Lecture Notes in Biomathematics

Managing Editor: S. Levin

15

Thomas Nagylaki

Selection in One- and Two-Locus Systems

Springer-Verlag
Berlin · Heidelberg · New York 1977

Author

Prof. Dr. Thomas Nagylaki
Dept. of Biophysics and Theoretical Biology
The University of Chicago
920 East 58th Street
Chicago, Illinois 60637/USA

Library of Congress Cataloging in Publication Data

Nagylaki, Thomas, 1944-
Selection in one- and two-locus systems.

(Lecture notes in biomathematics ; 15)
Bibliography: p.
Includes index.
1. Population genetics--Mathematical models.
2. Natural selection--Mathematical models. I. Title.
II. Series.
QH455.N33 575.1 77-8295

AMS Subject Classifications (1970): 92-01, 92-02, 92A05, 92A10.

ISBN-13:978-3-540-08247-7 e-ISBN-13:978-3-642-93061-4
DOI:10.1007/978-3-642-93061-4

2145/3140-543210

PREFACE

Most of these notes were presented as part of a two-quarter course on theoretical population genetics at The University of Chicago. Almost all the students were either undergraduates in mathematics or graduate students in the biological sciences. The only prerequisites were calculus and matrices. As is done in these notes, biological background and additional mathematical techniques were covered when they were required. I have included the relevant problems assigned in the course.

My aim in these notes is to formulate the various models fairly generally, making the biological assumptions quite explicit, and to perform the analyses relatively rigorously. I hope the choice and treatment of topics will enable the reader to understand and evaluate detailed analyses of specific models and applications in the literature. No attempt has been made to review the literature or to assign credit. Most of the references are to papers directly germane to the subjects and approaches covered here. Frequency of reference is not intended to reflect proportionate contribution.

I am very grateful to Professor James F. Crow for helpful comments and to Mrs. Adelaide Jaffe for her excellent typing. I thank the National Science Foundation for its support (Grant No. DEB76-01550).

June 1976 Thomas Nagylaki

CONTENTS

1. INTRODUCTION

Population genetics concerns the genetic structure and evolution of natural populations. The genetic composition of a population is usually described by genotypic proportions, which may depend on space and time. These genotypic frequencies are determined by a few elementary genetic principles and the following evolutionary factors.

Various genotypes may have different probabilities of surviving to adulthood and may reproduce at different rates. Differential mortality and fertility are the components of *selection*. Unless the population is in equilibrium, selection will change the genotypic and allelic frequencies in accordance with the expected number of progeny, called fitness, of the various genotypes. Natural selection has been recognized since Darwin as the directive force of adaptive evolution.

The action of selection is strongly affected by the *mating system*. If mating occurs without regard to the genotypes under consideration, we say it is random. This is the simplest situation and, at least approximately, appears to be frequently realized in nature. We say there is inbreeding if related individuals are more likely to mate than randomly chosen ones. Assortative mating refers to the tendency of individuals resembling each other with respect to the trait in question to mate with each other. Disassortative mating means that phenotypically dissimilar individuals mate more often than randomly chosen ones. Nonrandom mating influences genotypic frequencies. In the absence of selection, inbreeding does not change gene frequencies, but assortative and disassortative mating may. This will happen if the mating pattern is such that some genotypes have a higher probability of mating than others.

Mutation designates the change from one allelic form to another. Clearly, it directly alters gene frequencies.

In spatially structured populations, *migration* must be taken into account. It can affect not only the geographical composition of the population, but the amount of genetic variability as well.

Unless some of the parameters, such as the selection intensities, required to specify the elements of evolution described above fluctuate at random, the evolutionary forces will be deterministic. In a finite population, however, allelic frequencies will vary

probabilistically due to the random sampling of genes from one generation to the next. This process is called *random genetic drift*. Its causes are (nonselective) random variation in the number of offspring of different individuals and the stochastic nature of Mendel's Law of Segregation. Evidently, the smaller the population, the larger is the evolutionary rôle of random drift. No matter how large the population is, however, the fate of rare genes still depends strongly on random sampling.

All three current views of evolution identify mutation as the source of raw material for evolutionary change, but differ in the emphasis placed on the other factors.

The main underpinning of Fisher's theory is his Fundamental Theorem of Natural Selection--that the rate of change of the mean fitness of a population is equal to the additive component of its genetic variance in fitness (Fisher, 1930). Fisher held that evolution occurs primarily by the deterministic increase in fitness of large populations under the action of natural selection. We may imagine the mean fitness as a surface in the space of suitable dynamical variables like gene frequencies, and consequently envisage that the population is climbing a hill on this surface. In Fisher's picture, random drift is responsible only for small chance fluctuations in the trajectory of the population. These notes are essentially an exposition of the theory underlying Fisher's view at the level of one and two loci.

Wright (1931, 1970) stressed that, due to multiple effects (pleiotropy) and interactions (epistasis) of loci and selection for an intermediate optimum, the fitness surface has many selective peaks. Small populations can "test" this surface by random drift, sometimes crossing a saddle from a lower selective hill to a higher one. He pointed out that if a species is divided into many such small populations, which exchange relatively few migrants, dispersion of selectively favored individuals may enable it to reach the highest peak on the surface. A comprehensive analytical or numerical treatment of this complex theory awaits future research.

The neutral theory of Kimura (1968) and of King and Jukes (1969) ascribes much of evolution, especially at the molecular level, to mutation and random drift. In its strongest form, this theory attributes a lot of the variation even in morphological characters to these two forces (Nei, 1975, pp.251-253). This is still not inconsistent with the fact that natural selection determines the nature of adaptation. The weakest form of the theory holds only that most amino acid substi-

tutions are neutral. The neutral theory is mathematically quite highly developed.

It may be helpful to list with suitable general references the major aspects of theoretical population genetics not covered in these notes. Inbreeding and quantitative genetics do not require advanced mathematics and are discussed by Crow and Kimura (1970). The other topics require considerable use of probability, analysis, and differential equations. For random fluctuations in selection intensities, the reader may consult Karlin and Lieberman (1974) and Karlin and Levikson (1974). Geographical variation is discussed in Nagylaki (1977). Various aspects of random drift are treated by Moran (1962), Ewens (1969), Crow and Kimura (1970), and Kimura and Ohta (1971). Felsenstein and Taylor (1974) have compiled a bibliography of theoretical population genetics with almost complete coverage through the autumn of 1973.

The next chapter concerns selection in asexual haploid populations. Its purpose is to formulate and analyze in a simple setting many of the problems dealt with in these notes. Having obtained this perspective, the reader will be better prepared for the treatment of selection with the additional complexity of mating and Mendelian segregation and recombination in the remainder of the notes. Chapter 3 examines the structure of a randomly mating population in the absence of selection. The reader who already has a thorough understanding of Mendelism (including sex-linkage and recombination), which will be assumed in the subsequent chapters, may omit this chapter. The basic theory of selection at an autosomal locus is developed in Chapter 4. A few less common types of selection and modes of inheritance are discussed chiefly to provide more practice in formulating models. The remaining chapters are logically independent of each other.

We shall now give a minimum of very much simplified genetic background. More information will be introduced when it is required. Cavalli-Sforza and Bodmer (1971) present a good summary of the pertinent genetic background. Crow's (1976) lucid book is more detailed, but still quite concise.

The genetic material, deoxyribonucleic acid (DNA), consists of four bases, adenine (A), guanine (G), thymine (T), and cytosine (C), each of which is linked to a sugar and a phosphate, forming a *nucleotide*. These nucleotides are arranged in a double helix, the only possible pairings being A-T and G-C. Therefore, the information is in the sequence along a single helix. Three bases code for an amino acid, of which there are 20. A protein consists of at least several hundred

amino acids. Roughly, the region of DNA determining a protein is a *locus* or *gene*. A particular sequence there is an *allele*. In population genetics, however, "gene" is sometimes used in the sense of "allele", as defined above (e.g., in "gene frequency"). A *chromosome* is composed of proteins and a single thread-like molecule of DNA, along which genes are arranged linearly. The number of different chromosomes is characteristic of each species. If the chromosomes in a set are single, the organism is called *haploid*; if they are doubled, the organism is *diploid*. Generally, bacteria, algae, mosses, and fungi are examples of haploid organisms, while higher plants and animals are diploid. We shall not consider polyploids, in which chromosomes are at least tripled. Very crudely, asexual haploids just duplicate themselves. At *meiosis*, in diploids the chromosomes separate, and each *gamete* (sperm or egg) carries a single set of chromosomes. The sets from a sperm and egg unite at *fertilization* to form a diploid *zygote*, from which the individual develops. In plants, pollen fertilize ova to form seeds.

2. ASEXUAL HAPLOID POPULATIONS

The study of haploid populations in this chapter will enable us to formulate and analyze many of the problems which concern us without the additional complication of mating and Mendelian segregation and recombination. If alleles are interpreted as genotypes, and mutation rates refer to zygotes rather than alleles, the formalism applies, regardless of ploidy, also to asexual species like the dandelion. As explained in Chapter 3, it applies also to a Y-linked locus in sexually-reproducing diploids. We shall expound the basic selection model with discrete nonoverlapping generations in Section 2.1, include mutation and migration in Sections 2.2 and 2.3, and treat overlapping generations with continuous time in Section 2.4.

2.1 *Selection*

We consider a single locus with alleles A_i, $i = 1, 2, \ldots, k$, and assume generations are discrete and nonoverlapping. Thus, the adults are replaced by their offspring in each generation. Although this assumption will hold for some laboratory populations, for natural populations of haploids it should be viewed as a simple approximation. Let the number of offspring carrying A_i in generation t, where $t = 0, 1, 2 \ldots$, be $n_i(t)$. The total number of offspring,

$$N = \sum_i n_i, \tag{2.1}$$

must be sufficiently large to allow us to neglect random drift.

Let v_i designate the probability that an A_i offspring survives to reproductive age. The average number of progeny of an A_i adult is f_i. The *viabilities* v_i and *fertilities* f_i may be functions of the time t and the vector of population numbers, denoted by $\underset{\sim}{n}(t)$. The product $w_i = v_i f_i$ represents the *fitness* of an A_i individual. The $v_i n_i$ A_i adults in generation t contribute $f_i(v_i n_i) = w_i n_i$ A_i offspring to generation $t+1$. Therefore, the basic recursion relations

$$n_i(t+1) = w_i[t, \underset{\sim}{n}(t)]n_i(t) \tag{2.2}$$

depend only on the fitnesses, and not on the viabilities and fertilities

separately. Since w_i is the expected number of progeny of an A_i juvenile, this is not surprising. The fundamental difference equations (2.2) determine $\underset{\sim}{n}(t)$ iteratively in terms of $\underset{\sim}{n}(0)$, provided the fitnesses $w_i(t,\underset{\sim}{n})$ are specified.

If A_i is lethal, $v_i = 0$, or causes sterility, $f_i = 0$, then $w_i = 0$. Otherwise, the w_i will usually not differ from each other by more than a few percent. If the population size is approximately constant, the average of the w_i will be close to unity. The numbers w_i-1 are called *selection coefficients*.

The proportion or *frequency* of the allele A_i among offspring is

$$p_i = \frac{n_i}{N}. \tag{2.3}$$

Unless stated otherwise, a prime will always signify the next generation. Then (2.1), (2.2), and (2.3) yield

$$N' = \sum_i n_i' = \sum_i w_i\left(\frac{n_i}{N}\right)N = \bar{w}N, \tag{2.4}$$

where

$$\bar{w} = \sum_i w_i p_i \tag{2.5}$$

is the *mean fitness* of the population and gives its rate of growth. The mean fitness is of great conceptual and analytical importance in the theory of selection.

The gene frequencies satisfy the recursion relations

$$p_i' = \frac{n_i'}{N'} = \frac{w_i n_i}{\bar{w}N} = p_i\left(\frac{w_i}{\bar{w}}\right). \tag{2.6}$$

We see at once from (2.5) and (2.6) that the relation

$$\sum_i p_i = 1 \tag{2.7}$$

holds for all time if initially true, as it must be. It is also apparent from (2.5) and (2.6) that the gene frequencies depend only on ratios of the fitnesses. All the w_i may be multipled by the same constant without altering the evolution of the allelic frequencies p_i. Exploiting this scale invariance often simplifies the algebra. We shall always employ a Δ to indicate the change in one generation. Thus, (2.6) gives

$$\Delta p_i = p_i'-p_i = \frac{p_i(w_i-\bar{w})}{\bar{w}}. \tag{2.8}$$

If the fitnesses are functions only of time, it is useful to iterate (2.2):

$$n_i(t) = n_i(0) \prod_{\tau=0}^{t-1} w_i(\tau). \tag{2.9}$$

Then the gene frequencies read

$$p_i(t) = \frac{p_i(0) \prod_{\tau=0}^{t-1} w_i(\tau)}{\sum_j p_j(0) \prod_{\tau=0}^{t-1} w_j(\tau)}. \tag{2.10}$$

It is often assumed that the fitnesses are constant, meaning, really, that they vary much more slowly than the other pertinent evolutionary parameters. In that case, (2.9) and (2.10) reduce to

$$n_i(t) = n_i(0) w_i^t, \tag{2.11}$$

$$p_i(t) = \frac{p_i(0) w_i^t}{\sum_j p_j(0) w_j^t}. \tag{2.12}$$

Suppose A_1 is the fittest allele: $w_1 > w_i$, $i > 1$. Equation (2.12) informs us immediately that $p_1(t) \to 1$ as $t \to \infty$. Since the population number can be changed each generation without changing gene frequencies, this means that the population size will remain finite, and the fittest gene will be ultimately *fixed*, the others being *lost*. Of course, all statements of this sort presuppose the allele under consideration is initially present in the population. For instance, here we posit $p_1(0) > 0$.

It will be helpful to distinguish three levels of description of the evolution of a population. We shall refer to the specification of the variables of interest as functions of time as a *complete solution*. Equation (2.12) is an example. Often, even though one cannot obtain a complete solution, one can determine the fate of the population for all initial conditions. We may call this a *complete* or *global analysis*. The statement that with constant fitnesses the fittest allele is fixed, falls in this category. If we cannot carry out a complete analysis, we may still obtain some information of evolutionary interest by locating all the equilibria of the system and investigating its behavior in the neighborhood of these stationary states. In the problem

treated above, a part of such a *local analysis* would be to observe that $p_1 = 1$ is an equilibrium, and to show that if $p_1(0)$ is sufficiently close to 1, then $p_1(t) \to 1$ as $t \to \infty$.

We characterize the local stability of equilibria as follows. If, in terms of a suitable metric for the variables of the problem, the population will remain within an arbitrarily small preassigned distance of the equilibrium, provided it starts sufficiently close to the equilibrium, we say the equilibrium is *stable*. Otherwise, it is *unstable*. An equilibrium is *asymptotically stable* if it is stable, and a population starting sufficiently close to the equilibrium converges to it. The gene frequency equilibrium $p_1 = 1$ is globally asymptotically stable, while the equilibrium $p_2 = 1$ is unstable. If $w_i = 1$ for all i, every point is a stable equilibrium, but the stability is not asymptotic. The word "asymptotic" is frequently omitted in population genetics.

We proceed now to study the change in mean fitness. From (2.5) we have

$$\begin{aligned} \Delta\bar{w} &= \sum_i (p_i' w_i' - p_i w_i) \\ &= \sum_i [p'(\Delta w_i + w_i) - p_i w_i] \\ &= \overline{\Delta w} + \sum_i w_i \Delta p_i, \end{aligned} \tag{2.13}$$

where

$$\overline{\Delta w} = \sum_i p_i' \Delta w_i \tag{2.14}$$

is the mean of the fitness changes over the next generation. Substituting (2.8) into (2.13), we find

$$\Delta\bar{w} = \overline{\Delta w} + \bar{w}^{-1} \sum_i p_i w_i (w_i - \bar{w}). \tag{2.15}$$

But (2.5) informs us that

$$\sum_i p_i (w_i - \bar{w}) = 0,$$

so we may subtract $\bar{w}$ from the first w_i in (2.15) to obtain

$$\Delta\bar{w} = \overline{\Delta w} + \bar{w}^{-1} V, \tag{2.16}$$

where

$$V = \sum_i p_i(w_i-\bar{w})^2 \tag{2.17}$$

is the genic variance in fitness. The simple steps in this derivation are often useful.

Equation (2.16) is a simple case of Fisher's Fundamental Theorem of Natural Selection (Fisher, 1930). With constant fitnesses, $\overline{\Delta w} = 0$, so $\Delta\bar{w} = \bar{w}^{-1}V \geq 0$, i.e., the mean fitness is nondecreasing. Since $V = 0$ if and only if $p_i = 0$ or $w_i = \bar{w}$ for all i, (2.6) implies $\Delta\bar{w} = 0$ only at equilibrium. Thus, selection increases the mean fitness, using up the genic variance. If the selection coefficients are small, we may choose all the fitnesses to be close to 1. Then $\bar{w}$ will be approximately unity, and the rate of change of the mean fitness will be roughly equal to the genic variance.

With only two alleles, it is customary to put $p = p_1$ and $q = p_2 = 1-p$. From (2.8) we deduce

$$\Delta p = pq(w_1-w_2)/\bar{w}, \tag{2.18}$$

and the variance in fitness reduces to

$$V = pq(w_1-w_2)^2. \tag{2.19}$$

2.2 *Mutation and Selection*

Let us consider first mutation without selection. We designate the probability that an A_i individual has an A_j offspring for $i \neq j$ by the mutation rate u_{ij}. It will be convenient to use the convention $u_{ii} = 0$ for all i. Generally, mutation rates are quite small; 10^{-6} is a representative value. Mutation rates at the nucleotide level are on the order of 10^{-10}. The gene frequency change in one generation clearly reads

$$\Delta p_i = \sum_j p_j u_{ji} - p_i \sum_j u_{ij}. \tag{2.20}$$

By interchanging dummy variables in one of the sums, often a useful device, we observe directly that

$$\sum_i \Delta p_i = 0,$$

so that (2.7) is preserved. Since the total mutation rate must not

exceed unity, therefore $\Delta p_i \geq -p_i$, whence $p_i' \geq 0$, as required.

For two alleles, one commonly writes $u = u_{12}$ and $v = u_{21}$. With $p = p_1$, as above, (2.20) becomes

$$\Delta p = v-(u+v)p. \tag{2.21}$$

At equilibrium, $\Delta p = 0$, so the frequency of A_1 is

$$\hat{p} = \frac{v}{u+v}. \tag{2.22}$$

We shall follow convention and indicate equilibrium values by a caret. Equilibria like (2.22), with more than one allele present, are called *polymorphic*. As expected, if mutation is irreversible, i.e., $u = 0$ or $v = 0$, the allele whose frequency is decreasing is absent at equilibrium.

It is frequently convenient to study the deviation from equilibrium. Substituting $x = p-\hat{p}$ into (2.21), we find

$$x' = (1-u-v)x,$$

with the solution

$$x(t) = x(0)(1-u-v)^t. \tag{2.23}$$

Therefore, there is global convergence to (2.22) at the geometric rate $1-u-v$. Note that the time required for significant gene frequency change,

$$t = \frac{\ln[x(t)/x(0)]}{\ln(1-u-v)} \approx \frac{\ln[x(0)/x(t)]}{u+v},$$

is very long, typically about 10^6 generations.

It is often useful to approximate powers like the one in (2.23) by exponentials. For $|\varepsilon| << 1$ and $\varepsilon^2 t << 1$, $(1-\varepsilon)^t = \exp[t \ln(1-\varepsilon)] = \exp[-t(\varepsilon+\frac{1}{2}\varepsilon^2+\ldots)] \approx e^{-\varepsilon t}$. Thus, we may rewrite (2.23) as

$$x(t) \approx x(0)e^{-(u+v)t}. \tag{2.24}$$

The fact that (2.24) becomes inaccurate as $(u+v)^2 t$ approaches 1 is irrelevant because by that time $x(t)$ is extremely close to 0.

To include selection, we set up the formal scheme

$$\underset{p_i}{\text{Offspring}} \xrightarrow[\text{viability}]{} \underset{}{\text{Adult}} \xrightarrow[\text{fertility}]{} \underset{p_i^*}{\text{Offspring}} \xrightarrow[\text{mutation}]{} \underset{p_i'}{\text{Offspring}}$$

with the indicated gene frequencies. Let R_{ij} be the probability that a gamete from an A_i offspring carries A_j. Recalling (2.6), we have

$$p_i^* = p_i\left(\frac{w_i}{\bar{w}}\right), \tag{2.25a}$$

$$p_i' = \sum_j p_j^* R_{ji}, \tag{2.25b}$$

where $\bar{w}$ is still given by (2.5). It is important to note that (2.25) correctly describes the biological situation that, while selection acts on the phenotype, which develops from the offspring genotype, the germ cells mutate with no phenotypic effect at rates u_{ij}, related to R_{ij} by

$$R_{ij} = \delta_{ij}\left(1 - \sum_k u_{ik}\right) + u_{ij}. \tag{2.26}$$

The Kronecker delta, δ_{ij}, is defined by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

From (2.25b) and (2.26) we derive

$$\Delta p_i = p_i' - p_i = p_i^* - p_i + \sum_j p_j^* u_{ji} - p_i^* \sum_j u_{ij}.$$

If selection is weak, since $u_{ij} \ll 1$, we may neglect $(p_i^* - p_i)u_{ij}$ for all i and j to obtain

$$\Delta p_i \approx \Delta p_i(\text{selection}) + \Delta p_i(\text{mutation}),$$

where

$$\Delta p_i(\text{selection}) = p_i^* - p_i,$$

$$\Delta p_i(\text{mutation}) = \sum_j p_j u_{ji} - p_i \sum_j u_{ij}.$$

Let us analyze the diallelic case. Choosing, without loss of generality, $w_1 = 1$, $w_2 = 1+s$, $s > 0$, in the notation introduced above, (2.25) reduces to the *linear fractional transformation*

$$p' = \frac{\alpha+\beta p}{\gamma+\delta p} \tag{2.27}$$

with $\alpha = v(1+s)$, $\beta = 1-sv-u-v$, $\gamma = 1+s$, $\delta = -s$. Since (2.27) occurs in several models, we shall discuss it for arbitrary values of its parameters. The two solutions of $p' = p$ are

$$p_{\pm} = (2\delta)^{-1}(\beta-\gamma\pm Q^{1/2}), \tag{2.28}$$

where

$$Q = (\beta-\gamma)^2 + 4\alpha\delta. \tag{2.29}$$

The trivial case $\delta = 0$ corresponds to (2.21), so we suppose $\delta \neq 0$. We also assume $\alpha\delta \neq \beta\gamma$, for otherwise (2.27) shows that $p' = \alpha/\gamma$.

If $Q \neq 0$, then $p_+ \neq p_-$, hence

$$y = \frac{p-p_+}{p-p_-} \tag{2.30}$$

satisfies $y' = \lambda y$, where

$$\lambda = \frac{\beta+\gamma-Q^{1/2}}{\beta+\gamma+Q^{1/2}}. \tag{2.31}$$

Therefore,

$$y(t) = y(0)\lambda^t, \tag{2.32}$$

and (2.27) has the solution

$$p(t) = \frac{p_-y(t)-p_+}{y(t)-1}. \tag{2.33}$$

Of course, $y(0)$ is evaluated from (2.30). There are two cases.

1. $Q < 0$: Here p_+ and p_- are complex. Since $|\lambda| = 1$, (2.32) has the form

$$y(t) = y(0)e^{-i\theta t}, \tag{2.34}$$

where

$$\theta = 2\tan^{-1}\left[\frac{(-Q)^{1/2}}{\beta+\gamma}\right]. \tag{2.35}$$

2. $Q > 0$: Now p_+ and p_- are real, and there are three subcases.

   (a) $\beta+\gamma > 0$: From (2.31) we see that $|\lambda| < 1$, whence $y(t) \to 0$ as $t \to \infty$. Therefore, (2.33) implies that $p(t) \to p_+$ as $t \to \infty$.

   (b) $\beta+\gamma < 0$: Since $|\lambda| > 1$, therefore, $y(t) \to \infty$, and consequently $p(t) \to p_-$ as $t \to \infty$.

   (c) $\beta+\gamma = 0$: Here $\lambda = -1$, and hence $y(t) = y(0)(-1)^t$. Therefore, $p(t)$ alternates between $p(0)$ and $p(1)$. This also holds for $Q < 0$, for then (2.35) yields $\theta = \pi$.

It remains to analyze the case with $Q = 0$.

3. $Q = 0$: Equation (2.28) tells us that there is a single equilibrium

$$\hat{p} = (2\delta)^{-1}(\beta-\gamma). \tag{2.36}$$

We find that

$$z = (p-\hat{p})^{-1} \tag{2.37}$$

satisfies, for $\beta+\gamma \neq 0$,

$$z' = z+\frac{2\delta}{\beta+\gamma}, \tag{2.38}$$

with the obvious solution

$$z(t) = z(0)+\frac{2\delta t}{\beta+\gamma}. \tag{2.39}$$

Equations (2.37) and (2.39) inform us that

$$p(t) \sim \hat{p}+\frac{\beta+\gamma}{2\delta t} \tag{2.40}$$

as $t \to \infty$. Thus, the ultimate rate of convergence to $\hat{p}$ is algebraic. If $\beta+\gamma = 0$, (2.29) tells us that $\alpha\delta = \beta\gamma$, so that $p' = \alpha/\gamma$.

In genetic problems, we shall be concerned with the mapping of some interval $I:[a,b]$ into itself. Then we can restrict the possible equilibrium structures for any continuous map $p' = f(p)$. Suppose a and b are not equilibria: $f(a) \neq a$ and $f(b) \neq b$. Hence, $f(a) > a$ and $f(b) < b$, so $g(p) = f(p)-p$ must change sign an odd number of times in I. Since f maps I into itself, therefore $g(p)$ is finite for p in I, and hence, counting multiplicity, has an odd number of zeroes in I. Thus, still counting multiplicity, f has an odd number of fixed points in I. For (2.27), provided $p_\pm \neq 0,1$, this means cases 1 and 3 may be excluded, and in case 2 either p_+ or p_- is in I, but not both.

Let us apply the theory just developed to our mutation-selection problem. It is easy to see that $Q \geq 0$, equality holding if and only if $u = 0$ and $s = v(1+s)$. Since usually $s >> v$, we assume $Q > 0$. Furthermore, it is trivial to verify that $\beta+\gamma \geq 0$, with equality only in the unbiological situation $u = v = 1$. Therefore, $p(t)$ converges to p_+ globally. It follows that $0 \leq p_+ \leq 1$. From the formula

$$p_\pm = (2s)^{-1}\{s+sv+u+v\mp[(s+sv+u+v)^2-4sv(1+s)]^{1/2}\}, \tag{2.41}$$

with a bit of algebra one can show that $p_- > 1$. With the biologically trivial assumption $u+v \leq 1$, one can prove from (2.31) that $\lambda \geq 0$, so

that we conclude from (2.30) that the convergence is without oscillation. If $v = 0$, both selection and mutation decrease p, so we expect $p_+ = 0$. Indeed, (2.41) yields $p_+ = 0$ and $p_- = 1+us^{-1}$. With $u = 0$, (2.41) reduces to $p_+ = v(1+s)/s$, $p_- = 1$. In the biologically important case $u,v << 1,s$, linearizing (2.41) in u and v yields $p_+ \approx v(1+s)/s$ and $p_- \approx 1+us^{-1}$. As expected, the equilibrium frequency, $p_+ << 1$.

2.3 *Migration and Selection*

We assume a proportion m of the population is replaced each generation by migrants with fixed gene frequencies $\bar{p}_i$. More complicated migration-selection schemes than this island model will be discussed for diploids in Chapter 6. To write our recursion relations, we replace mutation by migration in the formal mutation-selection scheme of Section 2.2. Then (2.25b), which we may rewrite as

$$p_i' = p_i^* + \sum_j p_j^* u_{ji} - p_i^* \sum_j u_{ji}, \tag{2.42}$$

becomes

$$p_i' = p_i^* + m(\bar{p}_i - p_i^*). \tag{2.43}$$

But the substitution $u_{ij} = m\bar{p}_j$, $i \neq j$, reduces (2.42) to (2.43), showing that migration is a special case of mutation.

2.4 *Continuous Model with Overlapping Generations*

Our formulation will be based on that of Cornette (1975) for diploids. Time, measured in arbitrary units, flows continuously. Let $\nu_i(t,x)\Delta x$ be the number of A_i individuals between the ages of x and $x+\Delta x$ at time t. The total number of A_i individuals at time t is

$$n_i(t) = \int_0^\infty \nu_i(t,x)dx. \tag{2.44}$$

If no individual survives beyond age X, then $\nu_i(t,x) = 0$ for $x > X$. The total population size is

$$N(t) = \sum_i n_i(t). \tag{2.45}$$

We set up equations for our fundamental variables, $\nu_i(t,x)$, with the aid of the life table, $l_i(t,x)$. We start observing the population