_{11} - e_{12} - e_{21} + e_{22}\}^2,
\end{aligned}$$

where

$$\begin{aligned}
e_{11} &= m_{11} - m_{12} - m_{21} + m_{22},\\
e_{12} &= m_{12} - m_{13} - m_{22} + m_{23},\\
e_{21} &= m_{21} - m_{22} - m_{31} + m_{32},\\
e_{22} &= m_{22} - m_{23} - m_{32} + m_{33}.
\end{aligned}$$

These expressions, given more generally to include the effect of inbreeding, were derived by Cockerham (1954). It is sometimes convenient to write

$$\sigma_A^2 = z_1^2 + z_3^2, \quad \sigma_D^2 = z_2^2 + z_4^2, \quad \sigma_{AA}^2 = z_5^2, \quad \sigma_{AD}^2 = z_6^2 + z_7^2, \quad \sigma_{DD}^2 = z_8^2,$$

so that

$$\sigma^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2. \qquad (2.123)$$

A slightly shorter representation collects the final three terms as a single term σ_E^2 (epistatic variance), but for our purposes this is not useful, since the final three terms in (2.123) are involved differently in the correlation between relatives, and are therefore best kept separate.

Consider now the father-son and the full sib correlations in the measurement. It is possible to write down all 81 father-son genotypic combinations and, using a table extending Table 1.1, arrive at a father-son covariance. By doing this and a parallel procedure for full sibs, it is found that if the A and B loci are unlinked,

$$\text{corr(father-son)} = \left(\tfrac{1}{2}\sigma_A^2 + \tfrac{1}{4}\sigma_{AA}^2\right)/\sigma^2, \qquad (2.124a)$$

$$\text{corr(full sibs)} = \left(\tfrac{1}{2}\sigma_A^2 + \tfrac{1}{4}\sigma_D^2 + \tfrac{1}{4}\sigma_{AA}^2 + \tfrac{1}{8}\sigma_{AD}^2 + \tfrac{1}{16}\sigma_{DD}^2\right)/\sigma^2. \qquad (2.124b)$$

Cockerham (1956) demonstrated that, when the two loci are linked, the former expression remains unchanged but that the latter must be replaced by

$$\begin{aligned}\text{corr(full sibs)} = \{&\tfrac{1}{2}\sigma_A^2 + \tfrac{1}{4}\sigma_D^2 + \tfrac{1}{8}(3 - 4R + 4R^2)\sigma_{AA}^2 \\ &+ \tfrac{1}{4}(1 - 2R + 2R^2)\sigma_{AD}^2 \\ &+ \tfrac{1}{4}(1 - 2R + 2R^2)^2\sigma_{DD}^2\}/\sigma^2. \end{aligned} \tag{2.125}$$

The effect of linkage is always to increase the full sib correlation compared to the value for the unlinked case. We derive these formulas later in Chapter 7 as particular cases of correlations where the trait in question depends on an arbitrary number of loci, using a more efficient approach.

The analysis in this section has assumed a discrete-time model, and it is expected that qualitatively similar conclusions would hold for a continuous model. One possible complication for such models does, however, occur. In the discrete models the frequency of any genotype is found immediately from the frequencies of the gametes making up this genotype, so that, for example,

$$\text{freq}(A_1A_1B_1B_2) = 2\,\text{freq}(A_1B_1)\,\text{freq}(A_1B_2). \tag{2.126}$$

In the continuous-time model of Nagylaki and Crow (1974) the existence of linkage disequilibrium between the two loci implies that "Hardy–Weinberg" equations such as (2.126) are no longer true. This is of some interest since many theoretical analyses of continuous-time two-locus models have assumed the truth of equations like (2.126). However, Nagylaki (1976) has shown that when fitness differentials are small a state of "quasi-Hardy–Weinberg" soon emerges when genotypic frequencies can, to a very close approximation, be found from the constituent gametic frequencies.

2.11 Genetic Loads

A genetic load is said to arise if the population mean fitness is less than that of some optimal value which in some idealized sense it could take. The two forms of genetic load that have caused considerable controversy in the literature are the substitutional load and the segregational load. In both cases the load ℓ is defined by

$$\ell = (w_{\max} - \bar{w})/\bar{w}, \tag{2.127}$$

where $w_{\max}$ is the fitness of the most fit genotype and $\bar{w}$ is the mean fitness. If we normalize fitnesses so that the mean fitness is 1, we replace (2.127) by

$$\ell \approx w_{\max} - 1. \tag{2.128}$$

Our aim in this section is to analyze the formal calculations for both forms of load. These formal calculations have remained implicit rather than explicit in the analyses of proponents of genetic loads as calculated by the formula (2.128). Before doing this we briefly review the historical context.

The load concept was introduced by Haldane (1957, 1961) in the substitutional case. As a result of his load calculations, Haldane placed a quite conservative limit on the rate at which favorable new alleles at different loci, arising perhaps by mutations or perhaps by an environmental change rendering a previously unfavorable allele favorable, could spread throughout a population. Specifically, he came to the conclusion that as a result of what became known as the substitutional load (his "cost of natural selection"), substitutional processes at different loci could not start more frequently than about 300 generations apart.

As we observe below, a load in effect refers to a variance in fitness, not to a mean fitness. The essence of the substitutional load argument is that if many selectively driven substitutional processes are occurring in some population at any given time, then there will exist a substantial variance in fitness of this individuals in the population of interest at that time. Individuals carrying the favored allele at all the loci substituting will then have a very high fitness, that is will be required to produce an extremely large number of offspring. This is in effect the substitutional load placed on the population.

The load concept was subsequently extended to define a segregational load, the motivation being the observation, in the 1960's, that there exists considerable genetic variation in natural populations. The segregational load argument claimed that under a selective explanation for the variation, perhaps because of heterozygote advantage at many of the loci exhibiting genetic variation, the most fit individuals in the population would again have a very high fitness and thus would be required to produce an extremely large number of offspring. This led to comments such as that of Dobzhansky (1970, page 220), that "higher vertebrates and man do not possess enough 'load space' to maintain more than a few balanced polymorphisms," leading to the view (page 224) that selection favoring heterozygotes "cannot explain the polymorphisms observed in man." At about the same time, segregational load arguments and subsequently substitutional load arguments were used by Kimura (1968) to support his neutral theory of evolution. The aim of this section is to show that the (implicit) arguments of Haldane, Dobzhansky and Kimura are all unjustified.

The segregational and substitutional genetic load "problem" arises when segregation occurs or substitutions take place at many loci simultaneously. The implicit assumption made in load calculations by proponents of the load concept is that multilocus fitnesses are obtained by first constructing single locus fitnesses and then multiplying these over the loci segregating or substituting. We initially make this (surely unrealistic) assumption so as

to follow load calculations and arguments, but later discuss more realistic fitness models.

We start with a discussion of the segregational load. This load exists because of segregation at a number of loci arising from heterozygote selective advantage at each locus. For simplicity we assume two alleles segregating at each locus and with a fitness scheme where, at each locus, each homozygote has fitness $1 - \frac{1}{2}s$ and the heterozygote has fitness $1 + \frac{1}{2}s$. Thus with the multiplicative assumption, and with two loci segregating, the two-locus fitness scheme (2.91) would be

	B_1B_1	B_1B_2	B_2B_2
A_1A_1	$(1-\frac{1}{2}s)^2$	$(1-\frac{1}{2}s)(1+\frac{1}{2}s)$	$(1-\frac{1}{2}s)^2$
A_1A_2	$(1-\frac{1}{2}s)(1+\frac{1}{2}s)$	$(1+\frac{1}{2}s)^2$	$(1-\frac{1}{2}s)(1+\frac{1}{2}s)$
A_2A_2	$(1-\frac{1}{2}s)^2$	$(1-\frac{1}{2}s)(1+\frac{1}{2}s)$	$(1-\frac{1}{2}s)^2$

With many loci segregating the multilocus fitness scheme is the natural generalization of the two-locus fitness scheme above. We emphasize again that this model is discussed here since this is the model implicitly assumed in load calculations.

The equilibrium properties of this model are not straightforward. We shall see later (see (6.33)) that when the recombination fraction R between A and B loci is sufficiently large, the stable equilibrium frequencies of A_1, A_2, B_1 and B_2 are all $1/2$, and the mean fitness is 1, as a straightforward multiplication of single-locus values would suggest. However, when R is sufficiently small the picture is more complicated and the population mean fitness exceeds 1 at the stable equilibrium point of the system. We defer consideration of this case until later and assume for the moment the "loose linkage" case.

More generally, for m sufficiently loosely linked loci and a multiplicative fitness model generalizing the two-locus scheme above, the equilibrium frequencies of all alleles at all loci are $1/2$. Any individual is a heterozygote at j of these m loci with probability

$$\binom{m}{j}\left(\frac{1}{2}\right)^m,$$

so that the equilibrium population mean fitness is

$$\sum_{j=0}^{m}\binom{m}{j}\left(\frac{1}{2}\right)^m\left(1+\frac{1}{2}s\right)^j\left(1-\frac{1}{2}s\right)^{m-j} \;=\; 1. \tag{2.129}$$

An individual heterozygous at all loci has fitness $(1+\frac{1}{2}s)^m$, and a formal application of the definition (2.128) implies that the segregational load is

$$\left(1+\frac{1}{2}s\right)^m - 1 \;\approx\; e^{sm/2} - 1. \tag{2.130}$$

This can be substantial for large values of m, and this the formal calculation directly leads to the segregational load "problem".

We return to this calculation below, and turn next to the substitutional load. We consider first the substitution process at one single gene locus, and initially, to follow formal substitutional load calculations, we do *not* scale fitnesses to make the mean population fitness equal to 1.

Suppose that at the locus of interest, fitnesses of the form (1.25b) apply, with $s > 0$. It is convenient, and does not materially affect the substance of the argument, to assume that $h = 0.5$. Then because of natural selection, the frequency of the allele A_1 will steadily increase in the population. When the frequency of A_1 is x the population mean fitness is $1 + sx$, and the load as defined by (2.128), is $s(1 - x)$. The overall substitutional load L for the entire substitution process is defined as the sum of this quantity during the process when x increases from a small value x_1 (at time t_1) to a value x_2 close to unity (at time t_2). Thus

$$\begin{aligned} L &= \sum s(1-x) \\ &\approx \int_{t_1}^{t_2} s(1-x)\,dt \\ &= 2\int_{x_1}^{x_2} x^{-1}\,dx \qquad \text{from (1.27)} \\ &= 2\log(x_2/x_1). \end{aligned}$$

Since x_2 is close to 1, this differs only trivially from $-2\log x_1$. Unfortunately the value chosen for x_1 will depend to a large extent on the view one takes of the most likely form of genetic evolution, and the discussion in Section 1.7 becomes relevant to the argument. A value often chosen for evolutionary load arguments is $x_1 = 0.0001$, and this gives $L = 18.4$. When $h \neq 0.5$ the load as calculated using this form of calculation usually exceeds 18.4, and for operational purposes the "representative value" $L = 30$ is generally used in the load argument. We therefore adopt this value also.

What does this calculation mean for the offspring requirement of the individuals in any given generation? Suppose that all selection is through viability differences and the number of reproducing adults in each generation remains constant at N. A considerable proportion of the depletion in population numbers between birth and the age of reproduction is non-genetic. Taking only the genetic component, and supposing there is no depletion through genetic deaths of the optimal genotype A_1A_1, a straightforward calculation shows that when the frequency of A_1 is x, there must be $N(1+s)/(1+sx)$ individuals at birth, so that after differential viabilities operate there are N individuals at the age of maturity. Thus the average individual is required to leave approximately $1 + s(1 - x)$ offspring after

non-genetic deaths are taken into account, so that there will be $Ns(1-x)$ "genetic deaths" in each generation associated with the evolutionary process. Summed over the entire process this gives NL individuals in all. If each substitutional process takes T generations, this implies an average of NL/T such "deaths" in each generation.

Consider now a sequence of loci at which substitutions start regularly n generations apart. For convenience it is assumed that the same fitness parameters apply for all these loci as for the single locus discussed above. As in the segregational load argument, it is implicitly assumed in load arguments that fitnesses are multiplicative over loci, so initially we make this assumption also. As with the segregational load, the substitutional load relates to the fitness, or offspring requirement, of an individual of the most genotype. In this case this is an individual with the superior genotype "A_1A_1" at each locus undergoing substitution.

At any one time there will be T/n substitutions in progress and thus a total of $(NL/T)(T/n) = NL/n$ "selective deaths" per generation. From this it is found that the offspring requirement of the most fit individual, assuming the multiplicative model of fitness with and with linkage equilibrium always holding between loci, is

$$(1 + L/T)^{T/n} \approx \exp(L/n) \approx \exp(30/n) \tag{2.131}$$

if we take the "representative value" 30 for L as discussed above.

The value $n = 300$ reached by Haldane (1957), as described above, arises from the fact that with this value of n, the expression in (2.131) is about 1.1, conforming to his view that an "excess reproductive requirement" of 10% is the maximum that can be expected, at least in mammals.

Kimura and Ohta (1971a) estimated that in the evolutionary history of mammals approximately six substitutions have been completed per generation in any evolutionary line. This implies that $n = 1/6$, 1800 times smaller than the Haldane "limiting" value, or equivalently implying substitutions occurring at 1800 times the upper rate as calculated by Haldane. Insertion of the value $n = 1/6$ in (2.131) leads to a substitutional load of $e^{180} \approx 10^{78}$. This form of calculation was a major factor in the development of the neutral theory, since it was argued (Kimura (1968)) that the amount of genetic substitution estimated to have taken place in evolution, in particular in mammalian evolution, could not be explained by selective processes because of a claimed unbearable substitutional genetic load that selective substitutions would imply. Thus (Kimura and Ohta, 1971a) claimed that

> "to carry out mutant substitution at the above rate, each parent must leave $e^{180} \approx 10^{78}$ offspring for only one of the offspring to survive. This was the main reason why random fixation of selectively neutral mutants was first proposed by one of us as the main factor in molecular evolution."

Because of calculations and claims of this type, it is clearly necessary to discuss the assumptions, both explicit and implicit, in formal load calculations.

We start with the expression in (2.131), and observe that this expression refers not to the offspring requirement of every individual, as is implied in the above quotation, but to the requirement of an individual of the maximum possible fitness when the population mean fitness is now scaled to 1. It is therefore appropriate to focus on this individual and on his fitness.

Our calculations show that the fitness e^{180} is arrived at by assuming that fitnesses are multiplicative over loci. This is a quite unreasonable assumption, and the large offspring requirement of the most fit individual is a direct consequence of it. It is certainly true that in nature substantial epistasis occurs, and if this is so there will be a considerable reduction to the load from that calculated formally by using marginal fitnesses and multiplicativity, as discussed below. The unreasonableness of the multiplicative assumption was stressed long ago, in particular by Wright (1930).

The second, and more important, problem concerns the very existence of an individual of the optimal multilocus genotype. It is extremely unlikely that such an individual ever exists. To simplify the argument we continue to consider the multiplicative case discussed above. It can be shown that with the individual locus fitness values $1+s, 1+s/2, 1$ for "A_1A_1, A_1A_2 and A_2A_2", as is assumed above, and with $s = 0.01$, $n = 1/6$, initial frequency $= 0.0001$, final frequency $= 0.9999$, there will be 22,080 loci substituting at any one time. The various favored alleles at each of these 22,080 loci will take a variety of frequencies in $(0, 1)$, and in particular at those loci where the substitution has only recently started, the frequency of the favored allele will be quite low. By calculating the means of the frequencies x_1, $x_2, \ldots$ of the favored allele at the various loci substituting, using (1.28), it is found that the probability that an individual taken at random is of this optimal genotype is on the order of $10^{-23,200}$. This value is so extremely small that a theory basing its numerical computations on the offspring requirement of such an individual must demand reconsideration. This point also was stressed by Wright (1977, p. 481).

What is needed is a calculation of the fitness of the individuals who might reasonably be expected to occur in the population of interest. Here the finite size of any population is an important factor in the calculations. Some progress on amending load calculations for this purpose may be made by using the statistics of extreme values in a population of given finite size (Kimura (1969), Ewens (1970)). It is convenient, for purposes of illustration only, to maintain the multiplicativity assumption here so as to discuss the point at issue. The starting point is to find the variance of the distribution of the fitness of an individual taken at random from the population, if the population mean fitness is scaled to unity. In the case considered above this variance is s/n (Ewens (1970), Crow and Kimura (1970, p. 252)). For $s = 0.01$, $n = 1/6$, this is a variance of 0.06, so that the standard deviation

in fitness is approximately 0.245. The rather low value for this standard derivation arises because it is most unlikely that any individual will have a genetic constitution which differs markedly, in terms of the number of favored genes carried, from the average.

If s is extremely small we may suppose, to a first approximation, that the distribution of fitness is a normal distribution. The statistical theory of of extreme values (see Pearson and Hartley (1958, Table 28)) shows that, for example in a population of size 10^5, the most fit individual that is likely to occur will have a fitness approximately four standard deviations in excess of the mean. In the present case this implies a fitness of $1+4(0.245) = 1.98$. On average, then, the most fit individual that is likely to exist in the population is required to produce only about two offspring in order to effect the gene substitutions observed. This is clearly an easily achievable goal.

A parallel argument holds for the segregational load as calculated in (2.130). The segregational load is clearly the excess over the mean of the offspring of the most fit individual, in the segregation load case the multiple heterozygote. The probability that an individual chosen at random in the population is of this genotype is $(1/2)^m$, and when m is large it is extremely unlikely that any individual in a population even of size several million has this genotype. As with the substitutional load, it is more reasonable to consider the fitness of the most fit individual likely to arise in the population. This is done as follows.

The mean fitness of the population is calculated in (2.129). The variance in fitness then found as

$$\sum_{j=0}^{m} \binom{m}{j} (\tfrac{1}{2})^m (1 + \tfrac{1}{2}s)^{2j} (1 - \tfrac{1}{2}s)^{2(m-j)} - 1. \tag{2.132}$$

This expression reduces to

$$(1 + \frac{1}{4}s^2)^m - 1 \approx e^{ms^2/4} - 1. \tag{2.133}$$

For the case $m = 10,000, s = 0.01$ this is about 0.28. A fitness four standard deviations above the mean is only just in excess of 3, and arguing as above for the substitutional load, this clearly is an achievable fitness for the most fit individual likely to arise in a population of size 10^5.

The essence of the argument, in both the substitutional load and the segregational load cases, is that in a finite population only a minute proportion of all theoretically possible genotypes are realized, and that those that are realized are not normally very "extreme". In particular the fitness of the most fit existing genotype is not extreme, and in the substitution case, substitutions at the required rate can easily be achieved through each individual's producing as many offspring as this most fit existing genotype, with consequent differential viability effecting the required substitutions.

There are many further arguments that make the substitutional load calculations leading to the value e^{180} of dubious value. First, it has been

assumed in all the calculations that selection arises entirely through viability differences. To the extent that fertility selection occurs, the offspring requirement is correspondingly lowered, in the sense that the calculation of the offspring requirement of the most fit individual is not a calculation of any relevance to the average individual.

Second, it has been assumed so far that fitnesses are fixed constants, and are not, for example, frequency-dependent. It is possible to devise frequency-dependent selection schemes for which there is no segregational load at a stable equilibrium. Thus in the fitness scheme

$$\begin{array}{lll} A_1A_1 & A_1A_2 & A_2A_2 \\ 1+a(1-2x) & 1 & 1-a(1-2x) \end{array} \tag{2.134}$$

where x is the frequency of A_1 and a is a small parameter, the point $x = 0.5$ is a point of stable equilibrium, and at this point all genotypes have equal fitness and there is no genetic load. On the other hand, it is unlikely that frequency-dependent fitnesses can reduce the substitutional load to zero, since with a change in gene frequencies due to selection, some selective differentials are necessary and hence some load. Little information is available on the extent to which frequency-dependent selection can reduce substitutional load.

We now consider the effects of linkage disequilibrium, and later of epistasis and linkage disequilibrium jointly, on load calculations. Stationary points of an evolutionary system exhibiting linkage disequilibrium generally have a higher mean fitness than points where linkage equilibrium holds at stationarity, and thus have a lower genetic load than that at linkage equilibrium equilibria. This is particularly so when the selective system implies epistasis. However, even in the simple multiplicative case, where we can say there is no multiplicative epistasis, the stable equilibrium points of the evolutionary system can display linkage disequilibrium and thus a decreased segregational load. For example, the calculations of Franklin and Lewontin (1970) show that in the case of 36 equally spaced linked loci, a multiplicative fitness scheme generalizing the two-locus multiplicative fitness scheme above with $s = 0.1$, and with recombination fraction 0.0025 between adjacent loci, the load when calculated from (2.130) is about 5, but when calculated using the actual population mean fitness is about 1.6. The smaller load arises from the linkage disequilibrium arising for this model. This point has also in effect been made by Lewontin (1974, pp. 289–290) in the context in discussing the effect of linkage disequilibrium on mean fitness.

Next, the joint effects of epistasis and linkage disequilibrium can decrease the segregational load substantially. Thus, for example, numerical computation shows that with the epistatic scheme (2.96) and with $R = 0.001$, there is a stable equilibrium set of gametic frequencies at

$$c_1 = 0.013, \quad c_2 = 0.469, \quad c_3 = 0.503, \quad c_4 = 0.015. \tag{2.135}$$

At this point the population mean fitness is 1.0417 and thus the genetic load as defined by (2.128) is 0.0233.

Suppose now that marginal fitness values for this case are found from (2.111), and the load calculated according to (2.127) using these marginal values and the marginal genotypic frequencies. The loads so calculated are 0.0212 for the A locus and 0.0210 for the B locus. The sum of these is almost twice that of the true load: For $R = 0$ it would be exactly twice. Evidently for general fitness schemes involving tight linkage and epistasis, the procedure leading to the load calculation of e^{180}, namely the calculation of a multilocus segregational load through an amalgamation of single-locus segregational load calculations, can lead to serious errors.

If we take into account, then, the unreasonable multiplicative fitness requirement implicit in load calculations, the unreasonable concentration on the fitness requirement of essentially impossible genotypes, the possibility of very substantial linkage disequilibria, the possibility of frequency-dependent fitnesses and a variety of other ecological and evolutionary arguments concerning the real nature of selective processes, it appears that there is no reason for load arguments to imply very conservative bounds on the number of loci that can undergo simultaneous selective substitution processes, no "load space" argument limiting the number of balanced polymorphisms arising at any one time in a population, and no load theory support for the neutral theory of evolution.

2.12 Finite Markov Chains

Some of the arguments presented later in this book use the theory of finite Markov chains, and in this section a brief and informal introduction to the theory of these is presented.

Consider a discrete random variable X which at time points 0, 1, 2, 3, ... takes one or other of the values 0, 1, 2, ..., M. We shall say that X, or the system, is in state E_i if X takes the value i. Suppose that at some time t, the random variable X is in state E_i. Then if the probability p_{ij} that at time $t + 1$, the random variable is in state E_j is independent of t and also of the states occupied by X at times $t - 1$, $t - 2$, ..., the variable X is said to be Markovian, and its probability laws follow those of a finite Markov chain. If the initial probability (at $t = 0$) that X is in E_i is a_i then the probability that X is in the state E_i, E_j, E_k, E_ℓ, E_m ... at times 0, 1, 2, 3, 4 ... is $ap_ip_{ij}p_{jk}p_{k\ell}p_{\ell m}\cdots$.

Complications to Markov chain theory arise if periodicities occur, for example, if X can return to E_i only at the time points t_1, $2t_1$, $3t_1$, ... for some integer $t_1 > 1$. Further minor complications arise if the states E_0, E_1, ..., E_M can be broken down into noncommunicating subsets. To avoid unnecessary complications, which never in any event arise in genetical

applications, we suppose that no periodicities exist and that, apart from the possibility of a small number of absorbing states, (E_i is absorbing if $p_{ii} = 1$), no breakdown into noncommunicating subsets occur.

It is convenient to collect the p_{ij} into a matrix $P = \{p_{ij}\}$, so that

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0M} \\ p_{01} & p_{11} & \cdots & p_{1M} \\ \vdots & & & \\ p_{M0} & p_{M1} & \cdots & p_{MM} \end{pmatrix}. \tag{2.136}$$

The probability $p_{ij}^{(2)}$ that X is in E_j at time $t+2$, given it is in E_i at time t, is evidently

$$p_{ij}^{(2)} = \sum_k p_{ik} p_{kj}.$$

Since the right-hand side is the (i,j)th element in the matrix P^2, and if we write $P^{(2)} = \{p_{ij}^{(2)}\}$, then

$$P^{(t)} = P^t \tag{2.137}$$

for $t = 2$. More generally (2.137) is true for any positive integer t. In all cases we consider, P^t can be written in the spectral form

$$P^t = \lambda_0^t r_0 \ell_0' + \lambda_1^t r_1 \ell_1' + \cdots + \lambda_M^t r_M \ell_M' \tag{2.138}$$

where λ_0, $\lambda_1, \ldots, \lambda_M$ ($|\lambda_0| \geq |\lambda_1| \geq \cdots \geq |\lambda_M|$) are the eigenvalues of P and $(\ell_0, \ldots, \ell_M)$ and $(r_0, \ldots, r_M)$, normalized so that

$$\ell_i' \mathbf{r}_i = \sum_{j=0}^{M} \ell_{ij} r_{ij} = 1, \tag{2.139}$$

are the corresponding left and right eigenvectors, respectively. Suppose E_0 and E_M are absorbing states and that no other states are absorbing. Then $\lambda_0 = \lambda_1 = 1$ and if $|\lambda_2| > |\lambda_3|$ and $i, j = 1, 2, \ldots, M-1$,

$$p_{ij}^{(t)} = r_{2i} \ell_{2j} \lambda_2^t + o(\lambda_2^t) \tag{2.140}$$

for large t. Thus the leading nonunit eigenvalue λ_2 plays an important role in determining the rate at which absorption into either E_0 and E_M occurs.

Let π_j be the probability that eventually E_M (rather than E_0) is entered, given initially that X is in E_i. By considering values of X at consecutive time points it is seen that the π_i satisfy

$$\pi_i = \sum_{j=0}^{M} p_{ij} \pi_j, \quad \pi_0 = 0, \quad \pi_M = 1. \tag{2.141}$$

For the genetic model (1.48) (with $M = 2N$) the solution of (2.141) was $\pi_i = i/M$. The mean times $\bar{t}_i$ until absorption into E_0 or E_M occurs, given

that X is in E_i, similarly satisfy

$$\bar{t}_i = \sum_{j=0}^{M} p_{ij}\bar{t}_j + 1, \quad \bar{t}_0 = \bar{t}_M = 0. \tag{2.142}$$

Starting with X in E_i the members of the set of mean times $\{\bar{t}_{ij}\}$ that X is in E_j before absorption into either E_0 or E_M satisfy the equations

$$\bar{t}_{ij} = \sum_{k=0}^{M} p_{ik}\bar{t}_{kj} + \delta_{ij}, \quad \bar{t}_{0j} = \bar{t}_{Mj} = 0, \tag{2.143}$$

where $\delta_{ij} = 1$ and $i = j$ and $\delta_{ij} = 0$ otherwise. Further,

$$\bar{t}_{ij} = \sum_{n=0}^{\infty} p_{ij}^{(n)}, \quad \bar{t}_i = \sum_{j=1}^{M-1} \bar{t}_{ij}. \tag{2.144}$$

An expression can also be found for the variance σ_i^2 of the time before absorption, given initially X in E_i, namely

$$\sigma_i^2 = 2\sum_{j=1}^{M-1} \bar{t}_{ij}\,\bar{t}_j - \bar{t}_i - (\bar{t}_i)^2. \tag{2.145}$$

It is possible to derive the general form of the distribution of the time that X is in E_j if initially in E_i. Suppose that, starting in E_i, the probability that X ever enters E_j is α_{ij} and that once in E_j, the probability that X ever returns to E_j is r_j. Then the probability that E_j is occupied exactly n times before absorption takes place at E_0 or E_m is

$$\left.\begin{array}{ll} 1-\alpha_{ij} & \text{for } n = 0 \\ \alpha_{ij}(r_j)^{n-1}(1-r_j) & \text{for } n \geq 1. \end{array}\right\} \tag{2.146}$$

This is clearly a modified geometric distribution. The mean is thus

$$\begin{aligned} \bar{t}_{ij} &= \alpha_{ij}(1-r_j)\sum_{n=1}^{\infty} n r_j^{n-1} \\ &= \alpha_{ij}/(1-r_j) \end{aligned} \tag{2.147}$$

and the variance is

$$\begin{aligned} \sigma_{ij}^2 &= \alpha_{ij}(1-r_j)\sum_{n=1}^{\infty} n^2 (r_j)^{n-1} - \bar{t}_{ij}^2 \\ &= \bar{t}_{ij}\{1 - \bar{t}_{ij} + 2r_j/(1-r_j)\}. \end{aligned} \tag{2.148}$$

It is possible to find an expression for r_j and hence to calculate (2.148) but we do not enter into details here.

Consider now only those cases for which E_M is the absorbing state eventually entered. Writing X_t for the value of X at time t, we get

$$\begin{aligned} p^*_{ij} &= \text{Prob}\{X_{t+1} \text{ in } E_j \mid X_t \text{ in } E_i, E_M \text{ eventually entered}\} \\ &= \text{Prob}\{X_{t+1} \text{ in } E_j \text{ and } E_M \text{ eventually entered} \mid X_t \text{ in } E_i\} \\ &\quad \div \text{Prob}\{E_M \text{ eventually entered} \mid X_t \text{ in } E_i\} \\ &= p_{ij}\pi_j/\pi_i, \qquad (i,j = 1,2,\ldots,M), \end{aligned} \tag{2.149}$$

using conditional probability arguments and the Markovian nature of X. Let $\widetilde{P}$ be the matrix derived from P by omitting the first row and first column and let

$$V = \begin{pmatrix} \pi_1 & & & \\ & \pi_2 & & \\ & & \ddots & 0 \\ 0 & & & \pi_M \end{pmatrix}. \tag{2.150}$$

Then if $P^* = \{p^*_{ij}\}$, (2.149) shows that

$$P^* = V^{-1}\widetilde{P}V. \tag{2.151}$$

Standard theory shows that the eigenvalues of P^* are identical to those of P (with one unit eigenvalue omitted) and that if $\ell'(\mathbf{r})$ is any left (right) eigenvector of $\widetilde{P}$, then the corresponding left and right eigenvector of P^* are $\ell'V$ and $V^{-1}\mathbf{r}$. Further, if $P^{*(n)}$ is the matrix of conditional n step transition probabilities,

$$P^{*(n)} = (P^*)^n = V^{-1}\widetilde{P}^n V$$

so that

$$p^{*(n)}_{ij} = p^{(n)}_{ij}\pi_j/\pi_i, \tag{2.152}$$

a conclusion that can be reached directly as with (2.149). If $\bar{t}^*_{ij}$ is the conditional mean time spent in E_j, given initially X in E_i, then

$$\begin{aligned} \bar{t}^*_{ij} &= \sum_{n=0}^{\infty} p^{*(n)}_{ij} \\ &= (\pi_j/\pi_i)\sum_{n=0}^{\infty} p^{(n)}_{ij} \\ &= \bar{t}_{ij}\pi_j/\pi_i. \end{aligned} \tag{2.153}$$

If there is only one absorbing state interest centers solely on properties of the time until the state is entered. Taking E_0 as the only absorbing state and E_i as the initial state, the mean time t_i until absorption satisfies (2.142) with the single boundary condition $\bar{t}_0 = 0$, and the mean number of visits to E_j satisfies (2.143) with the single condition $\bar{t}_{0j} = 0$.

If there are no absorbing states P will have a single eigenvalue and all other eigenvalues will be strictly less than unity in absolute value. Equation (2.138) then shows that

$$\lim_{t\to\infty} P^t = \mathbf{r}_0 \ell_0' \tag{2.154}$$

and since r_0 is of the form $(1, 1, 1, \ldots, 1)'$,

$$\lim_{t\to\infty} p_{ij}^{(t)} = \ell_{0j} \quad \text{for all } i. \tag{2.155}$$

Using a slightly different notation we may summarize this by saying

$$\lim_{t\to\infty} p_{ij}^{(t)} = \phi_j, \tag{2.156}$$

where $\boldsymbol{\phi}' = (\phi_0, \phi_1, \ldots, \phi_M)$ is the unique solution of the two equations

$$\boldsymbol{\phi}' = \boldsymbol{\phi}' P, \quad \sum_{j=0}^{M} \phi_j = 1. \tag{2.157}$$

The vector $\boldsymbol{\alpha}$ is called the stationary distribution of the process and in genetical applications exists only if fixation of any allele is impossible (e.g. if all alleles mutate at positive rates).

If the matrix P is a continuant (so that $p_{ij} = 0$ if $|i - j| > 1$) explicit formulas can be found for most of these quantities. We write $p_{i,i+1} = \lambda_i$ and $p_{i,i-1} = \mu_i$ in conformity with standard notation in this case. If E_0 and E_M are both absorbing states the probability π_i in (2.141) becomes, explicitly,

$$\pi_i = \sum_{k=0}^{i-1} \rho_k \Big/ \sum_{k=0}^{M-1} \rho_k, \tag{2.158}$$

where

$$\rho_0 = 1, \quad \rho_k = \frac{\mu_1\mu_2\mu_3 \cdots \mu_k}{\lambda_1\lambda_2 \cdots \lambda_k}.$$

Further

$$\begin{aligned} \bar{t}_{ij} &= \frac{(1-\pi_i)\sum_{k=0}^{j-1}\rho_k}{\rho_{j-1}\mu_j}, \quad (j = 1, \ldots, i), \\ \bar{t}_{ij} &= \frac{\pi_i \sum_{k=j}^{M-1}\rho_k}{\rho_j \lambda_j}, \quad (j = 1 + 1, ; \ldots, M-1). \end{aligned} \tag{2.159}$$

Equations (2.144) and (2.153) then yield $\bar{t}_i$, $\bar{t}_{ij}^*$ and $\bar{t}_i^*$ immediately. When there is only one absorbing state (2.144) still holds, but now $\bar{t}_{ij}$ is defined

by

$$\bar{t}_{ij} = \begin{cases} \mu_j^{-1}\left\{1 + \dfrac{\lambda_{j-1}}{\mu_{j-1}} + \dfrac{\lambda_{j-1}\lambda_{j-2}}{\mu_{j-1}\mu_{j-2}} + \cdots + \dfrac{\lambda_{j-1}\lambda_{j-2}\cdots\lambda_1}{\mu_{j-1}\mu_{j-2}\cdots\mu_1}\right\} \\ \qquad (j = 1, 2, \ldots, i) \\ \bar{t}_{ii}\left(\dfrac{\lambda_i\lambda_{i+1}\cdots\lambda_{j-1}}{\mu_{i+1}\mu_{i+2}\cdots\mu_j}\right) \quad (j = i+1, \ldots, M) \end{cases} \tag{2.160}$$

if E_0 is the absorbing state and by

$$\bar{t}_{ij} = \begin{cases} \dfrac{\mu_{j+1}\mu_{j+2}\cdots\mu_i}{\lambda_j\lambda_{j+1}\cdots\lambda_{i-1}}\bar{t}_{ii} \quad (j = 0, 1, \ldots, i-1) \\ \lambda_i^{-1}\left\{1 + \dfrac{\mu_{i+1}}{\lambda_{i+1}} + \dfrac{\mu_{i+1}\mu_{i+1}}{\lambda_{i+1}\lambda_{i+2}} + \cdots + \dfrac{\mu_{i+1}\mu_{i+2}\cdots\mu_{M-2}}{\lambda_{i+1}\lambda_{i+2}\cdots\lambda_{M-1}}\right\} \\ \qquad (j = i, i+1, \ldots, M-1) \end{cases} \tag{2.161}$$

if E_M is the absorbing state. In this case of course there can be no further concept of a conditional mean absorption time.

Finally, when there are no absorbing states, the stationary distribution $\boldsymbol{\phi}$ is defined by

$$\phi_i = \phi_0 \frac{\lambda_0\lambda_1 \ \cdots \ \lambda_{i-1}}{\mu_1\mu_2 \ \cdots \ \mu_i}, \tag{2.162}$$

where ϕ_0 is chosen so that $\sum \alpha_i = 1$.

Various further results are possible for continuant Markov chain models, an accessible summary being given in Kemeny and Snell (1960). We shall draw on the formulas given above on a number of occasions throughout this book.

We conclude our discussion of finite Markov chains by introducing the concept of time reversibility. Consider a Markov chain admitting a stationary distribution $\{\phi_0, \phi_1, \ldots, \phi_M\}$. Then we define the process to be reversible if, at stationarity,

$$\text{Prob}\{X_t, X_{t+1}, \ldots, X_{t+n}\} = \text{Prob}\{X_t, X_{t-1}, \ldots, X_{t-n}\} \tag{2.163}$$

for every t and n. A necessary and sufficient condition for this is that the stationary state has been reached and that the equation

$$\phi_i p_{ij} = \phi_j p_{ji} \tag{2.164}$$

hold for all i, j. Certain classes of Markov chains are always reversible. For example, if the transition matrix is a continuant, (2.162) and (2.163) jointly show that the Markov chain at stationarity is reversible. Certain other chains, in particular several having genetical relevance, are reversible: we shall consider these later when discussing the uses to which the concept of reversibility can be put.

http://www.springer.com/978-0-387-20191-7

Mathematical Population Genetics 1
Theoretical Introduction
Ewens, W.J.
2004, XX, 418 p., Hardcover
ISBN: 978-0-387-20191-7