Mathematical Population Genetics

Lecture Notes[1] (Part 1), Winter Semester 2020

Reinhard Bürger

University of Vienna, Department of Mathematics

Contents

[1] November 5, 2020

1 Introduction

Population genetics is concerned with the study of the genetic composition of populations. This composition may be changed by segregation, selection, mutation, recombination, mating structure, migration, and other genetic, ecological, and evolutionary factors. Therefore, in population genetics these mechanisms and their interactions and evolutionary consequences are investigated. Traditionally, population genetics has been applied to animal and plant breeding, to human genetics, and more recently to ecology and conservation biology. It also has important interfaces with molecular biology, systematics, natural history, mathematics, statistics, and computing. One of the main subjects is the investigation of the mechanisms that generate and maintain genetic variability in populations, and the study of how this genetic variation, shaped by environmental influences, leads to evolutionary change, adaptation, and speciation. Therefore, research in population genetics relies on empirical observations, on experiments, and on theoretical considerations. In particular, population genetics provides the basis for understanding the evolutionary processes that have led to the diversity of life we encounter and admire.

Since so many factors interact and determine the evolutionary fate of a population, a proper understanding of the relevant processes requires a good deal of abstraction in planning experiments and in devising mathematical models. A good mathematical model, as well as a good experiment, takes into account the relevant biological mechanisms for studying a particular phenomenon and disregards the less relevant ones. As in other sciences, good model building must rest on an adequate knowledge of the basic biological reality and requires a clear formulation of the underlying hypotheses. The process of abstraction that is involved entails generality which, sometimes, may appear to be unnecessary. However, general methods or models, devised to study a particular phenomenon, may reveal the essence and the underlying structure more clearly and can often be applied to questions not anticipated before. Examples include the study of gene families and microsatellites, where classical population genetic modeling has been successfully applied.

Mathematical models and methods have a long history in population genetics, tracing back to Gregor Mendel, who used elementary mathematics to calculate the expected frequencies of the genes in his experiments. Francis Galton and the biometricians, notably Karl Pearson, developed new statistical methods to describe the distribution of trait values in populations and to predict their change between generations. The foundations of modern population genetics were laid by the work of Ronald A. Fisher, J.B.S. Haldane, and Sewall Wright, who reconciled Mendelism with Darwinism during the second and third decades of the twentieth century. They demonstrated that the theory of evolution by natural selection, proposed by Charles Darwin (1859), can be justified on the basis

of genetics as governed by Mendel's laws. The work of Fisher, Haldane, and Wright was highly mathematical for the biology of that time and was properly understood by only a small number of people. Nevertheless, their influence was enormous and they set the standards for mathematical modeling and for rigor of theoretical investigations for the subsequent decades.

Prior to 1900, the year when Mendel's work was rediscovered and then rapidly accepted, the hereditary mechanisms were unknown. Darwin believed in blending inheritance, according to which the hereditary material itself blended. However, as already noted by Darwin, blending inheritance produces uniformity and destroys variation that is so ubiquitous. In modern terms, heritable variance would be halved in each generation of random mating with blending inheritance (Fisher 1930). Therefore, one half of the heritable variance maintained in a population would have to arise anew in each generation. There were controversial lines of thought about the nature of this huge amount of new variation and its consequences for evolution. The 'gradualists', to which Darwin and the biometricians adhered, considered the changes across generations as gradual and incremental, whereas the 'saltationists' (e.g., T.H. Huxley and Galton) held that evolutionary changes occurred in 'jumps' of considerable magnitude. Much of the scientific dispute about Darwin's theory of evolution originated from the ignorance of the true hereditary mechanisms.

Despite the early work of Yule (1902), Hardy (1908), and Weinberg (1908), who showed that under the particulate mode of inheritance proposed by Mendel (1866), genetic variability is preserved under random mating, it was not before 1918 that the synthesis between genetics and the theory of evolution through natural selection began to take shape through Fisher's (1918) work (see Provine (1971) for a detailed account of the history of population genetics).

Today, the hereditary mechanisms have been firmly established and our knowledge about the molecular biology of the genes is rapidly increasing. Mutations are known to be the ultimate source of genetic variability, and many different processes at the chromosomal and molecular level have been identified that generate mutations. On the phenotypic level, the role of selection in shaping evolutionary change has been amply documented, whereas on the molecular level, a significant amount of neutral evolution appears to take place, its extent still being disputed. Nevertheless, there remain many open problems, some of which are qualitative in nature and some quantitative. Questions concerning the processes involved in speciation events or in the evolution of sex belong to the first class, whereas questions concerning the prediction of the expected evolutionary change of a population subject to selection belong to the second class. Such predictions are highly nontrivial, unless confined to one or a few generations, because there exist many different forms of selection and the response to selection depends on the pattern and amount of genetic

variability in the population. This variation, however, is a function of many genetic details (such as number of genes determining a trait, mutational properties, degree of linkage), of the demography (population size, mating structure), and of the selective forces acting. Therefore, the genetic variability may change from one generation to the next.

Mendel's (1866) prime achievement was the recognition of the particulate nature of the hereditary determinants, now called genes. A gene may have different forms, called *alleles*. From his experiments with peas he concluded that genes are present in pairs, one member of each pair having been inherited from the maternal parent, the other from the paternal. The allelic composition is called the *genotype*, and the set of observable properties derived from the genotype is called the *phenotype*. Thus, supposing that there are two alleles $\mathcal{A}_1$ and $\mathcal{A}_2$, there are three possible genotypes, $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, and $\mathcal{A}_2\mathcal{A}_2$. In the first and third case, the organism's genotype is *homozygous* (for $\mathcal{A}_1$ or $\mathcal{A}_2$, respectively), in the second case it is *heterozygous*. In general, the genotypes $\mathcal{A}_1\mathcal{A}_2$ and $\mathcal{A}_2\mathcal{A}_1$ cannot be distinguished. When the phenotype of the heterozygote $\mathcal{A}_1\mathcal{A}_2$ is the same as one of the homozygotes, say $\mathcal{A}_1\mathcal{A}_1$, allele $\mathcal{A}_1$ is called *dominant* and $\mathcal{A}_2$ is called *recessive*.

Mendel's first law states that when pure-bred (homozygous) strains are crossed, the hybrid progeny constituting the F_1 generation (the letter F stands for filial) are uniform (their genotype being $\mathcal{A}_1\mathcal{A}_2$ if the parents were $\mathcal{A}_1\mathcal{A}_1$ and $\mathcal{A}_2\mathcal{A}_2$) and usually express one of the two phenotypes (the one controlled by the dominant allele). According to Mendel's second law, recessive characters, which are masked in the heterozygous F_1, reappear in the F_2 in the proportion $1:3$ of the dominant character. This leads to the *Principle of Segregation*, stating that each reproductive cell (*gamete*) contains only one of the two alleles and that each gamete is equally likely to contain either one. The separation of the paired alleles from one another and their distribution to different cells, the gametes, is called *segregation* and occurs during *meiosis*. Meiosis is the process of formation of gametes from somatic cells. At mating, two reproductive cells fuse and form a *zygote* (fertilized egg), which contains the full (diploid) genetic information.

Mendel also performed experiments with pure-bred lines that differed in two characters, round versus wrinkled seed shape and yellow versus green color. From previous experiments he knew that 'round' was dominant over 'wrinkled' and 'yellow' dominant over 'green' because their F_2 ratios were $3:1$ each. The F_1 seeds, from crosses of lines having round and yellow seeds with lines having wrinkled and green seeds, were all round and yellow. In F_2 progeny from the dihybrid cross, all four phenotypes reappeared, approximately in the proportions 9/16 'round yellow', 3/16 'wrinkled yellow', 3/16 'round green', and 1/16 'wrinkled green'. Thus, the proportion of the four phenotypes is $9:3:3:1$, as expected when two pairs of alleles segregate independently, so that the $3:1$ ratios are combined at random. This is called Mendel's third law or the *Principle of Independent*

Assortment.

Since the 1940s it has been known that the genetic material is *deoxyribonucleic acid* (*DNA*). It consists of four bases: adenine (A), guanine (G), thymine (T), and cytosine (C). Each base is linked to a sugar and a phosphate group, yielding a *nucleotide*. The nucleotides are arranged along two chains to form a double-stranded helix in which the pairings A–T and G–C between the strands are formed. Therefore, all the genetic information is contained in each of the two strands. Three bases code for one amino acid, which are the building blocks of polypeptide chains and proteins. A gene typically represents a contiguous region of DNA coding for one polypeptide chain. Its position along the DNA is called the *locus*, and a particular sequence there is called an allele. Thus, two genes at the same locus, sampled from a population, may or may not be of the same allelic type. A double-stranded helix of DNA forms the backbones of the *chromosomes*, which are contained in the nucleus of each cell. In *diploid* organisms (higher plants and animals) chromosomes form homologous pairs, each one inherited from one parent. The exceptions are the *sex chromosomes*, which are involved in the genetic determination of sex. Usually, this is one pair of chromosomes which differ from each other, one called the X-chromosome, the other the Y-chromosome. In all mammals, in *Drosophila*, and in many other species and taxa, but not in birds, XX is female and XY is male. The term *autosome* is used for chromosomes that are not sex chromosomes. The number of different chromosomes per nucleus is characteristic of each species.

Any heritable change in the genetic material is called a *mutation*. Mutations are the ultimate source of genetic variability, and form the raw material upon which selection can act. Although the term mutation includes changes in chromosome structure and number, the vast majority of genetic variation is caused by gene mutations. Modern genetics has revealed that at the molecular level (gene) mutations occur in many different ways, for instance as base substitutions, in which one pair of nucleotides is replaced by another, as insertions or deletions of DNA, as inversions of sequences of nucleotides, or as transpositions. The latter are mainly caused by transposable elements changing their position from one site to another. For many population-genetic models, however, the molecular origin of a mutant is not necessarily of relevance. What often counts is only the rate at which mutations occur and a mutant's effect on fitness or, more generally, on the character under consideration. Typically, spontaneous mutation rates per locus per generation are of the order of 10^{-4} to 10^{-6}, and genomic mutation rates summed over all loci may be on the order of one per generation, but can vary substantially between species.

During meiosis, different chromosomes assort independently and *crossing over* between two homologous chromosomes may occur. Consequently, the newly formed gamete contains maternal alleles at one set of loci and paternal alleles at the complementary set.

This process is called *recombination*. Since it leads to random association between alleles at different loci, recombination has the potential to combine favorable alleles of different ancestry in one gamete and to break up combinations of deleterious alleles. These properties are thought to confer a substantial evolutionary advantage to sexual species relative to asexuals.

The mating pattern may have a substantial influence on the evolution of gene frequencies. The simplest and most important mode is *random mating*. This means that matings take place without regard to ancestry or the genotype under consideration. It seems to occur frequently in nature. For example, among humans, matings within a population appear to be random with respect to blood groups and allozyme phenotypes, but are non-random with respect to height. Random mating conserves allele frequencies and, after one generation, genotypic frequencies.

Selection occurs when individuals of different genotype leave different numbers of progeny because they differ in their probability to survive to reproductive age (*viability*), in their mating success, or in their average number of produced offspring (*fertility*). Darwin (1859) recognized and documented the central importance of selection as the driving force for adaptation and evolution. Since selection affects the entire genome, its consequences for the genetic composition of a population may be complex. Selection is measured in terms of *fitness* of individuals, i.e., by the number of progeny contributed to the next generation. There are different measures of fitness, and it consists of several components because selection may act on each stage of the life cycle.

The main purpose of these lecture notes is to provide the basic mathematical theory for understanding and predicting the evolutionary change within populations or species under the action of selection, mutation, and recombination. In this first part, we focus on deterministic models, which are the appropriate tool to understand these mechanisms if they act in a very large population so that random effects can be ignored. In a second part, we will introduce stochastic models that describe the interaction of so-called *random genetic drift* with selection, mutation, and recombination. Random genetic drift occurs naturally because reproduction has inherent stochastic components caused by random sampling of the genes transmitted from parents to offspring. The smaller the population size, the more important becomes random genetic drift.

As a general introduction to evolutionary biology, we recommend Barton et al. (2007). An excellent comprehensive introduction to theoretical evolutionary genetics and population genetics is the book by Charlesworth and Charlesworth (2010). Classical texts on theoretical population genetics are Crow and Kimura (1970) and Ewens (1979). More recent treatments of (various aspects of) theoretical population genetics include Nagylaki (1992), Bürger (2000), on which most of the material in these lecture notes is based, Ewens

(2004), and Wakeley (2008). The latter two books are primarily dedicated to stochastic models.

2 The Hardy–Weinberg Law

With the blending theory of inheritance, variation in a population declines rapidly, and this was one of the arguments against Darwin's theory of evolution. With Mendelian inheritance there is no such dilution of variation, as was shown independently by the famous British mathematician Hardy and the German physician Weinberg. In fact, only two years after the rediscovery of Mendelian heredity, Yule (1902) had pointed out that the ratio $1:2:1$ of the frequency of genotypes $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, $\mathcal{A}_2\mathcal{A}_2$, as obtained in the F_2 generation of a cross of $\mathcal{A}_1\mathcal{A}_1$ and $\mathcal{A}_2\mathcal{A}_2$ individuals, persisted in further random bred generations. Castle (1903) extended this to other gene frequencies. The general principle, on which these observations are based, and which is now called the Hardy–Weinberg Law, was discovered independently by Hardy (1908) and Weinberg (1908). We first derive the simple original version of this law, where at a given gene locus only two alleles, $\mathcal{A}_1$ and $\mathcal{A}_2$, occur. Thereafter, we shall treat some extensions.

2.1 Two alleles

We consider a random-mating population with discrete, nonoverlapping generations (as in annual plants and many insects) that is either *monoecious* (i.e., every individual has both male and female sexual organs, as in most plants and some animals) or *dioecious* (i.e., admits two sexes) with initially identical genotype frequencies in both sexes. Then the relative frequencies of the genotypes $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, and $\mathcal{A}_2\mathcal{A}_2$ in the population can be described by one set of variables, labeled P, $2Q$, and R, respectively, and $P+2Q+R=1$. The heterozygous genotype $\mathcal{A}_1\mathcal{A}_2$ has been assumed to be unordered, so that $2Q$ is the combined frequency of the ordered genotypes $\mathcal{A}_1\mathcal{A}_2$ and $\mathcal{A}_2\mathcal{A}_1$. We assume, furthermore, that the population is so large that gene and genotype frequencies may be treated as deterministic, and relative frequency can be identified with probability.

We want to derive the frequencies of the three genotypes $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, and $\mathcal{A}_2\mathcal{A}_2$ in the next generation. This can be achieved by calculating the frequencies of all possible matings and their offspring produced. For example, with random mating (with respect to the locus under consideration), the probability of the mating $\mathcal{A}_1\mathcal{A}_1\times\mathcal{A}_1\mathcal{A}_2$ is $4PQ$, because $\mathcal{A}_1\mathcal{A}_1$ can be male or female (and $\mathcal{A}_1\mathcal{A}_2$, thus, female or male), and the probabilities of the genotypes $\mathcal{A}_1\mathcal{A}_1$ and $\mathcal{A}_1\mathcal{A}_2$ are P and $2Q$, respectively. According to Mendel's laws, half of the progeny of such a mating are $\mathcal{A}_1\mathcal{A}_1$ and half are $\mathcal{A}_1\mathcal{A}_2$. Table 2.1 summarizes all possibilities.

Table 2.1: Mating table

Mating	Mating prob.	Cond. prob. of progeny		
		$\mathcal{A}_1\mathcal{A}_1$	$\mathcal{A}_1\mathcal{A}_2$	$\mathcal{A}_2\mathcal{A}_2$
$\mathcal{A}_1\mathcal{A}_1 \times \mathcal{A}_1\mathcal{A}_1$	P^2	1	0	0
$\mathcal{A}_1\mathcal{A}_1 \times \mathcal{A}_1\mathcal{A}_2$	$4PQ$	$\frac{1}{2}$	$\frac{1}{2}$	0
$\mathcal{A}_1\mathcal{A}_1 \times \mathcal{A}_2\mathcal{A}_2$	$2PR$	0	1	0
$\mathcal{A}_1\mathcal{A}_2 \times \mathcal{A}_1\mathcal{A}_2$	$4Q^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$\mathcal{A}_1\mathcal{A}_2 \times \mathcal{A}_2\mathcal{A}_2$	$4QR$	0	$\frac{1}{2}$	$\frac{1}{2}$
$\mathcal{A}_2\mathcal{A}_2 \times \mathcal{A}_2\mathcal{A}_2$	R^2	0	0	1

Therefore, the frequency of $\mathcal{A}_1\mathcal{A}_1$ homozygotes in the next generation is[2]

$$\begin{aligned} P' &= P^2 \cdot 1 + 4PQ \cdot \tfrac{1}{2} + 2PR \cdot 0 + 4Q^2 \cdot \tfrac{1}{4} + 4QR \cdot 0 + R^2 \cdot 0 \\ &= P^2 + 2PQ + Q^2 = (P+Q)^2 \end{aligned} \tag{2.1a}$$

and, similarly,

$$2Q' = 2PQ + 2PR + 2Q^2 + 2QR = 2(P+Q)(Q+R) \tag{2.1b}$$

and

$$R' = Q^2 + 2QR + R^2 = (Q+R)^2 . \tag{2.1c}$$

Here we have assumed that no mutation occurs and that no evolutionary forces, such as viability selection, differential fertility, geographical dispersal, or separate sexes, change the genotype frequencies. By substituting P', Q', and R' into the right-hand sides of (2.1a)–(2.1c), and observing the fact that $P + 2Q + R = 1$, we obtain after another generation of random mating

$$P'' = (P' + Q')^2 = (P+Q)^2 = P' \tag{2.2a}$$

and, similarly,

$$Q'' = Q' \qquad \text{and} \qquad R'' = R' . \tag{2.2b}$$

Thus, the genotype frequencies established after one generation of random mating are maintained under random mating in all subsequent generations.

Now let us consider the gene frequencies p and $q = 1-p$ of the alleles $\mathcal{A}_1$ and $\mathcal{A}_2$. Since all the genes in $\mathcal{A}_1\mathcal{A}_1$ individuals, and half the genes in $\mathcal{A}_1\mathcal{A}_2$ individuals, are $\mathcal{A}_1$ genes,

[2]Unless stated otherwise, a prime, $'$, signifies the next generation.

therefore $p = \frac{1}{2}(2P + 2Q)$ and similarly $q = Q + R$. Hence, we can rewrite the equations (2.1) as

$$P' = p^2, \quad 2Q' = 2pq, \quad R' = q^2 \,. \tag{2.3}$$

These are the famous Hardy–Weinberg proportions, and the Hardy–Weinberg Law states that after one generation of random mating, the genotype frequencies remain constant and can be expressed in terms of the allele frequencies according to (2.3). In particular, the allele (gene) frequencies remain constant and no genetic variability is lost by random mating.

2.2 The case of k alleles

Next, we generalize the Hardy–Weinberg Law to the case of k alleles, as first obtained by Weinberg (1909). We denote the alleles by $\mathcal{A}_i$, $i = 1, \ldots, k$, the frequency of the ordered genotype $\mathcal{A}_i\mathcal{A}_j$ by P_{ij} so that the frequency of the unordered genotype $\mathcal{A}_i\mathcal{A}_j$ is $P_{ij} + P_{ji} = 2P_{ij}$. Then the frequency of allele $\mathcal{A}_i$ in the population is

$$p_i = \sum_{j=1}^{k} P_{ij} \,. \tag{2.4}$$

The unordered genotype $\mathcal{A}_i\mathcal{A}_j$ $(i \neq j)$ can result from the unordered matings $\mathcal{A}_i\mathcal{A}_m \times \mathcal{A}_l\mathcal{A}_j$. It is convenient to classify these matings according to the number (and kind) of heterozygous genotypes involved: $\mathcal{A}_i\mathcal{A}_i \times \mathcal{A}_j\mathcal{A}_j$, $\mathcal{A}_i\mathcal{A}_m \times \mathcal{A}_j\mathcal{A}_j$, $\mathcal{A}_i\mathcal{A}_i \times \mathcal{A}_l\mathcal{A}_j$, $\mathcal{A}_i\mathcal{A}_j \times \mathcal{A}_i\mathcal{A}_j$, and $\mathcal{A}_i\mathcal{A}_m \times \mathcal{A}_l\mathcal{A}_j$, where $m \neq i$, $l \neq j$, and $(m, l) \neq (j, i)$. These matings occur with probabilities $2(P_{ii}P_{jj})$, $2(2P_{im} \cdot P_{jj})$, $2(P_{ii} \cdot 2P_{lj})$, $2P_{ij} \cdot 2P_{ij}$, $2(2P_{im} \cdot 2P_{lj})$, respectively. The conditional probabilities that an offspring of such a mating is of genotype $\mathcal{A}_i\mathcal{A}_j$ are 1, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively. Therefore, the total probability of an $\mathcal{A}_i\mathcal{A}_j$ genotype in the next generation is

$$2P'_{ij} = 2P_{ii}P_{jj} + 2\sum_{m\neq i} P_{im}P_{jj} + 2\sum_{l\neq j} P_{ii}P_{lj} + 2P_{ij}^2 + 2\sum_{m\neq i}\sum_{\substack{l\neq j \\ (m,l)\neq(j,i)}} P_{im}P_{lj} \,.$$

Since this is exactly $2\sum_m \sum_l P_{im}P_{lj} = 2(\sum_m P_{im})(\sum_l P_{il})$[3], it follows that

$$P'_{ij} = p_i p_j \qquad \text{for every } i \text{ and } j \,, \tag{2.5}$$

because the case $i = j$ can be proved in a similar but easier manner.

A population in which the genotype frequencies satisfy these equations is said to be in *Hardy–Weinberg equilibrium*. A mathematically trivial but biologically important

[3]Throughout, sums without limits run over all possible indices. Thus, $\sum_m P_{im} = \sum_{m=1}^{k} P_{im}$.

consequence of (2.5) is that gene frequencies remain constant across generations, i.e.,

$$p'_i = p_i \quad \text{for every } i\,. \tag{2.6}$$

If random union of gametes is posited, as may be realistic for some marine organisms, then (2.5) follows from the definition of random union.

2.3 Separate Sexes

Hardy–Weinberg equilibrium is also attained, though delayed by one generation, in a population with separate sexes and different initial genotypic frequencies at an autosomal (not sex-linked) locus. Let P_{ij} and Q_{ij} be the frequencies of the ordered genotype $\mathcal{A}_i\mathcal{A}_j$ of males and females, respectively. The gene frequencies in the two sexes are $p_i = \sum_j P_{ij}$ and $q_i = \sum_j Q_{ij}$. An argument similar to that leading to (2.5) shows that after one generation of random mating the genotypic frequencies in the two sexes are equal and

$$P'_{ij} = Q'_{ij} = \tfrac{1}{2}(p_i q_j + p_j q_i)\,, \tag{2.7}$$

and the gene frequencies are

$$p'_i = q'_i = \tfrac{1}{2}(p_i + q_i)\,. \tag{2.8}$$

Thus, as shown above, another generation of random mating yields Hardy–Weinberg ratios

$$P''_{ij} = Q''_{ij} = p'_i q'_j\,. \tag{2.9}$$

2.4 X-Linkage

In most higher organisms sex is determined by a pair of non-homologous chromosomes, the sex chromosomes. One sex has chromosomes XX, the other XY. We assume that the males are the heterogametic sex XY and females are XX. Genes carried on the X-chromosome are called X-linked and genes on the Y-chromosome are said to be Y-linked. Formally, the dynamics of Y-linked genes is identical to that in haploid populations. X-linked loci are of considerable importance in human genetics, and we shall now investigate the validity of the Hardy–Weinberg Law for such genes.

Let the relative frequency of the allele $\mathcal{A}_i$ be p_i in males and q_i in females. The frequency of ordered genotypes $\mathcal{A}_i\mathcal{A}_j$ in females is denoted by Q_{ij}. Since a male inherits its gene from the mother, we have

$$p'_i = q_i\,. \tag{2.10}$$

Under the assumption of random mating, the genotype frequencies among females in the next generation are

$$Q'_{ij} = \tfrac{1}{2}\sum_l (p_i Q_{jl} + p_j Q_{il}) = \tfrac{1}{2}(p_i q_j + p_j q_i)\,, \tag{2.11}$$

because $q_i = \sum_j Q_{ij}$. It follows that

$$q_i' = \tfrac{1}{2}(p_i + q_i)\,. \tag{2.12}$$

Therefore, the frequency of $\mathcal{A}_i$ in the male gene pool satisfies the recursion relation

$$p_i'' = \tfrac{1}{2}(p_i' + p_i)\,. \tag{2.13}$$

Let $x_i = \frac{1}{3}(p_i + 2q_i)$ denote the average frequency of $\mathcal{A}_i$ in the entire population, and let $y_i = p_i - q_i$ be the difference between male and female gene frequencies. Then (2.10) and (2.12) imply $x_i' = x_i$ and $y_i' = -\frac{1}{2}y_i$. Therefore, $x_i(t) = x_i(0)$ and $y_i(t) = y_i(0)(-\frac{1}{2})^t$. It follows that

$$p_i(t) = x_i(0) + \tfrac{2}{3}(-\tfrac{1}{2})^t y_i(0)\,, \tag{2.14a}$$

$$q_i(t) = x_i(0) - \tfrac{1}{3}(-\tfrac{1}{2})^t y_i(0)\,. \tag{2.14b}$$

The same result is, of course, obtained by directly solving (2.13). Equations (2.14) show that the allele frequencies converge to Hardy–Weinberg proportions,

$$p_i = q_i = x_i(0) \quad \text{and} \quad Q_{ij} = x_i(0)x_j(0)\,, \tag{2.15}$$

in an oscillatory manner. Convergence is rapid, but no longer occurs in one or two generations.

3 Selection at a single locus

Selection is the major driving force of evolution. It occurs when genotypes in a population differ in their fitnesses, i.e., in their viability, mating success, or fertility and, therefore, leave different numbers of progeny. The basic mathematical models of selection were developed and investigated in the 1920s and early 1930s by Fisher (1930), Wright (1931), and Haldane (1932). Here, we shall be concerned mainly with the consequences of selection caused by differential viability for the evolution of a population. Our treatment concentrates on the diploid one-locus case in discrete time. We will focus on the equilibrium properties and the evolutionary dynamics of allele frequencies. A central result concerns Fisher's Fundamental Theorem of Natural Selection about the increase of mean fitness. Finally, the dynamics in continuous time will be derived and briefyl investigated.

3.1 Viability selection in diploid populations with discrete generations

We assume discrete and nonoverlapping generations, random mating, and that genotype frequencies are the same in both sexes (as is the case if individuals are monoecious, or if

they are dioecious with the same viabilities in both sexes and the same sex ratio in all matings). Suppose that at an autosomal locus the alleles $\mathcal{A}_1, \ldots, \mathcal{A}_k$ can occur. We count individuals at the zygote stage and denote the (relative) frequency of $\mathcal{A}_i\mathcal{A}_i$ homozygotes by P_{ii} and that of (unordered) $\mathcal{A}_i\mathcal{A}_j$ heterozygotes by $2P_{ij}$ (cf. Section 2.2). Then the frequency of the allele $\mathcal{A}_i$ is

$$p_i = \sum_j P_{ij} \,. \tag{3.1}$$

Since mating is at random, the genotype frequencies P_{ij} are in Hardy–Weinberg proportions (2.5). Let us suppose that selection acts solely through differential viabilities, and denote the fitness (viability) of $\mathcal{A}_i\mathcal{A}_j$ individuals by W_{ij}. The fitnesses W_{ij} satisfy $W_{ij} \geq 0$ and $W_{ij} = W_{ji}$, because they belong to the same (unordered) genotype $\mathcal{A}_i\mathcal{A}_j$. Then the frequency of $\mathcal{A}_i\mathcal{A}_j$ genotypes among adults that have survived selection is

$$P^*_{ij} = \frac{W_{ij}P_{ij}}{\overline{W}} = \frac{W_{ij}p_ip_j}{\overline{W}} \,, \tag{3.2}$$

where

$$\overline{W} = \sum_{i,j} W_{ij}P_{ij} = \sum_{i,j} W_{ij}p_ip_j = \sum_i W_ip_i \tag{3.3}$$

is the mean fitness and

$$W_i = \sum_j W_{ij}p_j \tag{3.4}$$

is the *marginal fitness* of allele $\mathcal{A}_i$. In particular, the frequency of $\mathcal{A}_i$ after selection is $p^*_i = \sum_j P^*_{ij} = W_ip_i/\overline{W}$. Because of random mating, the allele frequency p'_i among zygotes of the next generation is also p^*_i, so that allele frequencies evolve according to the *selection equation*

$$p'_i = p_i \frac{W_i}{\overline{W}} \quad \text{for } i = 1, \ldots, k \,. \tag{3.5}$$

This recursion equation preserves the relation

$$\sum_i p_i = 1 \,, \tag{3.6}$$

and describes the evolution of allele frequencies at a single autosomal locus in a diploid population.

We note that the right-hand side of (3.5) remains unchanged if all fitness values W_{ij} are multiplied by the same constant. It is often convenient to introduce such a scaling, and to use *relative fitness* instead of *absolute fitness*, which is measured by the expected number of progeny of individuals of a given genotype. In particular, evolution of the gene frequencies in the population is independent of the growth rate of the population, and constant population size can be assumed (this can be shown to be true even if population regulation acts identically on all alleles).

Remark 3.1. The consideration of differential fertilities leads to a much more complex model, because fertilities have to be assigned to each mating pair $\mathcal{A}_i\mathcal{A}_j \times \mathcal{A}_k\mathcal{A}_l$. Such a general model is derived in Nagylaki (1992), where some of its features are explored. For multiplicative fertilities, this model reduces to the present one, (3.5). Otherwise, its analysis requires following the change of genotype frequencies across generations. A lucid account of the mathematical properties of the diallelic pure fertility equation can be found in Hofbauer and Sigmund (1998), who also provide further references.

3.2 Multiplicative fitnesses and selection on haploids

If fitnesses are *multiplicative*, i.e., if constants v_i, $i = 1, \ldots, k$, exist such that $W_{ij} = v_i v_j$ for every i, it is easily shown that (3.5) reduces to the allele frequency dynamics in haploid populations,

$$p_i' = p_i \, \frac{v_i}{\bar{v}} \,, \quad \text{for every } i \,. \tag{3.7}$$

For haploid populations, (3.7) is readily derived because fitnesses (viabilities) $v_i \geq 0$ can be assigned directly to alleles $\mathcal{A}_i$. The mean fitness of the population is $\bar{v} = \sum_j v_j p_j$, which is linear in the p_i (in contrast to the general diploid case). The difference equation (3.7) can be solved explicitly, and its solution is

$$p_i(t) = \frac{p_i(0) v_i^t}{\sum_j p_j(0) v_j^t} \,, \quad i = 1, \ldots, k \,, \tag{3.8}$$

Formally, the haploid selection dynamics (3.7) can therefore be considered as a special case of the diploid dynamics (3.5). However, it should be noted that the mean fitness $\overline{W}$ in the diploid multiplicative model is $\overline{W} = \bar{v}^2$.

Remark 3.2. In haploid populations, the best allele becomes fixed. We assume (3.7). If one allele, say $\mathcal{A}_1$ has higher fitness than all others ($v_1 > v_i$ for every $i \neq 1$), then $(v_i/v_1)^t \to 0$ for $i \neq 1$ as $t \to \infty$. Therefore, (3.8) shows (prove this!) that $p_1(t) \to 1$ as $t \to \infty$, i.e., in the long run the best allele becomes fixed.

As we shall see, this is not necessarily so in diploids.

Example 3.3.[4] Here we show by way of examples, how efficient selection can be in the long run. Consider two alleles, $\mathcal{A}_1$ and $\mathcal{A}_2$. If $\mathcal{A}_1$ is the wild type and $\mathcal{A}_2$ is a new beneficial allele, we may set (without loss of generality!) $v_1 = 1$ and $v_2 = 1 + s$. Then we obtain from (3.8):

$$\frac{p_2(t)}{p_1(t)} = \frac{p_2(0)}{p_1(0)} \left(\frac{v_2}{v_1} \right)^t = \frac{p_2(0)}{p_1(0)} (1+s)^t \,.$$

Thus, $\mathcal{A}_2$ increases geometrically relative to $\mathcal{A}_1$.

[4] The end of an example is signified by the sign $\lhd$

If $s = 0.5$, then after 10 generations the frequency of $\mathcal{A}_2$ has increased by a factor of $(1+s)^t = 1.5^{10} \approx 57.7$ relative to $\mathcal{A}_1$.

If $s = 0.05$ and $t = 100$, this factor is $(1+s)^t = 1.05^{100} \approx 131.5$.

Thus, slight fitness differences may have a big long-term effect! $\lhd$

3.3 The case of two alleles

It is instructive to consider the diallelic case in some detail, because it exhibits several of the basic properties of the multiallelic case, but can be analyzed by elementary means. We shall write p and $1-p$ instead of p_1 and p_2. We use relative fitnesses and assume

$$W_{11} = 1\,, \quad W_{12} = 1 - hs\,, \quad \text{and } W_{22} = 1 - s\,, \tag{3.9}$$

where s is called the *selection coefficient* and h describes the degree of dominance. Because fitnesses are nonnegative, we assume $s \le 1$ and $hs \le 1$. The allele $\mathcal{A}_1$ is called *dominant* if $h = 0$, partially dominant if $0 < h < \frac{1}{2}$, *recessive* if $h = 1$, and partially recessive if $\frac{1}{2} < h < 1$. The expressions *additivity*, or *no dominance*, refer to $h = \frac{1}{2}$. From (3.4), the marginal fitnesses of the two alleles are

$$W_1 = 1 - hs + hsp \quad \text{and} \quad W_2 = 1 - s + s(1-h)p\,, \tag{3.10}$$

and the mean fitness is

$$\overline{W} = 1 - s + 2s(1-h)p - s(1-2h)p^2\,. \tag{3.11}$$

It is easily verified that the allele-frequency change from one generation to the next can be written as

$$\Delta p = p' - p = \frac{p(1-p)s}{\overline{W}}\,[1 - h - (1-2h)p] \tag{3.12a}$$

$$= \frac{p(1-p)}{2\overline{W}}\,\frac{\mathrm{d}\overline{W}}{\mathrm{d}p}\,. \tag{3.12b}$$

For our analysis we exclude the trivial case of no selection ($s = 0$) and assume $s > 0$. First we observe that $p = 0$ and $p = 1$ are always equilibria of the gene-frequency dynamics (3.12). This is also biologically obvious because we have ignored evolutionary forces such as mutation and migration that could introduce new (or lost) alleles into the population. Any other equilibrium must be a critical point of $\overline{W} = \overline{W}(p)$ and, since $\overline{W}$ is quadratic in p, there can be at most one further equilibrium. There exists a third equilibrium if h is such that $1 - h - (1-2h)p = 0$ for some $0 < p < 1$. This can occur if and only if either $h > 1$ or $h < 0$. In both cases, the equilibrium frequency is

$$\hat{p} = \frac{1-h}{1-2h}\,, \tag{3.13}$$

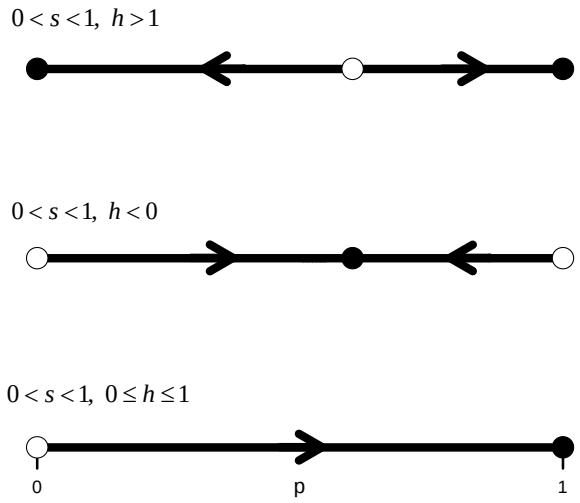

Figure 3.1: Convergence patterns for selection at a diallelic locus. Full circles signify stable equilibria, open circles signify unstable equilibria.

and $\hat{p}$ is a *polymorphic equilibrium*, because both alleles are present in the population with nonzero frequency.

The investigation of the selection equation (3.12) requires consideration of three cases according to the three different convergence patterns that can occur (Figure 3.1).

(*i*) $0 \leq h \leq 1$. Then $\overline{W}$ is an increasing function of p for $p \in [0,1]$ and it follows that $p(t) \to 1$ as $t \to \infty$. Thus, the favored allele $\mathcal{A}_1$ eventually goes to fixation and selection removes all genetic variability. Two particularly simple special cases are that of multiplicative fitnesses, i.e., h such that $1-s = (1-hs)^2$, leading to the asexual dynamics (3.7) and its explicit solution (3.8), and that of *additive* fitnesses (or no dominance), where $h = 1/2$ and $\overline{W}$ is a linear function of p. Although the precise value of h does not influence the eventual results of selection, it has a significant influence on the rate of evolution toward equilibrium. As shown by Figure 3.2, an initially rare, advantageous allele, that is dominant or intermediate, sweeps through the population much faster than a recessive allele. The obvious reason is that a rare allele occurs almost exclusively in heterozygotes, where recessiveness hides it from selection. For analytical results concerning the rate of convergence toward equilibrium in the diallelic case, the reader may consult Nagylaki (1992, Chapter 4.2).

(*ii*) $h < 0$. In this case of *overdominance*, or *heterozygote advantage*, the mean fitness function $\overline{W}(p)$ is concave and the equilibrium $\hat{p}$ (3.13) is the (local) maximum of $\overline{W}$. Using

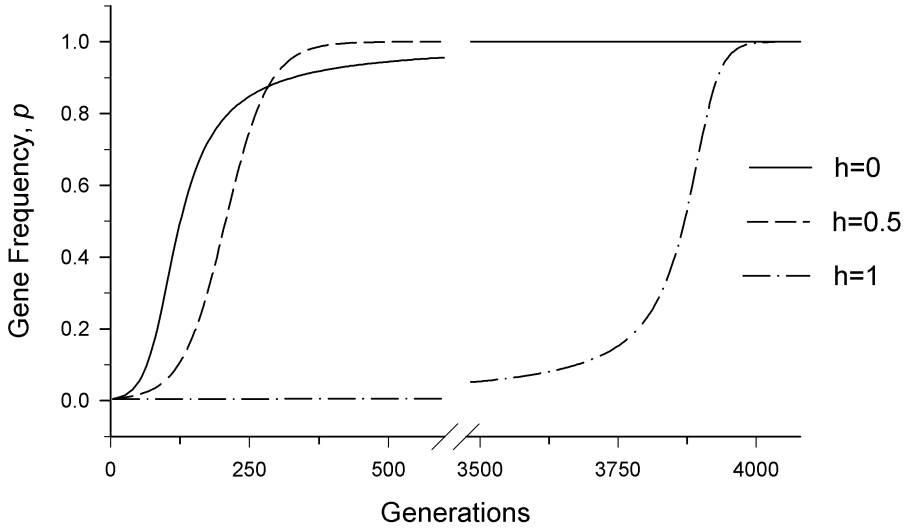

Figure 3.2: Selection of a dominant ($h = 0$, solid line), intermediate ($h = 1/2$, dashed), and recessive ($h = 1$, dash-dotted) allele. The initial frequency is $p_0 = 0.005$ and the selective advantage is $s = 0.05$.

(3.13), we can write (3.12a) as

$$\Delta p = \frac{sp(1-p)}{\overline{W}}(1-2h)(\hat{p}-p)\,. \tag{3.14}$$

As $0 < sp(1-p)(1-2h)/\overline{W} < 1$ for $0 < p < 1$, it follows that convergence to $\hat{p}$ occurs and is monotone, i.e., nonoscillatory. Therefore, the polymorphic equilibrium $\hat{p}$ is *globally asymptotically stable* (see Appendix A for a precise definition).

(*iii*) $h > 1$. In this case of *underdominance*, the polymorphic equilibrium $\hat{p}$ is unstable because fitness is minimized there. Indeed, (3.14) shows that the sign of Δp is the same as that of $p - \hat{p}$, and that $p(t)$ converges monotonically to 0 if $0 < p(0) < \hat{p}$, and to 1 if $\hat{p} < p(0) < 1$. Therefore, $\hat{p}$ is an *unstable* equilibrium and the outcome of evolution depends on the initial state $p(0)$. As in case (i), selection eventually removes any genetic variability and the population becomes *monomorphic*.

3.4 The Fundamental Theorem of Natural Selection

The results above imply that for two alleles mean fitness increases from generation to generation and remains unchanged only at equilibrium states. This is a special case of

Fisher's Fundamental Theorem of Natural Selection (often abbreviated as FTNS), which he formulated as follows:

"The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time." (Fisher 1930)

What Fisher called the genetic variance is now called the genic, or additive genetic, variance. Since a variance is always nonnegative, this implies that mean fitness is nondecreasing. Fisher's statement of the theorem was not based on an explicit dynamical model for gene-frequency change and gave rise to ambiguities. The classical interpretation has been that it is a theorem about the (rate of) increase of mean fitness, stating that this increase is (approximately) equal to the additive genetic variance; cf. (3.23) below. However, as we shall see below, mean fitness is increasing only under restrictive assumptions such as random mating, a single locus, etc., whereas Fisher viewed it as a very general theorem.

In fact, Fisher strongly opposed the interpretation of his FTNS that selection acts to maximize mean fitness (see the discussion on pp. 229-230 in Bennett 1983). Ewens and Lessard (2015) discuss interpretations of the Fundamental Theorem that indeed are very general (by admitting multiple loci, epistasis, strong selection, and also nonrandom mating), but that concern certain partial changes in mean fitness. These general results still await applications that provide biological or mathematical insight into evolutionary processes. Whenever we refer without further specification to the FTNS, we mean its classical interpretation (3.23) as a theorem specifying the change of mean fitness caused by selection as a function of the additive genetic variance. Here, we shall prove the following weaker result:

Theorem 3.4 (Mean fitness is nondecreasing)**.** *We assume the general multiallelic selection dynamics* (3.5)*. Then*

$$\overline{W}' = \sum_{i,j} W_{ij} p_i' p_j' \geq \overline{W}\,, \tag{3.15}$$

and $\overline{W}' = \overline{W}$ *holds only at equilibria of* (3.5)*, i.e., if and only if* $p_i' = p_i$ *or, equivalently, if* $p_i(W_i - \overline{W}) = 0$ *for every* i*. Thus, mean fitness increases (strictly) along nonconstant solutions (trajectories) of* (3.5)*. It remains constant only at equilibrium.*

Remark 3.5. The first proofs that mean fitness is nondecreasing under the general multiallele dynamics (3.5) were given by Scheuer and Mandel (1959), Mulholland and Smith (1959), Atkinson *et al.* (1960), and Kingman (1961a). Note that Theorem 3.4 shows that $\overline{W}$ is a strict Lyapunov function for (3.5).

Below, we shall reproduce the elegant proof of Kingman which is based on two well

known inequalities. The first is the elementary inequality

$$\sqrt{ab} \le \tfrac{1}{2}(a+b)\,, \qquad a,b \ge 0\,. \tag{3.16}$$

The second is a special case of Jensen's inequality

$$f\Big(\sum_i p_i x_i\Big) \le \sum_i p_i f(x_i)\,, \tag{3.17}$$

where f is a convex function on some interval I, $x_i \in I$, and $p_i \ge 0$ such that $\sum_i p_i = 1$. For the function $f(x) = x^\alpha$, $\alpha > 1$, which is convex on $[0,\infty)$, Jensen's inequality yields

$$\Big(\sum_i p_i x_i\Big)^\alpha \le \sum_i p_i x_i^\alpha\,. \tag{3.18}$$

Proof of Theorem 3.4. It is sufficient to show $\overline{W}^{\,2}\,\overline{W}' \ge \overline{W}^{\,3}$ (we exclude the trivial case $\overline{W} = 0$). Using (3.3), (3.5), and (3.4), we obtain

$$\begin{aligned}
\overline{W}^{\,2}\,\overline{W}' &= \overline{W}^{\,2} \sum_{i,j} p_i' p_j' W_{ij} \\
&= \sum_{i,j} (p_i W_i)(p_j W_j) W_{ij} \\
&= \sum_{i,j,\ell} p_i p_j p_\ell W_{ij} W_{i\ell} W_j
\end{aligned} \tag{3.19}$$

and, by exchanging the roles of j and ℓ,

$$\overline{W}^{\,2}\,\overline{W}' = \sum_{i,j,\ell} p_i p_j p_\ell W_{ij} W_{i\ell} W_\ell\,. \tag{3.20}$$

Taking the arithmetic mean of the left-hand sides of (3.19) and (3.20), and that of the right-hand sides, we obtain

$$\begin{aligned}
\overline{W}^{\,2}\,\overline{W}' &= \sum_{i,j,\ell} p_i p_j p_\ell W_{ij} W_{i\ell} \tfrac{1}{2}(W_j + W_\ell) \\
&\ge \sum_{i,j,\ell} p_i p_j p_\ell W_{ij} W_{i\ell} W_j^{1/2} W_\ell^{1/2} && \text{by (3.16)} \\
&= \sum_i p_i \Big(\sum_j W_{ij} p_j W_j^{1/2}\Big)^2 \\
&\ge \Big(\sum_i p_i \sum_j W_{ij} p_j W_j^{1/2}\Big)^2 && \text{by (3.15) with } \alpha = 2 \\
&= \Big(\sum_j p_j W_j^{3/2}\Big)^2 && \text{by (3.4)} \\
&\ge \overline{W}^{\,3} && \text{by (3.18) with } \alpha = \tfrac{3}{2}\,.
\end{aligned}$$

This proves that *mean fitness is nondecreasing.* □

One of the important consequences of this result is, as discussed in more detail in Section 3.5, that the population will evolve steadily to an equilibrium, and that the equilibrium will be asymptotically stable.

We proved that mean fitness is nondecreasing, but Fisher stated that the rate of increase is given by the genic variance. Therefore, let us calculate the change of mean fitness $\Delta\overline{W} = \overline{W}' - \overline{W}$:

$$\begin{aligned}\Delta\overline{W} &= \overline{W}^{-2}\sum_{i,j} p_i p_j W_{ij}(W_i W_j - \overline{W}^2)\\ &= \overline{W}^{-2}\sum_{i,j} p_i p_j W_i W_j (W_{ij} - \overline{W})\\ &= \overline{W}^{-1}\sigma_{\mathrm{A}}^2 + \overline{W}^{-2}\sum_{i,j} p_i p_j (W_i - \overline{W})(W_j - \overline{W})(W_{ij} - \overline{W}),\end{aligned} \tag{3.21}$$

where

$$\sigma_{\mathrm{A}}^2 = 2\sum_i p_i (W_i - \overline{W})^2 \tag{3.22}$$

is the *additive genetic, or genic, variance in fitness.*

For *additive fitnesses*, or absence of dominance, there are constants v_i such that $W_{ij} = v_i + v_j$ for every i and j. Then a simple calculation reveals that the last term in (3.21) vanishes, and

$$\Delta\overline{W} = \sigma_{\mathrm{A}}^2/\overline{W} \tag{3.23}$$

is obtained. This can also be shown in the haploid case.

The fact that the additive genetic variance determines the selection response is an extremely important insight for evolution. Below, it is shown that this is true for general fitness values provided selection is weak. In general, the second term on the right-hand side of (3.21), below denoted by R, does not vanish. However, upper and lower bounds for R can be derived, as indicated in the following remark.

Remark*3.6.[5] [Estimates on the increase of mean fitness in terms of the additive genetic variance] Following Nagylaki (1991), we decompose $W_{ij} - \overline{W} = (W_i - \overline{W}) + (W_j - \overline{W}) + \vartheta_{ij}$. Then R becomes

$$R = \overline{W}^{-2}\sum_{i,j} p_i p_j (W_i - \overline{W})(W_j - \overline{W})\vartheta_{ij}\,, \tag{3.24}$$

and the Cauchy–Schwarz inequality gives

$$\overline{W}^2\,|R| \le \left[\sum_{i,j} p_i p_j \vartheta_{ij}^2\right]^{1/2}\left[\sum_{i,j} p_i p_j (W_i - \overline{W})^2 (W_j - \overline{W})^2\right]^{1/2} = \tfrac{1}{2}\sigma_{\mathrm{D}}\sigma_{\mathrm{A}}^2\,,$$

[5] Remarks with a $*$ contain additional, more advanced material for interested readers, but are not required for the exam. The end of such remarks is marked by the sign [$*$].

where σ_{D} is the standard deviation of the dominance effects of fitness, i.e.,

$$\sigma_{\mathrm{D}}^2 = \sum_{i,j} \vartheta_{ij}^2 p_i p_j = \sum_{i,j} (W_{ij} - W_i - W_j + \overline{W})^2 p_i p_j \tag{3.25}$$

is the *dominance variance*. In addition, we define the *(total) genetic variance*, i.e., the variance of fitnesses of the genotypes as

$$\sigma_{\mathsf{G}}^2 = \sum_{i,j} (W_{ij} - \overline{W})^2 p_i p_j \,. \tag{3.26}$$

It is easy to show that

$$\sigma_{\mathsf{G}}^2 = \sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{D}}^2 \,, \tag{3.27}$$

i.e., the genetic variance can be decomposed into independent additive and dominance effects. (Fisher had invented the method of analysis of variance in 1918 to derive this result.)

Further, we denote the largest and smallest fitness coefficients among the W_{ij} by $W_{\max}$ and $W_{\min}$, exclude lethality by assuming $W_{\min} > 0$, and define the selection coefficient s by $s = (W_{\max} - W_{\min})/W_{\min}$. Then $\sigma_{\mathrm{D}} \le \sigma_{\mathrm{G}} \le s W_{\min}$ and $|R| \le \frac{1}{2} s \sigma_{\mathrm{A}}^2 / \overline{W}$. Therefore, (3.21) becomes

$$\Delta \overline{W} = \frac{\sigma_{\mathrm{A}}^2}{\overline{W}} (1 + E) \,, \tag{3.28}$$

where the relative error E obeys the estimate

$$|E| = \left| R \overline{W} / \sigma_{\mathrm{A}}^2 \right| \le \tfrac{1}{2} s \tag{3.29}$$

(Nagylaki 1991). For $s < 2$, (3.28) and (3.29) yield the lower bound

$$\Delta \overline{W} \ge (1 - \tfrac{1}{2} s) \sigma_{\mathrm{A}}^2 / \overline{W} \tag{3.30a}$$

for the change in mean fitness. For $s > \frac{1}{2}$, the bound

$$\Delta \overline{W} \ge \tfrac{3}{4} \sigma_{\mathrm{A}}^2 / W_{\max} \,, \tag{3.30b}$$

derived by Lyubich *et al.* (1976, 1980) (cf. Lyubich 1992, Chapter 4.2) is better. [$*$]

For weak selection ($s \ll 1$), (3.28) and (3.29) imply the *asymptotic form of the Fundamental Theorem of Natural Selection*,

$$\Delta \overline{W} = \sigma_{\mathrm{A}}^2 / \overline{W} + O(s^3) \,. \tag{3.31}$$

Here, the order (Landau) symbol O is defined as follows: $f(x) = O(g(x))$ as $x \to a$ means $f(x) = g(x) h(x)$, for some function $h(x)$ that is bounded in an interval containing

$x = a$. For instance, $f(s) = g(s) + O(s^n)$ if $|f(s) - g(s)|/s^n \leq$ const as $s \to 0$. If $\lim_{x\to a} |f(x)/g(x)| = 0$, this is often written as $f(x) = o(g(x))$ as $x \to a$.

If, in addition, $W_{\max}$ is normalized to one, then $\overline{W} \approx 1$ and $\Delta\overline{W} \approx \sigma_{\mathrm{A}}^2$. This shows that Fisher's fundamental statement is valid to a close approximation if selection is weak (see also Section 3.6).

3.5 Equilibria and dynamics

Although mean fitness is a nondecreasing function, the equilibrium structure of the diploid selection equation (3.5) may be complicated, ranging from the completely degenerate case of no selection, when every state is an equilibrium, to multiple locally stable equilibria, and to selection regimes with a unique globally asymptotically stable equilibrium. Here we summarize some of the main results, often from a dynamical systems point of view. For definitions and properties of the basic mathematical concepts, the reader is referred to Appendix A.

It is natural to consider the recursion relation (3.5) as a discrete dynamical system on the *simplex*

$$S_k = \left\{ \boldsymbol{p} = (p_1, \ldots, p_k) \in \mathbb{R}^k : \sum_i p_i = 1,\, p_i \geq 0,\, i = 1, \ldots, k \right\}, \tag{3.32}$$

where $\mathbb{R}^k$ denotes the k-dimensional Euclidean space. The state of the population in a given generation is represented by the (column) vector of allele frequencies, $\boldsymbol{p} = (p_1, \ldots, p_k) \in S_k$, the state in the next generation by $\boldsymbol{p}' = (p_1', \ldots, p_k')$, where the p_i' are calculated according to (3.5). Obviously, if $\boldsymbol{p} \in S_k$, then $\boldsymbol{p}' \in S_k$. Thus, the map $\boldsymbol{p} \to \boldsymbol{p}'$ leaves the simplex S_k invariant and defines a discrete dynamical system on S_k. This map can be iterated, and the sequence $\boldsymbol{p}$, $\boldsymbol{p}'$, $\boldsymbol{p}'' = (\boldsymbol{p}')', \ldots, \boldsymbol{p}^{(n)}, \ldots$ is called the *orbit*, or *trajectory*, of $\boldsymbol{p}$. An orbit characterizes the evolution of the allele frequencies in the population.

The simplex S_k is a $(k-1)$-dimensional convex subset of $\mathbb{R}^k$. Its vertices, $\boldsymbol{e}_i = (0, 0, \ldots, 1, 0, \ldots, 0)$, the 1 being the ith component, correspond to the monomorphic states of the population with allele $\mathcal{A}_i$ fixed ($p_i = 1$) and other alleles absent. The interior of S_k consists of all $\boldsymbol{p}$ with $p_i > 0$ for every $i = 1, \ldots, k$, i.e., of all completely polymorphic states. For every proper subset $J \subset \{1, \ldots, k\}$, the set of all $\boldsymbol{p}$, with $p_i = 0$ for $i \in J$, is called a (boundary) face of S_k.

Occasionally, it will be necessary to indicate that marginal and mean fitness depend on $\boldsymbol{p}$, in which case we will write $W_i(\boldsymbol{p})$ and $\overline{W}(\boldsymbol{p})$, respectively. We shall denote the *fitness matrix*, i.e., the $k \times k$ matrix of fitness values W_{ij}, by W. Then $W_i(\boldsymbol{p}) = (\mathsf{W}\boldsymbol{p})_i$ (the ith component of the vector $\mathsf{W}\boldsymbol{p}$) and $\overline{W}(\boldsymbol{p}) = \boldsymbol{p}^\top \mathsf{W}\boldsymbol{p} = \sum_i p_i (\mathsf{W}\boldsymbol{p})_i$.[6]

[6]The superscript $^\top$ denotes transposition of a vector or a matrix.

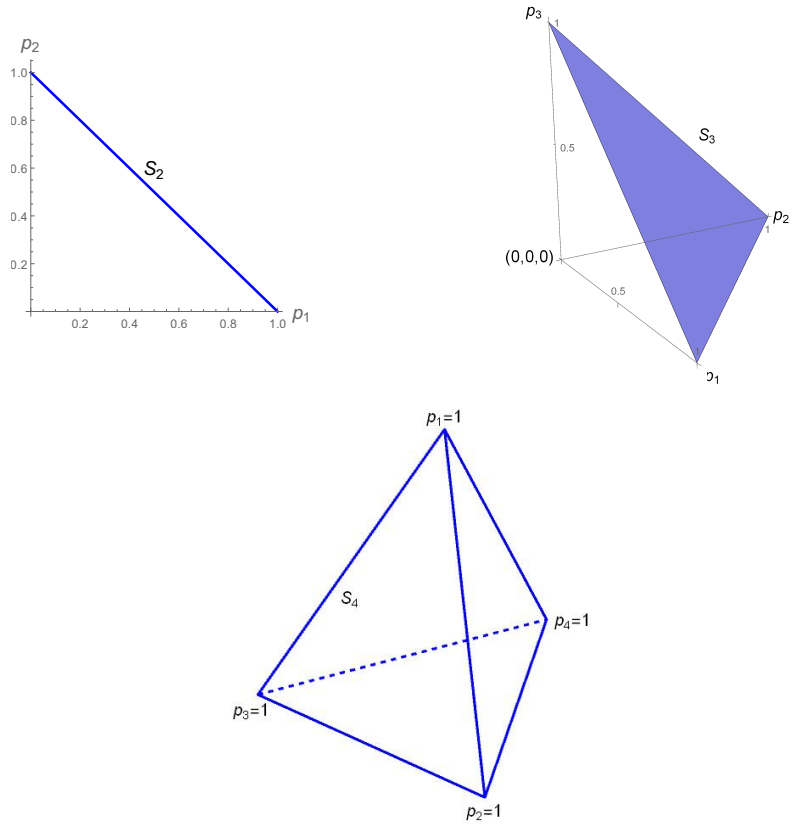

Figure 3.3: Visualization of the simplices S_2, S_3, and S_4

First, we investigate the equilibria, or fixed points, of (3.5). By definition, these are the points $\hat{\boldsymbol{p}}$ satisfying $\hat{\boldsymbol{p}}' = \hat{\boldsymbol{p}}$. From (3.5), we get the equilibrium conditions

$$p_i(W_i - \overline{W}) = 0\,, \qquad 1 \le i \le k\,, \tag{3.33}$$

whence it follows that $\hat{\boldsymbol{p}}$ is an equilibrium if and only if $W_i(\hat{\boldsymbol{p}}) = \overline{W}(\hat{\boldsymbol{p}})$ for every i with $\hat{p}_i > 0$. The monomorphic states $\boldsymbol{e}_i$ are always equilibrium points. A completely polymorphic equilibrium $\hat{\boldsymbol{p}}$ is an equilibrium in the interior of S_k. Therefore, it satisfies $\hat{p}_i > 0$ for every i, and is characterized by the equations

$$W_i(\hat{\boldsymbol{p}}) = \overline{W}(\hat{\boldsymbol{p}}) \quad \text{ for every } i = 1, \ldots, k\,. \tag{3.34}$$

A simple characterization of equilibria in terms of critical points of $\overline{W}$ is the following (Mandel, 1959):

Theorem 3.7. *A state $\hat{\boldsymbol{p}}$ is an equilibrium if and only if it is a critical point for the restriction of mean fitness $\overline{W}(\boldsymbol{p})$ to the minimal face of S_k that contains $\hat{\boldsymbol{p}}$.*

Proof. For the proof, it is clearly sufficient to assume that $\hat{\boldsymbol{p}}$ is an internal equilibrium. The critical points of $\overline{W}(\boldsymbol{p})$ subject to the constraint $\sum_i p_i = 1$ are found by the method of Lagrange multipliers, i.e., by determining the critical points of the auxiliary function $g(\boldsymbol{p}) = \overline{W}(\boldsymbol{p}) - \lambda \sum_i p_i$. These are given by $\partial g/\partial p_i = 2W_i(\boldsymbol{p}) - \lambda = 0$, $i = 1, \ldots, k$. It follows that $0 = \sum_i p_i(\partial g/\partial p_i) = 2\overline{W} - \lambda$. Therefore, $W_i = \overline{W}$ for every i, and (3.34) yields the assertion. □

The following is essentially a reformulation of Theorem 3.4 on the increase of mean fitness:

Theorem 3.8. *A state $\hat{\boldsymbol{p}}$ is an equilibrium if and only if $\overline{W}'(\hat{\boldsymbol{p}}) = \overline{W}(\hat{\boldsymbol{p}})$. Otherwise, $\Delta\overline{W} > 0$ and the mean fitness is strictly increasing along orbits.*

Therefore, $\overline{W}$ is a strict Lyapunov function (cf. Appendix A) for the selection dynamics (3.5) and, hence, the central tool for deriving stability results.

Concerning the number of possible equilibria, we have already noted that all monomorphic states are equilibrium points. Thus, *there are at least k equilibria.* In general, continua of equilibria may exist. However, if the fitness matrix W is such that only a finite number equilibria exists, an upper bound is easily derived:

Theorem 3.9. *If the number of equilibria is finite, then it is less or equal than $2^k - 1$.*

Proof. Indeed, let $S \subseteq \{1, \ldots, k\}$ be nonempty, and denote the smallest element of S by $i(S)$. We consider the system of k linear equations

$$W_i(\boldsymbol{p}) - W_{i(S)}(\boldsymbol{p}) = 0\,, \quad i \in S,\ i > i(S)\,, \tag{3.35a}$$

$$p_i = 0\,, \quad i \notin S\,, \tag{3.35b}$$

$$\sum_{i \in S} p_i = 1\,. \tag{3.35c}$$

This linear system has zero, one, or infinitely many solutions. Since there exist $2^k - 1$ nontrivial subsets S, our assertion is proved. □

A simple example of a system with exactly $2^k - 1$ equilibria is obtained, if we choose $W_{ii} = 1$ for every i, and $W_{ij} = 0$ for $i \neq j$. Then the equilibria are exactly the centers of each face of S_k. Elementary linear algebra tells us that (3.35) admits exactly one solution if the principal submatrix W_S of W is nonsingular, i.e., if $\det(\mathsf{W}_S) \neq 0$. ($\mathsf{W}_S$ is the matrix formed by the rows and columns of W that correspond to S.) However, this solution does not necessarily satisfy $p_i \geq 0$ for every i. Therefore, the number of possible equilibria is finite and, hence, bounded by $2^k - 1$, if all principal submatrices W_S are nonsingular. Otherwise, lines or, more generally, linear manifolds of equilibria may exist.

Example 3.10 (An example with three alleles, adapted from Example 4.10 in Nagylaki and Lou, 2006)**.** We explore the equilibrium and stability structure of the 3-allele system with the fitness matrix

$$\mathsf{W} = \begin{array}{c} \begin{matrix} A_1 & A_2 & A_3 \end{matrix} \\ \begin{pmatrix} 1 & 1 & 1+s_1 \\ 1 & a & 1+s_2 \\ 1+s_1 & 1+s_2 & a \end{pmatrix} \end{array} \begin{matrix} A_1 \\ A_2 \\ A_3 \end{matrix}$$

Throughout, we assume $1 > a \geq 0$, $s_1 > 0$, and $s_2 > 0$. The three monomorphic equilibria $p_1 = 1$, $p_2 = 1$, and $p_3 = 1$ (also called corner or vertex equilibria) exist always.

Next, we study existence of equilibria with two alleles are present. They lie on the edges of the simplex. For the (A_1, A_2) system, we have

$$W_{11} = 1\,,\ W_{12} = 1\,,\ W_{22} = a\,,$$

whence A_1 is (completely) dominant and will reach fixation in this subsystem.

For the (A_1, A_3) system, we have

$$W_{11} = 1\,,\ W_{13} = 1 + s_1\,,\ W_{33} = a\,,$$

so that there is overdominance and the polymorphic equilibrium

$$\hat{p}^{(13)} = \left(\frac{1 - a + s_1}{1 - a + 2s_1}, 0, \frac{s_1}{1 - a + 2s_1} \right)$$

is globally asymptotically stable in this subsystem.

For the (A_2, A_3) system, we have

$$W_{22} = a\,,\ W_{23} = 1 + s_2\,,\ W_{33} = a\,,$$

so that there is overdominance and the polymorphic equilibrium

$$\hat{p}^{(23)} = \left(0, \frac{1}{2}, \frac{1}{2} \right)$$

is globally asymptotically stable in this subsystem.

Next, we determine if an internal, or fully polymorphic, equilibrium exists. For this purpose we have to solve the system of equations

$$W_1 = \overline{W}\,,\ W_2 = \overline{W}\,,\ W_3 = \overline{W}\,.$$

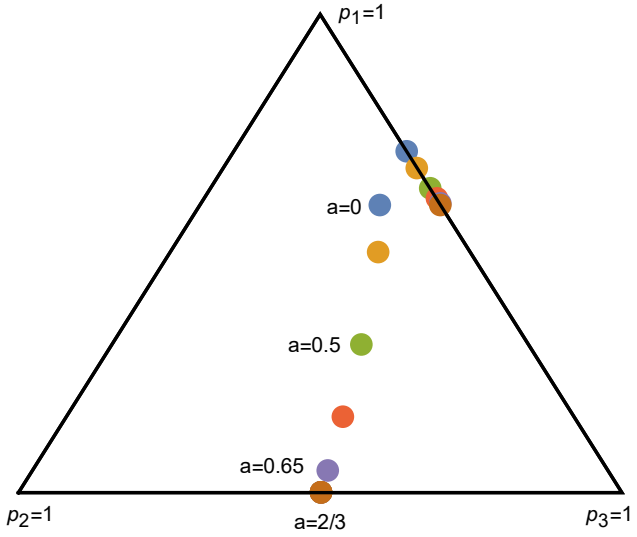

Figure 3.4: Polymorphic equilibria for $a = 0, 0.25, 0.5, 0.6, 0.65, \frac{2}{3}$. The parameters $s_1 = \frac{2}{3}$ and $s_2 = 1$ are fixed. The bifurcation occurs at $a = 1 + s_1 - s_2 = \frac{2}{3}$

Because of the constraint $p_1 + p_2 + p_3$, it is sufficient to solve two of them. After some computation, we obtain the following potential solution $\hat{p} = (\hat{p}_1, \hat{p}_2, \hat{p}_3)$:

$$\begin{aligned}
\hat{p}_1 &= \frac{(1 - a + s_2)(1 - a + s_1 - s_2)}{u}\,, \\
\hat{p}_2 &= \frac{s_1(s_2 - s_1)}{u}\,, \\
\hat{p}_3 &= \frac{(1 - a)s_1}{u}\,,
\end{aligned}$$

where $u = (1-a)^2 - (s_1 - s_2)^2 + 2s_1(1-a)$. It is not difficult to check that the equilibrium $\hat{p}$ is in the interior of the simplex if and only if

$$1 + s_1 > s_2 > s_1 > 0 \quad \text{and} \quad a < 1 + s_1 - s_2\,.$$

It is easily shown that $\hat{p}$ coincides with the polymorphic equilibrium on the edge $p_1 = 0$ if and only if $a = 1 + s_1 - s_2$. Thus, for increasing a, there is a bifurcation at $a = 1 + s_1 - s_2$ and $\hat{p}$ leaves the state space through the overdominant equilibrium $\hat{p}^{(23)}$ at $p_1 = 0$.

Now we examine the *stability of the equilibria*. From the above analysis, we immediately infer that the three monomorphic equilibria are unstable.

To determine the stability (in the full state space) of the other equilibria, we apply a linear stability analysis, for which we need the Jacobian matrix $J = (\partial p_i'/\partial p_j)$ for all (i,j). This Jacobian has always the eigenvalue 0 because $(1,1,1)$ is a left eigenvector and $\sum_i \frac{\partial p_i'}{\partial p_j} = 0$ because $\sum_i p_i' = 1$. (Note that $(1,1,1)$ is perpendicular to S_3.) Therefore, only the two other eigenvalues are relevant for stability. In general, the Jacobian is rather complicated. However, it is easily computed with *Mathematica*. Because we analyze a difference equation, an equilibrium is linearly stable if all eigenvalues λ of J evaluated at this equilibrium satisfy $|\lambda| < 1$ (see Appendix A.1).

At the equilibrium $\hat{p}^{(13)} = \left(\frac{1-a+s_1}{1-a+2s_1}, 0, \frac{s_1}{1-a+2s_1}\right)$, we obtain the eigenvalues

$$\lambda_1 = 0\,,\ \lambda_2 = \frac{1-a(1-s_1)+s_1}{(1+s_1)^2-a}\,,\ \lambda_3 = \frac{1-a+s_1(2+s_2)}{(1+s_1)^2-a}\,.$$

Therefore, $0 < \lambda_2 < 1$, $\lambda_3 > 0$, and $\lambda_3 < 1$ if and only if $s_2 < s_1$. Consequently, $\hat{p}^{(13)}$ is asymptotically stable if $s_2 < s_1$, and it is a saddle point if $s_2 > s_1$.

At the equilibrium $\hat{p}^{(23)} = \left(0, \frac{1}{2}, \frac{1}{2}\right)$, we obtain the eigenvalues

$$\lambda_1 = 0\,,\ \lambda_2 = \frac{2a}{1+a+s_2}\,,\ \lambda_3 = \frac{2+s_1}{1+a+s_2}\,.$$

Then $0 < \lambda_2 < 1$, $\lambda_3 > 0$, and λ_3 ¡1 if and only if $1+s_1-s_2 < a < 1$, which is possible only if $s_2 > s_1$. Consequently, $\hat{p}^{(23)}$ is asymptotically stable if $s_2 > s_1$ and $1+s_1-s_2 < a < 1$. Otherwise, it is a saddle point.

In summary, all boundary equilibria are unstable if

$$s_2 > s_1 > 0 \ \text{ and } \ 0 \le a < 1 + s_1 - s_2\,.$$

The linear stability analysis of the internal equilibrium $\hat{p}$ is cumbersome. The characteristic polynomial is given by

$$xq(x)/[a^2 - a(2+s_1(2+s_1)) + (1+2s_1-s_2)(1+s_2)]^2$$

where (with some help of *Mathematica*)

$$\begin{aligned} q(x) = {} & (1+s_1-s_2)(1+s_2)(1+2s_1-s_1^3+s_1s_2(2+s_1)-s_2^2) \\ & - a\Big[4-4s_2^2-s_1^4(2+s_2)+s_1^3s_2(5+2s_2)+s_1(1+s_2)(8-(3-s_2)s_2) \\ & \quad + s_1^2(4+s_2(3-s_2(4+s_2)))\Big] \\ & + a^2\Big[6+s_1(6+(2-s_1)s_1)+(1-s_1)s_1s_2-(2-s_1)s_2^2\Big] \end{aligned}$$

$$
\begin{aligned}
&- a^3\Big[4 - s_1(2s_1^2 + s_2 - s_1 s_2)\Big] \\
&+ a^4(1 - s_1) \\
&- x\Big[a^2(2 - s_1) - a(4 + 2s_1^2 + s_1(2 - s_2)) + (2 + 3s_1 - 2s_2)(1 + s_2)\Big] \\
&\quad \cdot \Big[a^2 - a(2 + s_1(2 + s_1)) + (1 + 2s_1 - s_2)(1 + s_2)\Big] \\
&+ x^2\Big[a^2 - a(2 + s_1(2 + s_1)) + (1 + 2s_1 - s_2)(1 + s_2)\Big]^2 .
\end{aligned}
$$

We observe that

$$
q(1) = (1 - a)s_1^2(1 - a + s_2)(1 - a + s_1 - s_2)(s_2 - s_1)
$$

and

$$
\begin{aligned}
\frac{dq}{dx}(x = 1) &= (1 - a)s_1^2(1 - a + s_2) \\
&\quad \cdot \Big[(1 - a + s_1 - s_2)^2(1 + 2s_2) + s_1(2s_2 - s_1 - a s_1)\Big] > 0 .
\end{aligned}
$$

In particular,

$$
q(1) = 0 \ \text{ if } a = 1 + s_1 - s_2 ,
$$

and

$$
q(1) > 0 \ \text{ if } a < 1 + s_1 - s_2 \text{ and } s_2 > s_1 .
$$

Therefore, 1 is an eigenvalues if $a = 1 + s_1 - s_2$, which is in accordance with the bifurcation that we already noticed.

In addition, because (i) $q(x)$ is convex (the coefficient of x^2 is positive), (ii) $\frac{dq}{dx}\big|_{x=1} > 0$, and (iii) $q(1) > 0$ if $a < 1 + s_1 - s_2$ and $s_2 > s_1$, there can be no eigenvalue > 1.

Hence, if there are two positive eigenvalues, the equilibrium $\hat{p}$ is linearly stable. This can indeed be shown, but is even more complicated and not shown here (but see the special case $a = 0$ below). However, we can conclude the following:

If $a < 1 + s_1 - s_2$ and $s_2 > s_1$ is not satisfied, then $q(1) < 0$ and there is an eigenvalue > 1. Therefore, the internal equilibrium is unstable.

If $a = 0$, the analysis is considerably simplified because then $q(x) = (1 + s_2)r(x)$, where

$$
\begin{aligned}
r(x) &= c_0 + c_1 x + c_2 x^2 , \\
c_0 &= (1 - s_1 + s_2)[(1 + s_1)^2 - s_2^2 + s_1^2(s_2 - s_1) + s_1(2s_2 - s_1)] , \\
c_1 &= -(2 + 3s_1 - 2s_2)(1 + 2s_1 - s_2)(1 + s_2) , \\
c_2 &= (1 + s_2)(1 + 2s_1 - s_2)^2 .
\end{aligned}
$$

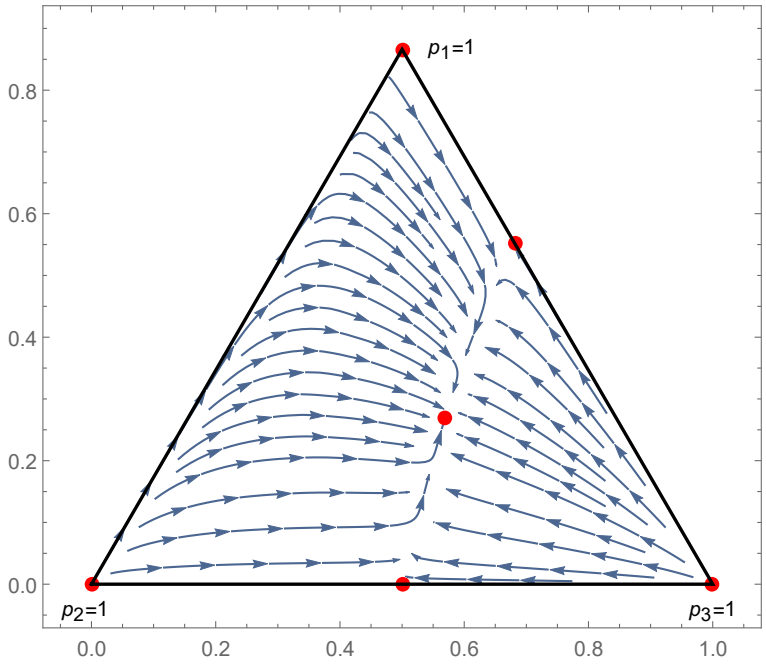

Figure 3.5: Stream plot (generated with *Mathematica*) for the following parameters in W: $a = \frac{1}{2}$, $s_1 = \frac{2}{3}$, $s_2 = 1$.

With $a = 0$, the condition $a < 1 + s_1 - s_2$ and $s_2 > s_1$ simplifies to $0 < s_1 < s_2 < 1 + s_1$.

Therefore, $r(0) = c_0 > 0$ if $0 < s_1 < s_2 < 1 + s_1$. In addition, $q(x)$ assumes a negative value at its minimum $-c_1/(2c_2) \in (0, 1)$ if and only if $c_1^2 > 4c_0c_2$. The latter condition simplifies to

$$s_1^2[1 + 4s_1 - 3s_2 + 4(s_1 - s_2)^2] > 0\,,$$

which is satisfied if $1 + s_1 - s_2 > 0$. Hence, we have proved linear stability of the fully polymorphic equilibrium $\hat{p}$ if $a = 0$ and $0 < s_1 < s_2 < 1 + s_1$.

In summary, the equilibrium $\hat{p}$ is linearly stable if $0 \le a < 1 + s_1 - s_2$ and $s_2 > s_1 > 0$.

Because $\overline{W}$ is a Lyapunov function and $\overline{W}$ is a quadratic polynomial, the equilibrium $\hat{p}$ is globally asymptotically stable in this case, i.e., it attracts all solutions that start in the interior of S_3.

The result about local stability can be deduced with very little calculation either from rather general index theorems and degree theory (e.g., Hofbauer 1990) or from results on multiallelic systems in Nagylaki and Lou (2006). ◁

Remark 3.11. For triallelic systems, Nagylaki and Lou (2006) proved the following results about stability of a fully polymorphic equilibrium.

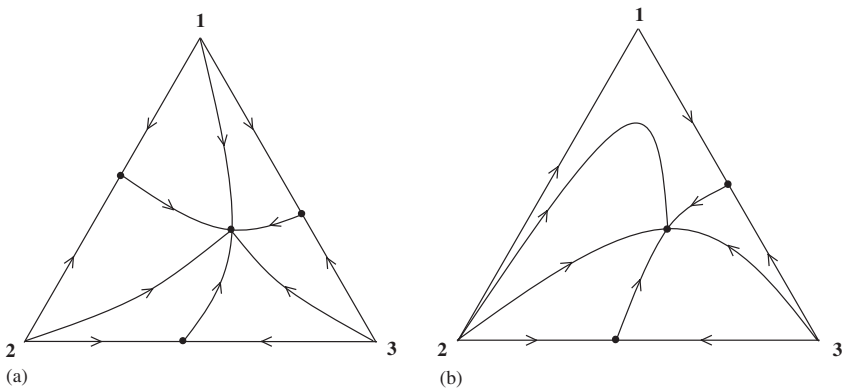

Fig. 3. Phase portraits for three alleles with an isolated, asymptotically stable, internal equilibrium point.

Figure 3.6: This figure illustrates the two possible cases in which a triallelic system can have an asymptotically stable internal equilibrium. In (a) there is pairwise overdominance, and in (b) two pairs of alleles are overdominant and the third pair has intermediate or complete dominance.

- Suppose there is pairwise overdominance. If there exists an isolated internal equilibrium point, it is globally asymptotically stable.
- Suppose there exists an isolated, asymptotically stable, internal equilibrium point. Then there is either

  (a) pairwise overdominance or

  (b) two pairs of alleles are overdominant and the third pair has intermediate or complete dominance.

The corresponding phase portraits are shown in Fig. 3.6.

Of particular evolutionary interest are stable equilibria. To avoid pathologies, we assume $W_{ii} > 0$ for every i. This implies $\overline{W}(\boldsymbol{p}) > 0$ for $\boldsymbol{p} \in S_k$, and invariance of every boundary face. Therefore, alleles which are present initially do not disappear in finite time (but they may disappear as $t \to \infty$). Two consequences of the fact that mean fitness is strictly increasing along solutions, except at equilibrium points, are the following:

Theorem 3.12. *1. An internal equilibrium, i.e., a completely polymorphic equilibrium, is asymptotically stable if and only if it is an isolated local maximum of* $\overline{W}$.

2. Every solution converges to the set of equilibrium points. If there is only a finite number of equilibria, every solution converges to exactly one equilibrium point.

By more sophisticated methods, much stronger results can be derived (cf. Kingman 1961b, Lyubich *et al.* 1980, Losert and Akin 1983, Lyubich 1992). We state some of the most important ones without proof:

Theorem 3.13. *1. An equilibrium point is stable if and only if it is a local, not necessarily isolated, maximum of* $\overline{W}$.

2. If an internal equilibrium exists, it is stable if and only if, counting multiplicities, the fitness matrix W *has exactly one positive eigenvalue.*

3. If an asymptotically stable internal equilibrium exists, then every *solution starting in the interior of* S_k *converges to this equilibrium.*

4. If the matrix W *has* i *positive eigenvalues, at least* $(i-1)$ *alleles will be absent at a stable equilibrium.*

5. Every solution converges to some, uniquely determined, equilibrium point (*even if stable linear manifolds of equilibria exist*).

A fairly detailed treatment of the dynamics and the equilibrium behavior of the discrete-time selection model is contained in Lyubich (1992, Chapter 9). Proofs of the above results as well as of many more results may be found there. Other important references, with additional results, include Karlin (1984) and Nagylaki (1992).

A natural question is about the maximum possible number of coexisting stable equilibria, but this is unsolved except for special cases (Broom *et al.* 1993, Hofbauer and Sigmund 1998, Chapter 19). As we have seen above, there can be at most one asymptotically stable, completely polymorphic equilibrium, which then is globally stable relative to all completely polymorphic initial states. Vickers and Cannings (1988) showed that in a system of k alleles there can be at most two stable equilibria involving $k-1$ alleles. However, they gave an example of a system of four alleles having three stable equilibria, one with three alleles, the others with two alleles. An important result that can be proved is the following (cf. Kingman 1961b, Karlin and Lessard 1984, Vickers and Cannings 1988):

Theorem 3.14. *If* $\hat{\boldsymbol{p}}$ *is a stable equilibrium when there are* $(k-1)$ *alleles and* $\hat{p}_i > 0$ *for every* i*, and if a new allele is introduced which causes the equilibrium to be unstable, then there is a unique stable equilibrium in the enlarged space.*

Therefore, if new alleles are introduced sequentially into a population, such that the time between introductions is sufficient to allow convergence to equilibrium, then the future equilibria are uniquely determined, given the sequence of introduction. This result does not hold if two or more alleles are introduced within a short period. At an externally stable

boundary equilibrium (see Appendix A.2), $\hat{\boldsymbol{p}}$, an allele $\mathcal{A}_k$ not present at this equilibrium will die out if introduced at low frequency. This occurs if and only if $W_k(\hat{\boldsymbol{p}}) \leq \overline{W}(\hat{\boldsymbol{p}})$.

3.6 Discrete- versus continuous-time models

Most higher animal species have overlapping generations because birth and death occur continuously in time. Because a rigorous derivation from biological principles is difficult (see e.g., Nagylaki and Crow 1974), below we derive the continuous-time model as an approximation to the discrete dynamics for the case of weak selection.

In a continuous-time model, the (Malthusian) fitness m_{ij} of a genotype $\mathcal{A}_i\mathcal{A}_j$ is defined as its birth rate minus its death rate. The marginal fitness of allele $\mathcal{A}_i$ is $m_i = \sum_j m_{ij}p_j$ and the mean fitness of the population is $\overline{m} = \sum_i m_i p_i = \sum_{i,j} m_{ij}p_i p_j$. Then the dynamics of allele frequencies is given by[7]

$$\dot{p}_i = p_i(m_i - \overline{m}) \quad \text{for every } i\,. \tag{3.36}$$

This is the analog of the discrete-time selection dynamics (3.5). The equilibria are obtained from the condition $\dot{p}_i = 0$ for every i. Note that the dynamics remains unchanged if the same constant is added to all fitnesses m_{ij}.

The difference equation (3.5) and the differential equation (3.36) have the same equilibria, provided there are constants a and $s > 0$ such that

$$W_{ij} = a + sm_{ij} \qquad \text{for every } i, j\,. \tag{3.37}$$

This is obvious by noting that (3.37) implies $W_i = a + sm_i$ and $\overline{W} = a + s\overline{m}$. As a consequence, the results in the section above on the number of equilibria apply without change to the dynamics (3.36).

For weak selection the discrete-time model (3.5) can be approximated by the continuous model (Nagylaki (1992, p. 99). Assume that W_{ij} is given by (3.37) with $a = 1$. We rescale time according to $t = \lfloor \tau/s \rfloor$, where $\lfloor\ \rfloor$ denotes the closest smaller integer, so that s may be interpreted as generation length, and we set $q_i = q_i(\tau) = p_i(t)$, where p_i satisfies the difference equation (3.5). Then we have

$$\frac{d}{d\tau}q_i = \lim_{s\to 0}\frac{1}{s}\left[q_i(\tau + s) - q_i(\tau)\right] = \lim_{s\to 0}\frac{1}{s}\left[p_i(t+1) - p_i(t)\right]\,.$$

From (3.5) and (3.37) we obtain $p_i(t+1) - p_i(t) = sp_i(t)(m_i - \overline{m})/(1 + s\overline{m})$ and, therefore, $\dot{q}_i = q_i(m_i - \overline{m})$. This proves the assertion because $\Delta p_i \approx s\dot{q}_i = sp_i(m_i - \overline{m})$.

[7] Throughout, we denote the time derivative $\frac{\mathrm{d}x}{\mathrm{d}t}$ of any function $x = x(t)$ by $\dot{x}$.

Remark 3.15. Summarizing, this means that if selection is weak in the sense that terms of order $O(s^2)$ can be neglected, where $s = \max_{i,j} W_{ij} - \min_{i,j} W_{ij}$, then the discrete-time dynamics (3.5) can be approximated by the continuous-time dynamics, i.e.,

$$\Delta p_i \approx \dot{p}_i = p_i(W_i - \overline{W}) \quad \text{for every } i\,. \tag{3.38}$$

Therefore, both dynamical systems have the same equilibria, and even the stability properties of the equilibria are the same, at least if selection is weak. The latter statement follows from the easily derived relation between the Jacobians for both cases and the fact that $\overline{W}$ is a Lyapunov function in both cases (see below). Frequently, the differential equations are easier to analyze.

Remark 3.16. For the differential equation (3.36), it is very easy to show that (Exercise!)

$$\dot{\overline{m}} = 2\sum_i p_i(m_i - \overline{m})^2 = \sigma_{\mathrm{A}}^2\,. \tag{3.39}$$

In the following remark, which presents optional material, we show that the continuous-time selection dynamics is a so-called *generalized gradient system*. Such dynamical systems have simple dynamic behavior, e.g., periodic orbits or chaos cannot occur (see Appendix A.3).

Remark* 3.17. To show that (3.36) is a generalized gradient system on the simplex S_k, we define the indicator variables

$$f_i(k,l) = \begin{cases} 1\,, & \text{if } k = l = i\,, \\ \frac{1}{2}\,, & \text{if } k \neq l, \text{ and } k = i \text{ or } l = i\,, \\ 0\,, & \text{otherwise}\,, \end{cases} \tag{3.40}$$

that associate to the genotype $\mathcal{A}_k\mathcal{A}_l$ the values 1, $\frac{1}{2}$, or 0, depending on the number of $\mathcal{A}_i$ alleles this genotype contains. Then the expectation of f_i with respect to the frequency distribution of genotypes is

$$\bar{f}_i = \sum_{k,l} f_i(k,l) P_{kl} = p_i\,, \tag{3.41}$$

and the covariance of f_i and f_j, denoted by $g^{ij} = \mathrm{Cov}[f_i, f_j]$, is calculated to be

$$g^{ij} = \tfrac{1}{2} p_i(\delta_{ij} - p_j)\,. \tag{3.42}$$

Here, δ_{ij} is the Kronecker delta defined by $\delta_{ij} = 1$, if $i = j$, and $\delta_{ij} = 0$, otherwise. We shall denote the (symmetric, positive definite) matrix formed by the entries g^{ij} by $\mathsf{G}_{\boldsymbol{p}}$. With this notation and the simple identity $\frac{\partial \overline{m}}{\partial p_j} = 2m_j$, the selection dynamics (3.36) can be rewritten in the form of a generalized gradient system, i.e.,

$$\dot{p}_i = \sum_{j=1}^{k} g^{ij} \frac{\partial \overline{m}}{\partial p_j}\,, \tag{3.43a}$$

or

$$\dot{\boldsymbol{p}} = \tfrac{1}{2}\mathsf{G}_{\boldsymbol{p}} \operatorname{grad} \overline{m}\,, \tag{3.43b}$$

where $\operatorname{grad} \overline{m} = (\frac{\partial \overline{m}}{\partial p_1}, \ldots, \frac{\partial \overline{m}}{\partial p_k})$ denotes the gradient operator. We shall call gradient systems on S_k of this form *Svirezhev–Shahshahani gradients*, because they are gradient systems with respect to the metric introduced by Svirezhev (1972); cf. also Shahshahani (1979) and Svirezhev and Passekov (1990).

Because the matrix $\mathsf{G}_{\boldsymbol{p}}$ is positive definite, it follows immediately from (3.43b) that mean fitness is nondecreasing:

$$\dot{\overline{m}} = (\operatorname{grad} \overline{m})^{\top} \dot{\boldsymbol{p}} = (\operatorname{grad} \overline{m})^{\top} \mathsf{G}_{\boldsymbol{p}} \operatorname{grad} \overline{m} \geq 0\,. \tag{3.44}$$

A simple calculation reveals that (3.43a) can be expressed as

$$\dot{p}_i = \operatorname{Cov}[f_i, m]\,, \tag{3.45}$$

where m is the random variable $m(\mathcal{A}_k\mathcal{A}_l) = m_{kl}$. This is the covariance formula of Li (1967) and Price (1970). It holds under much more general assumptions (see Lessard 1997), and is a special case of Robertson's Secondary Theorem of Natural Selection (see Bürger 2000, Chaps. II.5, II.6).

Generalized gradient systems have properties analogous to (usual) gradient systems. The main difference is that they form gradients with respect to a Riemannian (but non-Euclidean) metric. Using the inverse of the matrix $\mathsf{G}_{\boldsymbol{p}}$, such a metric can be defined on the simplex. It takes into account that small changes in gene frequencies near the boundary of the simplex have much more important consequences for the dynamics than equally small changes in the interior. For details, see Bürger (2000, Appendix A.3). [$*$]

In the case of two alleles, it is convenient to write $p = p_1$ and $q = p_2 = 1 - p$. Then (3.43a) becomes

$$\dot{p} = \frac{pq}{2}\left(\frac{\partial \overline{m}}{\partial p} - \frac{\partial \overline{m}}{\partial q}\right)\,, \tag{3.46a}$$

or after substituting $q = 1 - p$ in $\overline{m}$,

$$\dot{p} = \frac{p(1-p)}{2}\frac{\mathrm{d}\overline{m}}{\mathrm{d}p}\,. \tag{3.46b}$$

This representation of gene-frequency change goes back to Wright (1935), and is equivalent to (3.12b) if $\overline{m}$ is replaced by $\ln \overline{W}$ and $\dot{p}$ by $\overline{W}\Delta p$.

If we assign the (relative) fitnesses 0, $-hs$, and $-s$ to the genotypes $\mathcal{A}_1\mathcal{A}_1$, $A_1\mathcal{A}_2$, and $\mathcal{A}_2\mathcal{A}_2$, (3.46b) simplifies to

$$\dot{p} = sp(1-p)^2 \quad \text{if } h = 0 \text{ ($\mathcal{A}_1$ is dominant)}\,, \tag{3.47}$$

$$\dot{p} = \tfrac{1}{2} sp(1-p) \quad \text{if } h = \tfrac{1}{2} \text{ (no dominance)}\,, \tag{3.48}$$

$$\dot{p} = sp^2(1-p) \quad \text{if } h = 1 \text{ ($\mathcal{A}_1$ is recessive)}\,. \tag{3.49}$$

Note that (3.48) can be solved explicitly.

It was Wright who proposed and advocated the idea of an *adaptive topography* in which evolution takes place like a hill-climbing process, and mean fitness plays the role of a potential-like function.

4 Mutation and selection

Natural selection and mutation are two central factors guiding biological evolution: mutation generates the genetic variability upon which selection can act. This was clearly recognized by the pioneers of population genetics, Fisher, Haldane, and Wright, who developed mathematical models quantifying the relative importance of selection and mutation in maintaining genetic variation. In traditional models, two alleles per locus, the wild type and a mutant, are considered, and the equilibrium frequencies of the alleles can be calculated under recurrent mutation and various assumptions on the selective values of the genotypes. In many instances, however, more than two alleles per locus may occur.

Prior to 1970, only few general results about mutation-selection models with multiple alleles per locus were available (see Crow and Kimura 1970). Moran (1976) demonstrated existence, uniqueness, and global stability of an equilibrium for a haploid mutation-selection model. Previously, Thompson and McBride (1974) had derived the solution of a system of differential equations occurring in the theory of the evolution of macromolecules that is formally equivalent to the haploid mutation-selection equation.

This section is devoted to the basic properties of classical models incorporating mutation and selection at a single gene locus. By classical models, we mean that only a finite number of alleles (mutant types) per locus is considered. For some purposes, models with an infinite number of alleles, even with a continuum of alleles, have been proved useful. But they are beyond the scope of this introduction. Throughout, populations are assumed to be sufficiently large that random genetic drift can be ignored, to mate at random (if sexual), and to have constant genotypic fitnesses.

4.1 Only mutation

We shall employ a simple concept of mutation, sufficient for most purposes in population genetics theory, by designating any change from one allelic type to another a mutation. Here, we assume that all mutations are neutral, i.e., all have the same fitness. Let us consider k alleles, $\mathcal{A}_1, \ldots, \mathcal{A}_k$, at a gene locus and label their frequencies by $p_1, \ldots, p_k$.

For $i \neq j$ we denote the probability that an $\mathcal{A}_i$ gene has an $\mathcal{A}_j$ offspring by the mutation rate μ_{ij}. We shall use the convention $\mu_{ii} = 0$ for every i. Then the fraction of $\mathcal{A}_i$ genes that do not mutate is $1 - \sum_j \mu_{ij}$, and $\mathcal{A}_j$ genes give rise to a mutant $\mathcal{A}_i$ with probability μ_{ji}. Therefore, the frequency p_i' of $\mathcal{A}_i$ in the next generation is

$$p_i' = p_i\left(1 - \sum_j \mu_{ij}\right) + \sum_j p_j \mu_{ji} \text{ for every } i\,. \tag{4.1}$$

We call this linear recursion the (*pure*) *mutation equation*. Due to the convention $\mu_{ii} = 0$, the index i may or may not be excluded in the above summations.

Linear algebra shows that there exists a unique equilibrium if all mutation rates are positive. In addition, convergence to this equilibrium occurs at a geometric rate, as will be proved below in much greater generality.

Example 4.1. Let us illustrate this for the simple case of two alleles. Denoting the mutation rate from $\mathcal{A}_1$ to $\mathcal{A}_2$ by μ, the reverse mutation rate by ν, and the frequency of $\mathcal{A}_1$ by p, the recursion (4.1) reduces to

$$p' = p(1 - \mu - \nu) + \nu\,.$$

If μ or ν is positive, there exists a unique equilibrium frequency (obtained from the condition $p' = p$). It is given by

$$\hat{p} = \frac{\nu}{\mu + \nu}\,.$$

The above recursion equation can be solved explicitly and, using $\hat{p}$, its solution can be expressed as

$$p(t) - \hat{p} = (p_0 - \hat{p})(1 - \mu - \nu)^t\,,$$

where $p_0 = p(0)$ is the initial frequency of $\mathcal{A}_1$. This shows that convergence to equilibrium occurs at a geometric rate, but is very slow because $\mu + \nu$ is typically very small.

4.2 Dynamics in haploid populations

We consider a haploid, asexually reproducing population, in which k types (alleles), labeled as $\mathcal{A}_1, \ldots, \mathcal{A}_k$, may occur. Let the fitness of $\mathcal{A}_i$ be W_i, let its relative frequency in generation t be $p_i = p_i(t)$, so that $\sum_i p_i = 1$, and let $\boldsymbol{p} = (p_1, \ldots, p_k)^\top$. As previously, allele frequencies in successive generations are denoted by p_i and p_i'. Frequencies are measured in offspring before selection. Thus, the life cycle begins with selection, which is followed by reproduction during which mutation occurs. We designate the probability that an $\mathcal{A}_i$ individual has an $\mathcal{A}_j$ offspring (where $j \neq i$) by the mutation rate μ_{ij}, and use the convention $\mu_{ii} = 0$ for every i.

After selection, the frequency of $\mathcal{A}_i$ is $p_i^* = p_i W_i/\overline{W}$. Then reproduction and mutation occur, and change the allele frequencies p_i^* according to (4.1). Therefore, substituting p_i^* for p_i in (4.1), we obtain the *mutation-selection equation*

$$p_i' = p_i \frac{W_i}{\overline{W}} + \frac{1}{\overline{W}} \sum_j (p_j W_j \mu_{ji} - p_i W_i \mu_{ij}) \,. \tag{4.2}$$

For our purposes, it is convenient to cast (4.2) into matrix form. Let us define the $k \times k$ mutation matrix $\tilde{\mathsf{U}} = (\tilde{u}_{ij})$ by

$$\tilde{u}_{ij} = \begin{cases} 1 - \sum_l \mu_{il} \,, & i = j \,, \\ \mu_{ji} \,, & i \neq j \,, \end{cases} \tag{4.3}$$

and the mutation-selection matrix $\mathsf{C} = (c_{ij})$ by

$$c_{ij} = \tilde{u}_{ij} W_j \,. \tag{4.4}$$

Then a simple calculation shows that

$$\bar{c} = \sum_i (\mathsf{C}\boldsymbol{p})_i = \overline{W} \,, \tag{4.5}$$

and (4.2) can be rewritten as

$$\boldsymbol{p}' = \frac{1}{\bar{c}} \mathsf{C}\boldsymbol{p} \,. \tag{4.6}$$

The state space is the simplex S_k. Observing that $\boldsymbol{n}(t) = \mathsf{C}^t \boldsymbol{n}(0)$ is the solution of $\boldsymbol{n}' = \mathsf{C}\boldsymbol{n}$, and $\boldsymbol{p}(t) = \boldsymbol{n}(t)/\sum_i n_i(t)$, it follows immediately that (4.2) has the explicit solution

$$p(t) = \mathsf{C}^t \boldsymbol{p}_0 \Big/ \sum_i (\mathsf{C}^t \boldsymbol{p}_0)_i \,, \tag{4.7}$$

where $\boldsymbol{p}(0) = \boldsymbol{p}_0 \in S_k$ designates the initial frequency distribution.

It is our aim to show that all solutions $\boldsymbol{p}(t)$ of (4.2), or (4.6), starting from an arbitrary initial value in S_k, converge to a unique equilibrium distribution $\hat{\boldsymbol{p}}$ that is completely polymorphic, i.e., $\hat{p}_i > 0$ for every i. Some positivity condition on the mutation rates and the fitnesses will be required to achieve such a result, because otherwise equilibria may exist with $\hat{p}_i = 0$ for one or several i.

Actually, more complicated behavior can occur, as the following example shows. Consider two alleles, $\mathcal{A}_1$ and $\mathcal{A}_2$, and assume $W_1 = W_2 = 1$ and $\mu_{12} = \mu_{21} = 1$. Inserting these values into (4.2) gives $p_1' = 1 - p_1$ and, furthermore, $p_1'' = p_1$. Thus, periodic solutions exist in this simple model. The appropriate condition for achieving uniqueness and stability is that a positive integer n exists so that through a series of n steps, every allele $\mathcal{A}_i$ gives

rise to descendants of every type $\mathcal{A}_j$ with positive probability. Mathematically, this means that for some integer $n \geq 1$ all entries of the matrix C^n must be positive (> 0). Such a matrix is called *primitive*. A matrix C^n can be positive[8] only if $W_i > 0$ for every i. In this case, C is primitive if and only if $\tilde{\mathsf{U}}$ is primitive. The following result was proved by Moran (1976):

Theorem 4.2. *If the matrix* C *defined in* (4.4) *is primitive, then the mutation-selection dynamics* (4.6) *admits a unique equilibrium,* $\hat{\boldsymbol{p}}$*, that satisfies* $\hat{p}_i > 0$ *for every* i*. This equilibrium is the unique solution of*

$$\hat{\overline{W}}\,\hat{\boldsymbol{p}} = \mathsf{C}\hat{\boldsymbol{p}}\,, \tag{4.8}$$

where $\hat{\overline{W}} = \sum_i W_i \hat{p}_i$ *is the equilibrium mean fitness, and it is globally asymptotically stable.*

Proof. By the Perron–Frobenius Theorem (see Appendix B), the primitive matrix C has an eigenvalue $r > 0$ with a positive eigenvector $\hat{\boldsymbol{q}}$, i.e., $\mathsf{C}\hat{\boldsymbol{q}} = r\hat{\boldsymbol{q}}$, and $\hat{\boldsymbol{q}}$ is unique except for multiplication by positive constants. In addition, r is the unique eigenvalue with this property. Assuming that $\hat{\boldsymbol{q}}$ is normalized, $\sum_i \hat{q}_i = 1$, we observe from this eigenvalue equation that $\hat{\bar{c}} = \sum_i (\mathsf{C}\hat{\boldsymbol{q}})_i = r$. From (4.5) we infer that the equilibrium mean fitness is $\hat{\overline{W}} = r$. Since the equilibrium solution $\hat{\boldsymbol{p}}$ must satisfy (4.8), and since $\hat{\boldsymbol{q}}$ is unique, we conclude that $\hat{\boldsymbol{p}} = \hat{\boldsymbol{q}}$, which proves the assertion about existence, uniqueness, and positivity of $\hat{\boldsymbol{p}}$. Global stability also follows from Perron–Frobenius theory, because for any initial vector $\boldsymbol{q}_0$, $r^{-t}\mathsf{C}^t\boldsymbol{q}_0$ converges to some positive multiple of $\hat{\boldsymbol{p}}$. This observation, together with the fact that a solution $\boldsymbol{p}(t)$ of (4.6) is normalized and given by (4.7), shows that any solution of (4.6), and hence of (4.2), converges to $\hat{\boldsymbol{p}}$. □

4.3 Dynamics in diploid populations

For diploid populations subject to mutation and selection, equilibria are not necessarily uniquely determined. Even periodic solutions and stable limit cycles may occur. Of particular interest are, therefore, conditions on the mutation and fitness parameters that ensure existence, uniqueness, and stability of polymorphic equilibria.

4.3.1 The mutation-selection equations

Let μ_{ij} denote the mutation rate from $\mathcal{A}_i$ to $\mathcal{A}_j$ and $\mu_{ii} = 0$. Allele frequencies are measured among zygotes before selection, and the life cycle begins with selection, which is followed by the production of germ cells, during which mutation occurs, and the formation of zygotes. The same argument that led to the haploid mutation-selection equation (4.2),

[8] Throughout, positive matrix or positive vector means that all components are > 0; nonnegative means ≥ 0. See also Appendix B.

but now by substituting (3.5) into (4.1), yields the diploid mutation-selection dynamics for a population with discrete generations,

$$p_i' = p_i \frac{W_i}{\overline{W}} + \frac{1}{\overline{W}} \sum_j (p_j W_j \mu_{ji} - p_i W_i \mu_{ij}) \text{ for every } i\,, \tag{4.9}$$

where W_i is the marginal fitness of the allele $\mathcal{A}_i$.

For a population with overlapping generations, we shall use the differential equation

$$\dot{p}_i = p_i(m_i - \overline{m}) + \sum_j (p_j \mu_{ji} - p_i \mu_{ij}) \text{ for every } i\,, \tag{4.10}$$

where m_i is the marginal Malthusian fitness of $\mathcal{A}_i$. Both models are classical; for continuous-time, see Crow and Kimura (1970, Chapter 6.4); for discrete time, see Moran (1976). For a derivation of the continuous-time model (4.10) consult Nagylaki (1974). It is also readily obtained as the weak-selection weak-mutation limit of (4.9).

Obviously, the dynamics of (4.9) remains unchanged if all fitnesses W_{ij} are multiplied by the same positive constant, and (4.10) remains unchanged if the same constant is added to every m_{ij}. For multiplicative fitnesses, $W_{ij} = W_i W_j$, the discrete-time recursion (4.9) reduces to the haploid recursion (4.2), and for additive fitnesses, $m_{ij} = m_i + m_j$, the continuous-time equation (4.10) reduces to the haploid equation.

Remark 4.3. For the haploid version of (4.10) or, equivalently, for additive fitnesses, an analog of Theorem 4.2 is valid. Define the matrix $\mathsf{A} = (a_{ij})$ by

$$a_{ij} = (m_i - \sum_\ell \mu_{i\ell})\delta_{ij} + \mu_{ji}\,, \tag{4.11}$$

and let $\bar{a} = \sum_i (\mathsf{A}\boldsymbol{p})_i = \sum_{i,j} a_{ij} p_i p_j$. Then (4.10) can be written as

$$\dot{\boldsymbol{p}} = \mathsf{A}\boldsymbol{p} - \bar{a}\boldsymbol{p}\,, \tag{4.12}$$

which has the solutions

$$\boldsymbol{p}(t) = \frac{e^{\mathsf{A}t}\boldsymbol{p}_0}{\sum_i (e^{\mathsf{A}t}\boldsymbol{p}_0)_i}\,. \tag{4.13}$$

One obtains global convergence to the unique equilibrium solution if A is irreducible (which is a slightly weaker assumption than primitivity).

4.3.2 The case of two alleles

For the present purpose it is convenient to parameterize the fitness values of the genotypes $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, $\mathcal{A}_2\mathcal{A}_2$ as $W_{11} = 1$, $W_{12} = 1 - hs$, $W_{22} = 1 - s$. Instead of p_1 and p_2, we write p and $q = 1 - p$. Then the marginal fitnesses and the mean fitness are given by

(3.10) and (3.11), respectively. For the mutation rates we write $\mu = \mu_{12}$ and $\nu = \mu_{21}$ and require $\mu + \nu < 1$. A straightforward calculation shows that the equilibria of the mutation-selection equation (4.9) are the solutions p of the cubic equation

$$\begin{aligned} p^3 s(2h-1) + p^2 s[2 - 3h + \mu h + \nu(1-h)] \\ + p[-s(1-h) + \mu(1-hs) + \nu(1-2s+hs)] - \nu(1-s) = 0 \end{aligned} \tag{4.14}$$

in the interval $[0,1]$. Depending on the parameters, it may have one, two, or three such solutions. Some elementary, but lengthy, algebra shows the following (Norman 1974, Nagylaki 1992):

Theorem 4.4. *If $0 < s < 1$ and $h \le \frac{1}{2}$, or $s < 0$ and $h \ge \frac{1}{2}$, then (4.14) has a unique solution in $[0,1]$. Because $\mu + \nu < 1$, this equilibrium is globally asymptotically stable. Convergence is monotonic.*

This result includes a number of interesting special cases, such as no dominance ($h = \frac{1}{2}$), complete dominance of $\mathcal{A}_1$ ($h = 0$), and overdominance ($h < 0$) (in all these cases $s > 0$ is assumed).

The equilibrium solutions of (4.14) are simple only in special cases. We restrict our attention to the case $\nu = 0$, in which back mutations from the deleterious (and thus rare) allele $\mathcal{A}_2$ to $\mathcal{A}_1$ are ignored. It will be convenient to give the precise formulas in terms of $q = 1 - p$. Obviously, $\hat{q}^{(0)} = 1$ is always an equilibrium, because if $\mathcal{A}_1$ is initially not present in the population, it will not arise by mutation. With $\nu = 0$, (4.14) reduces to a quadratic equation which, if $\mu/s \le \frac{1}{4}$, has the following solutions in $[0,1]$:

$$\hat{q}^{(1)} = \frac{h(1+\mu)}{2(2h-1)}\left[1 - \sqrt{1 - \frac{4\mu(2h-1)}{(1+\mu)^2 h^2 s}}\right] \qquad \text{if } h \notin \{0, \tfrac{1}{2}\}, \tag{4.15a}$$

$$\hat{q}^{(1)} = \sqrt{\frac{\mu}{s}} \qquad \text{if } h = 0, \tag{4.15b}$$

$$\hat{q}^{(1)} = \frac{2\mu}{s(1+\mu)} \qquad \text{if } h = \tfrac{1}{2}, \tag{4.15c}$$

and

$$\hat{q}^{(2)} = \frac{h(1+\mu)}{2(2h-1)}\left[1 + \sqrt{1 - \frac{4\mu(2h-1)}{(1+\mu)^2 h^2 s}}\right] \qquad \text{if } h > h_c, \tag{4.16}$$

where

$$h_c = \frac{1 - \mu/s}{1 - \mu} < 1. \tag{4.17}$$

If $h < h_c$, then $\hat{q}^{(2)}$ is biologically not meaningful because $\hat{q}^{(2)} > 1$. If this holds, then the equilibrium $\hat{q}^{(1)}$ is globally asymptotically stable. A further special case of this are

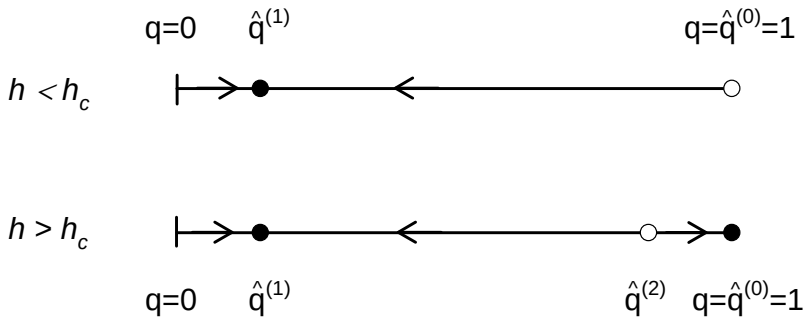

Figure 4.1: Equilibria and dynamics for the diallelic mutation-selection equation with one-way mutation. Stable equilibria are indicated by •, unstable ones by ∘.

multiplicative fitnesses: If $W_{11} = 1$, $W_{12} = 1 - t$, $W_{22} = (1 - t)^2$, one obtains (exactly)

$$\hat{q}^{(1)} = \frac{\mu}{t}\,. \tag{4.18}$$

The case $h > h_c$ includes recessiveness of A_1 ($h = 1$) and underdominance ($h > 1$). If $h > h_c$, then three equilibria coexist. They satisfy $0 < \hat{q}^{(1)} < \hat{q}^{(2)} < \hat{q}^{(0)} = 1$, and $\hat{q}^{(1)}$ and $\hat{q}^{(0)}$ are asymptotically stable, whereas $\hat{q}^{(2)}$ is unstable (see Figure 4.1).

Thus, for one-way mutation, a simple and explicit classification of the stability of equilibria is available. It can be shown that this is also valid for the case $0 < \nu \ll \mu$. Then, of course, $\hat{q}^{(0)} < 1$ and $\hat{q}^{(0)} \approx 1$. Analogous results hold for the differential equation (4.10).

We point out that for $h_c \leq h \leq 1$ the pure selection model has one globally asymptotically stable boundary equilibrium ($\hat{q} = 0$), but the introduction of mutation, however weak, leads to two stable and one unstable equilibria. Thus, already with two alleles, the diploid mutation-selection dynamics may be qualitatively different from the haploid dynamics.

The following simple approximations provide better intuition than the above exact results.

Corollary 4.5 (Approximations for mutation-selection balance)**.** *Assume that μ is of smaller order than s, i.e., such that terms of order $(\mu/s)^2$ can be neglected. Then the following approximations for the equilibrium frequency $\hat{q} = 1 - \hat{p}$ of the deleterious mutant A_2 hold.*

1. If $h \gg \sqrt{\mu/s}$, then

$$\hat{q} \approx \frac{\mu}{hs}\,. \tag{4.19a}$$

2. If $0 \le h \ll \sqrt{\mu/s}$, then

$$\hat{q} \approx \sqrt{\frac{\mu}{s}}\,. \tag{4.19b}$$

Importantly, the equilibrium frequency of a recessive deleterious mutant ($h \approx 0$) is much higher than that of an intermediate or dominant deleterious mutant ($h \gg \sqrt{\mu/s}$), because it occurs mostly in heterozygotes, against which selection is ineffective.

If $h > h_c$, then the unstable internal equilibrium is admissible and satisfies

$$\hat{q}^{(2)} \approx \frac{h}{2h-1} - \frac{\mu}{hs}\,. \tag{4.20}$$

The case of weak mutation can also be treated directly by applying perturbation arguments to the pure selection equation. If $s > 0$ and $h < 1$, then introduction of sufficiently weak mutation, i.e., such that $h < (1-\mu/s)/(1-\mu)$, cf. (4.17), leads to only very small disturbances of the equilibria which, in particular, maintain their stability properties. Then a stable equilibrium at the boundary will move inwards, and only the situation displayed in the upper drawing of Figure 4.1 can occur. The case $h = 1$ is nonregular and this simple perturbation argument does not apply. Generally, the case of weak mutation is biologically the most relevant.

Without proof (which is straightforward but tedious), we note that if backward mutation is included ($\nu > 0$) and if it is assumed that $\nu = O(\mu)$, then the approximations in Corollary 4.5 remain valid, i.e., the influence of ν is of smaller order than that of μ and can be ignored.

4.4 Some results about multiallelic models

With more than two alleles at a locus, the mutation-selection dynamics can be much more complicated than with selection alone because it has been proved that stable limit cycles may occur (see below). There are, however, also conditions that imply a simple equilibrium structure and a 'nice' dynamics.

For instance, Theorem 4.2 for haploids implies that in diploids with multiplicative fitnesses, global convergence to a fully polymorphic equilibrium occurs under the assumptions stated there. Next, we describe an other simple situation.

Following Kingman (1977, 1978), we assume that the mutation rates satisfy

$$\mu_{ij} = \mu_j \quad \text{for all pairs } j \neq i\,, \tag{4.21}$$

where the μ_j are constants. This means that the mutation rates depend only on the target genes. Kingman called this the *house-of-cards mutation model* because such mutations

to random types can bring down the house of cards built by evolution (a well adapted allele A_i may mutate to a very different allele, possibly with much lower fitness, and the parental nucleotide sequence has no influence on the result). In fact, it turns out that this is a reasonable approximation in situations, in which most mutants are deleterious and the variance of the distribution of mutational effects is large compared with the existing variance in a well adapted population. An important special case of the house-of-cards mutation model is obtained if all mutation rates are equal, i.e., if $\mu_{ij} = \mu$ for every $i \neq j$. Besides, (4.21) is automatically satisfied in the case of two alleles per locus.

It can be proved that with house-of-cards mutation, all solutions of the mutation-selection equation, both in discrete as well as in continuous time, converge to the set of equilibria because there exists a Lyapunov function (Passekov 1978, Hofbauer 1985). For the simpler differential equation, this will be proved in the exercises.

This result cannot be extended to more general mutation rates. Akin (1979) proved that for every mutation matrix (μ_{ij}) not satisfying (4.21), a fitness matrix (m_{ij}) can be chosen such that periodic orbits occur in the continuous-time dynamics. (This is also true for discrete time.) Hofbauer (1985) and Baake and Wiehe (1997) provided explicit classes of examples for which stable limit cycles occur (already in three-allele systems).

4.5 Applications and outlook

The models treated above are very simplistic. Therefore, many variants and extensions have been introduced and studied. Nevertheless, these simple models form the basis for many applications and important further developments.

4.5.1 Mutation load and Haldane's principle

The majority of mutations is (slightly) deleterious. Therefore, in general, mutations decrease the mean fitness of a population. The investigation of the influence of deleterious mutations on the mean fitness of a population is of considerable biological interest. In a sufficiently large population, deleterious mutants are kept at low frequency by selection, and their main effect is to reduce the mean fitness. In small populations, however, they exert additional effects, because they may become fixed by random genetic drift, and they may lead to a process, known as Muller's ratchet, during which deleterious mutants accumulate on chromosomes. Chromosomes with zero or few deleterious mutations will be lost by random genetic drift and, if there is little or no recombination (as in asexuals), chromosomes with fewer mutations will not be re-established. Muller's ratchet results in a continued decrease of mean fitness, and may lead to eventual extinction of a population. The protection of a population against a steady decrease of mean fitness caused by the permanent influx of deleterious mutations may be an important cause for the evolutionary

significance of sex and recombination (for recent reviews, see Hartfield and Keightley 2012, Sharp and Otto 2016).

Recurrent deleterious mutations impose a 'load' on a population by reducing their mean fitness below the maximum possible level. The genetic load is usually defined as the proportion by which the fitness of the average genotype in a population is reduced in comparison with the best genotype, i.e.,

$$L = \frac{W_{\max} - \overline{W}}{W_{\max}} \,. \tag{4.22}$$

The concept of genetic load has often been criticized on the grounds that a best genotype does not exist in a real population, because fitnesses change in time and may be frequency dependent; or if recombination is taken into account and the best type exists in theory, it is extremely unlikely to ever be produced. Of course, the idealizations on which every model is based have to be kept in mind when applying it. We use the term genetic load in a technical sense as a synonym for the reduction in mean fitness, caused by one or several genetic mechanisms, relative to the idealized situation when these mechanisms are absent or neglected.

The difference between the fitness of the best type and the equilibrium mean fitness at mutation-selection balance is called the mutation load. In the simplest case, when $\mathcal{A}_1$ is the wild type and $\mathcal{A}_2$ is a deleterious mutant that occurs at rate μ, the mutation load is (to leading order in μ)

$$L = 2\mu \,. \tag{4.23}$$

This result is exact if there is no back mutation (as is easily shown from our above theory). If the mutant is (nearly) recessive, a different result holds:

$$L \approx \mu \,. \tag{4.24}$$

The fact that the mutation load depends only on the mutation rate to deleterious alleles, but not on the strength of selection, is very remarkable and called Haldane's principle. Haldane (1937) discovered for the case of one-way mutation that

"... the loss of fitness to the species depends entirely on the mutation rate and not at all on the effect of the gene upon fitness of the individual carrying it ...."

Importantly, Haldane's principle also holds to leading order in μ if there are many alleles and the mutation pattern is arbitrary (subject to mild conditions); see Bürger and Hofbauer (1994).

4.5.2 Maintenance of genetic variation

A fundamental problem in evolutionary biology that has already plagued Darwin consists in answering the question: "What are the mechanisms that maintain genetic variability?".

Several mechanisms that can potentially contribute genetic variability under stabilizing selection have been proposed and investigated. In principle, variability can be maintained by either forces acting directly on the considered trait or as a side effect of genetic variation that is independent of the observed character. Among the direct mechanisms are overdominance, migration, mutation, frequency-dependent selection, fluctuating environments, genotype-environment interaction, or epistasis. The ultimate source of genetic variability is mutation; but can a balance between mutation and stabilizing selection account for a significant fraction of the observed levels of variability? This simple and appealing hypothesis was promoted by Lande (1975) on the basis of a mathematical analysis and a review of empirical data. Since then it has been the object of intense research and debate.

Analyses of such models are based on (quite far-reaching) extensions of the simple models discussed above because selection acts on phenotypic characters which frequently are determined by many genes (loci). Selection on phenotypic characters is often stabilizing, i.e., it selects for an intermediate optimum phenotype and, thus, erodes genetic variance. Therefore, multilocus models are needed and also more refined models about mutation. These go far beyond the scope of this introduction. Although much insight has been provided by mathematical models and empirical results, plenty of unresolved theoretical and empirical questions remain about the genetic architecture underlying quantitative traits, the strength and form of selection, and the relative importance of the genetic, evolutionary, and ecological mechanisms contributing to phenotypic, especially quantitative genetic, variation (for comprehensive treatments, see e.g. Bürger 2000, and Walsh and Lynch 2018).

5 Recombination

According to the Hardy–Weinberg Law, the genotype frequencies attain an equilibrium value after one generation of random mating if gene loci are considered separately. This is no longer true for genotypes with respect to two or more loci considered jointly. Consider two loci, $\mathcal{A}$ and $\mathcal{B}$, each with two alleles, $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{B}_1$, $\mathcal{B}_2$. Then there are ten possible genotypes. If, for instance, in the initial generation only the genotypes $\mathcal{A}_1\mathcal{B}_1/\mathcal{A}_1\mathcal{B}_1$ and $\mathcal{A}_2\mathcal{B}_2/\mathcal{A}_2\mathcal{B}_2$ are present, then in the next generation only these double homozygotes, as well as the two double heterozygotes $\mathcal{A}_1\mathcal{B}_1/\mathcal{A}_2\mathcal{B}_2$ and $\mathcal{A}_1\mathcal{B}_2/\mathcal{A}_2\mathcal{B}_1$ will be present. After further generations of random mating, all other genotypes will occur, but not immediately at their equilibrium frequencies. Of course, the formation of gametic types other than $\mathcal{A}_1\mathcal{B}_1$ or $\mathcal{A}_2\mathcal{B}_2$ requires that recombination between the two loci occur.

Recombination is a process by which pieces of DNA are broken and recombined to produce new combinations of alleles. In eukaryotic cells (i.e., cells with nucleus), recombi-

nation typically occurs during meiosis, which is the process by which (haploid) gametes are produced from diploid parental cells. During meiosis, the homologous pairs of maternal and paternal chromosomes align. During the alignment, the arms of the chromosomes can overlap and temporarily fuse, causing a crossover. Crossovers result in recombination and the exchange of genetic material between the maternal and paternal chromosomes.[9] Alleles at loci on different chromosomes will be combined with probability $\frac{1}{2}$ by Mendel's principle of independent assortment because each member of a pair of homologous chromosomes is transmitted to daughter cells with the same probability.

For a rigorous treatment, we consider more generally two loci, each with an arbitrary number of alleles. Let the frequencies of the alleles $\mathcal{A}_i$ at the $\mathcal{A}$ locus be denoted by p_i and those of the alleles $\mathcal{B}_j$ at the $\mathcal{B}$ locus by q_j. Let P_{ij} be the *frequency of the gamete* $\mathcal{A}_i\mathcal{B}_j$. Then $p_i = \sum_j P_{ij}$ and $q_j = \sum_i P_{ij}$. In general, these allele frequencies are no longer sufficient to describe the genetic composition of the population. *Linkage equilibrium* is defined as the state in which

$$P_{ij} = p_i q_j \tag{5.1}$$

holds for every i and j. Otherwise the population is said to be in *linkage disequilibrium* or, more precisely, gametic phase disequilibrium. It is equivalent to statistical dependence of allele frequencies among loci.

Let the parameter r denote the *recombination frequency* between the two loci, often called the *recombination rate*. This is not quite correct because (usually) r is the *probability* that a recombination event (crossing over) occurs between them. The value of r depends on the distance between the two loci along the chromosome. Loci with $r = 0$ are called completely linked (and may be treated as a single locus) and loci with $r = \frac{1}{2}$ are called unlinked. The maximum value of $r = \frac{1}{2}$ typically occurs for loci on different chromosomes, because then all four gametes are produced with equal frequency $\frac{1}{4}$. Thus, the recombination rate satisfies $0 \le r \le \frac{1}{2}$.

Given P_{ij}, we want to find the gametic frequencies P'_{ij} in the next generation after random mating. The derivation of the recursion equation is based on the following basic fact of Mendelian genetics: an individual with genotype $\mathcal{A}_i\mathcal{B}_j/\mathcal{A}_k\mathcal{B}_l$ produces gametes of parental type if no recombination occurs (with probability $1-r$), and recombinant gametes if recombination between the two loci occurs (with probability r). Therefore, the fraction of gametes $\mathcal{A}_i\mathcal{B}_j$ and $\mathcal{A}_k\mathcal{B}_l$ is $\frac{1}{2}(1-r)$ each, and that of $\mathcal{A}_i\mathcal{B}_l$ and $\mathcal{A}_k\mathcal{B}_j$ is $\frac{1}{2}r$ each. From these considerations, we see that the frequency of gametes of type $\mathcal{A}_i\mathcal{B}_j$ in generation $t+1$ produced without recombination is $(1-r)P_{ij}$. With recombination, $\frac{1}{2}$ of the gametes

[9] For detailed descriptions of recombination and meiosis consult, for instance, https://www.nature.com/scitable, especially https://www.nature.com/scitable/definition/recombination-226/ and the links shown there, or Wikipedia.

produced by individuals of type $\mathcal{A}_i\mathcal{B}_l/\mathcal{A}_k\mathcal{B}_j$ (for every l and k) is of type $\mathcal{A}_iB_j$. Because of random mating, the frequency of $\mathcal{A}_i\mathcal{B}_l/\mathcal{A}_k\mathcal{B}_j$ genotypes is $P_{il}P_{kj}$. Summing over all l and k, we conclude that the frequency of gametes $\mathcal{A}_i\mathcal{B}_j$ produced with recombination is rp_iq_j. Thus,

$$P'_{ij} = (1-r)P_{ij} + rp_iq_j\,. \tag{5.2}$$

This shows that the gene frequencies are conserved, but the gamete frequencies are not, unless the population is in linkage equilibrium, (5.1). Commonly, linkage disequilibrium between alleles $\mathcal{A}_i$ and B_j is measured by the parameter

$$D_{ij} = P_{ij} - p_iq_j\,. \tag{5.3}$$

The D_{ij} are often called simply linkage disequilibria, although no single D_{ij} is a complete measure of linkage disequilibrium. From (5.2) and (5.3) we infer that

$$D'_{ij} = (1-r)D_{ij} \tag{5.4}$$

and, hence,

$$D_{ij}(t) = (1-r)^tD_{ij}(0)\,. \tag{5.5}$$

Therefore, unless $r = 0$, linkage disequilibria decay at the geometric rate $1-r$ and linkage equilibrium is approached gradually without oscillation. With unlinked loci, $r = \frac{1}{2}$, linkage disequilibrium is halved each generation.

For two alleles at each locus, it is more convenient to label the frequencies of the gametes $\mathcal{A}_1\mathcal{B}_1$, $\mathcal{A}_1\mathcal{B}_2$, $\mathcal{A}_2\mathcal{B}_1$, and $\mathcal{A}_2\mathcal{B}_2$ by x_1, x_2, x_3, and x_4, respectively. A simple calculation reveals that in this case the difference of the frequency of coupling genotypes, $\mathcal{A}_1\mathcal{B}_1/\mathcal{A}_2\mathcal{B}_2$, and repulsion genotypes, $\mathcal{A}_1\mathcal{B}_2/\mathcal{A}_2\mathcal{B}_1$,

$$D = x_1x_4 - x_2x_3\,, \tag{5.6}$$

satisfies

$$D = D_{11} = -D_{12} = -D_{21} = D_{22}\,. \tag{5.7}$$

Thus, the recursion equations for the gamete frequencies, (5.2), may be rewritten as

$$x'_1 = x_1 - rD\,, \tag{5.8a}$$
$$x'_2 = x_2 + rD\,, \tag{5.8b}$$
$$x'_3 = x_3 + rD\,, \tag{5.8c}$$
$$x'_4 = x_4 - rD\,. \tag{5.8d}$$

The state space of this discrete dynamical system is the simplex S_4 because $x_i \geq 0$ and $x_1 + x_2 + x_3 + x_4 = 1$. Therefore, the two-locus gametic frequencies may be represented

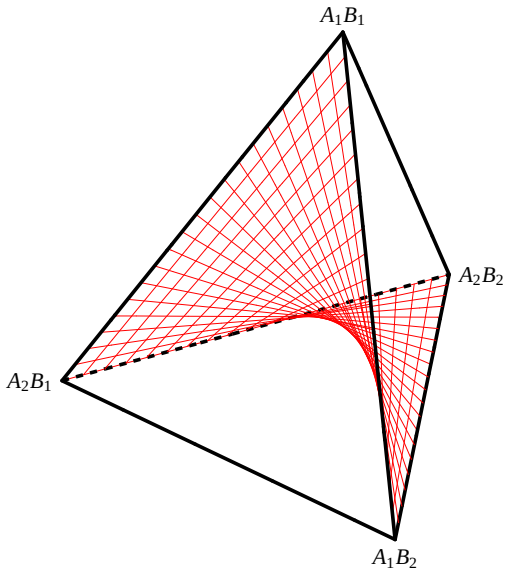

Figure 5.1: The tetrahedron represents the state space of the two-locus two-allele model. The vertices correspond to fixation of the labeled gamete, and frequencies are measured by the (orthogonal) distance from the opposite boundary face. Thus, at the center of the simplex all gametes have frequency $\frac{1}{4}$. The two-dimensional (red) surface is the linkage-equilibrium manifold corresponding to the states in linkage equilibrium, $D = 0$. The states of maximum linkage disequilibrium, $D = \pm\frac{1}{4}$, are the centers of the edges connecting $\mathcal{A}_1\mathcal{B}_2$ to $\mathcal{A}_2\mathcal{B}_1$ and $\mathcal{A}_1\mathcal{B}_1$ to $\mathcal{A}_2\mathcal{B}_2$.

geometrically by the points in a tetrahedron. The subset where $D = 0$ forms a two-dimensional manifold and is called the linkage equilibrium, or Wright, manifold. It is displayed in Figure 5.1.

It follows from (5.5) that, if $r > 0$, all solutions of (5.8) converge to the linkage-equilibrium manifold along straight lines, because the allele frequencies, x_1+x_2 and x_1+x_3, remain constant, and sets of the form $x_1 + x_2 = \text{const.}$ represent planes in this geometric picture. In the present simple model, the linkage-equilibrium manifold is invariant under the dynamics (5.8). With selection or mutation, this is generally not the case.

If there are more than two loci, linkage disequilibria among any group of at least two loci have to be considered. However, it can again be proved that under random mating all gametic combinations eventually reach equilibrium proportions. The rate of decay can be shown to be $1 - r_{\min}$, where $r_{\min}$ is the smallest of all two-locus recombination fractions. In particular, higher-order disequilibria decay faster than the two-locus disequilibria. The two-locus results were first derived by Robbins (1918), but the simple derivation given

above is due to Malécot (1948). The multilocus case was first treated by Geiringer (1944).

Many, if not most, traits of evolutionary or ecological significance are determined by multiple loci. Typically, these traits are under selection, which is transmitted to the underlying loci. Selection on such traits causes not only temporal changes in allele frequencies at these loci but also induces nonrandom associations among them, linkage disequilibria, which will change, too. Research in multilocus population-genetic theory is an important and active field. However, it is also notoriously complex. Already simple diallelic two-locus models, whose state space is the simplex S_4, can exhibit complex dynamics, such as stable cycling, and are much more difficult to study than one-locus models with four alleles. In the absence of recombination, any n-locus model reduces to a one-locus multiallelic selection model of the same dimension (by formally identifying each multilocus gamete with an allele). If recombination is much stronger than selection, then the dynamics can be approximated by that of independent loci, i.e., by the system of corresponding one-locus equations. However, they are not independent of each other but linked by the mean fitness, which depends on all loci. Because recombination probabilities are bounded by $\frac{1}{2}$, this situation can often be realized only by assuming that selection is very weak, in particular if there are loci on the same chromosome. Many important aspects of this theory are treated in Bürger (2000); for a recent and concise review, consult Bürger (2020).

A Basics from dynamical systems

The study of the evolution of gene frequencies requires the mathematical investigation of difference or differential equations that describe gene-frequency change across generations. Here, we summarize the basic concepts needed for exploring the dynamical and equilibrium properties of solutions of such equations. A highly recommended elementary introduction to the theory of differential equations and dynamical systems is the book by Hirsch and Smale (1974). A concise introduction, focused on stability of difference and differential equations, is LaSalle's (1976) text. Hofbauer and Sigmund (1998) develop the theory of dynamical systems hand in hand with topics from evolutionary biology.

A.1 Difference equations

Let $\mathbb{R}^k$ denote the k-dimensional Euclidean space and let X be a subset of $\mathbb{R}^k$. A *discrete dynamical system* consists of a map T of X into itself, $\boldsymbol{x} \to T\boldsymbol{x}$, where $\boldsymbol{x}$ is a vector in X. We shall call X the state space. The map T can be iterated, and the sequence $\boldsymbol{x}$, $T\boldsymbol{x}$, $T^2\boldsymbol{x} = T(T\boldsymbol{x}), \ldots, T^n\boldsymbol{x}, \ldots$, is called the *orbit*, or *trajectory*, of $\boldsymbol{x}$. The vector $\boldsymbol{x}$ may be interpreted as the initial state and n as the number of time intervals elapsed. We assume

that T is differentiable. Associated with the map $\boldsymbol{x} \to T\boldsymbol{x}$ is the *difference equation*, or *recursion relation*,

$$\boldsymbol{x}' = T\boldsymbol{x}\,, \tag{A.1}$$

which stands for $\boldsymbol{x}(n+1) = T(\boldsymbol{x}(n))$.

Of primary interest is the behavior of $T^n\boldsymbol{x}$ for large values of n. A point $\boldsymbol{y}$ is a *limit point* (accumulation point) of $T^n\boldsymbol{x}$ if there is a sequence of integers n_i ($n_i \to \infty$) such that $T^{n_i}\boldsymbol{x} \to \boldsymbol{y}$. The *$\omega$-limit* $\omega(\boldsymbol{x})$ of (the orbit $T^n\boldsymbol{x}$ of) $\boldsymbol{x}$ is the set of all limit points of $T^n\boldsymbol{x}$. An orbit $T^n\boldsymbol{x}$ is called *periodic* (or *cyclic*) if for some $k > 0$, $T^k\boldsymbol{x} = \boldsymbol{x}$. The least such integer is the *period* of the cycle. A point $\boldsymbol{x}$ such that $T\boldsymbol{x} = \boldsymbol{x}$ is called an *equilibrium*, a *fixed point*, or a *stationary state*. Frequently, ω-limits consist of periodic orbits or equilibria. To describe the basic properties of ω-limits, we need two further notions. A set H is said to be *positively invariant* if $T(H) \subseteq H$, and *invariant* if $T(H) = H$. Finally, a closed invariant set H is said to be *invariantly connected* if it is not the union of two nonempty disjoint closed invariant sets. An invariant set with a finite number of elements is invariantly connected if and only if it is a periodic motion.

ω-limits have the following two important properties (LaSalle 1976).

Theorem A.1. *1. Every ω-limit is closed and positively invariant.*

2. If $T^n\boldsymbol{x}$, $n \geq 1$, is bounded, then $\omega(\boldsymbol{x})$ is nonempty, compact, invariant, invariantly connected, and is the smallest closed set that $T^n\boldsymbol{x}$ approaches as $n \to \infty$.

This implies, for example, that if $T^n\boldsymbol{x}$ approaches a set with finitely many elements, $\omega(\boldsymbol{x})$ is a periodic set. In particular, if $T^n\boldsymbol{x}$ converges (i.e., to a single point $\boldsymbol{y}$), then $\boldsymbol{y} = \omega(\boldsymbol{x})$ is an equilibrium.

The central tool for obtaining information about the location of ω-limits are *Lyapunov functions*. Let $V : X \to \mathbb{R}$ and define $\Delta V(\boldsymbol{x}) = V(\boldsymbol{x}') - V(\boldsymbol{x})$. This can be computed without knowing the solution of (A.1). The function V is called a Lyapunov function of (A.1) on a subset $Y \subseteq X$ if (i) V is continuous in $\boldsymbol{x}$ and (ii) $\Delta V(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in Y$. (Condition (ii) could be replaced by $\Delta V(\boldsymbol{x}) \geq 0$.) We denote by $\overline{Y}$ the closure of Y, i.e., the smallest set containing Y and all of its accumulation points. Now we can formulate an extended version of Lyapunov's invariance principle (LaSalle 1976).

Theorem A.2. *If V is a Lyapunov function of (A.1) on Y and if $T^n\boldsymbol{x}$ is bounded and in Y for every $n \geq 0$, then $\omega(\boldsymbol{x})$ is contained in the maximal invariant subset of $\{\boldsymbol{y} \in \overline{Y} : \Delta V(\boldsymbol{y}) = 0\}$. In particular, there is a number c such that $V(\boldsymbol{y}) = c$ for all $\boldsymbol{y} \in \omega(\boldsymbol{x})$.*

Next we consider questions of stability. A set H is said to be *stable* if, given a neighborhood U of H (i.e., an open set containing the closure $\overline{H}$), there is a neighborhood

W of H such that $T^n(W) \subseteq U$ for every $n \geq 0$. This means that any orbit through W remains in U. This is a fairly weak concept of stability. A stronger and more important notion of stability is that of asymptotic stability. First, a set H is an *attractor* if there is a neighborhood U of H such that $x \in U$ implies $\omega(\boldsymbol{x}) \subseteq \overline{H}$. The set H is said to be *asymptotically stable* if it is stable and an attractor. H is called *globally asymptotically stable* if $T^n\boldsymbol{x} \to \overline{H}$ as $n \to \infty$ for all $x \in X$. The *basin of attraction* of a set H is the set of all $\boldsymbol{x}$ such that $\omega(\boldsymbol{x}) \in H$. The following criterion shows how a Lyapunov function can be used to prove (global) asymptotic stability and to obtain estimates for the basin of attraction.

Theorem A.3. *Let Y be a bounded open positively invariant set in X, and denote by M the largest invariant set in $\{\boldsymbol{y} \in \overline{Y} : \Delta V(\boldsymbol{y}) = 0\}$. If* (i) *$V$ is a Lyapunov function of* (A.1) *on Y and* (ii) *$M \subset Y$, then M is an attractor and the basin of attraction contains $\overline{Y}$. If, in addition,* (iii) *V is constant on M, then M is asymptotically stable and globally stable relative to Y.*

We notice that condition (iii) is automatically satisfied if M is a single point or if M is an invariantly connected set with a finite number of elements. The main difficulty in applying the above results consists, of course, in finding an appropriate Lyapunov function. Asymptotic stability can also be inferred by linear approximation of (A.1). Suppose that $\hat{\boldsymbol{x}}$ is an equilibrium of (A.1) and denote by A the $k \times k$ matrix that is the linear approximation to T at $\hat{\boldsymbol{x}}$. Thus, $\mathsf{A} = D_{\boldsymbol{x}}T(\hat{\boldsymbol{x}})$ is the *Jacobian matrix*[10] of T, evaluated at $\hat{\boldsymbol{x}}$. Then (A.1) can be written as

$$\boldsymbol{x}' = \hat{\boldsymbol{x}} + \mathsf{A}(\boldsymbol{x} - \hat{\boldsymbol{x}}) + \boldsymbol{h}(\boldsymbol{x} - \hat{\boldsymbol{x}})\,, \tag{A.2}$$

where $\boldsymbol{h}(\boldsymbol{x} - \hat{\boldsymbol{x}})$ is the remainder term. Denote by $r(\mathsf{A})$ the spectral radius of A (cf. Appendix B). Then the following holds:

Theorem A.4. *If $\boldsymbol{h}(\boldsymbol{x})$ is $o(\boldsymbol{x})$ as $\boldsymbol{x} \to 0$ and if $r(\mathsf{A}) < 1$ (i.e., the modulus of all eigenvalues is less than one), then $\hat{\boldsymbol{x}}$ is an asymptotically stable equilibrium of* (A.2). *If $r(\mathsf{A}) > 1$, then $\hat{\boldsymbol{x}}$ is unstable.*

If no eigenvalue of A has absolute value one, then the equilibrium is called *hyperbolic*. In this case, the orbits of (A.1) near an equilibrium $\hat{\boldsymbol{x}}$ look like those of the linearization $\boldsymbol{x}' = \hat{\boldsymbol{x}} + \mathsf{A}(\boldsymbol{x} - \hat{\boldsymbol{x}})$ near $\hat{\boldsymbol{x}}$. Hyperbolic equilibria are either sinks, sources, or saddles.

A powerful tool to prove that a given function is a Lyapunov function is the following result of Baum and Eagon (1967):

[10] For a continuously differentiable map $f : \mathbb{R}^k \to \mathbb{R}^m$ the Jacobian (matrix) $D_{\boldsymbol{x}}f(\boldsymbol{y})$ is the $m \times k$ matrix of first-order partial derivatives $\partial f_i/\partial x_j$ of $f = (f_1, \ldots, f_m)$ evaluated at $\boldsymbol{y}$.

Theorem A.5. *Let $P(\boldsymbol{x})$ be a polynomial with nonnegative coefficients, homogeneous of degree d in its variables $x_1, \ldots, x_k$. Let $\boldsymbol{x} = (x_1, \ldots, x_k)$ be any point satisfying $x_i \geq 0$ for every i and $\sum_{i=1}^{k} x_i = 1$, and let $\boldsymbol{y}(\boldsymbol{x}) = (y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x}))$ be given by*

$$y_i(\boldsymbol{x}) = x_i \frac{\partial P}{\partial x_i} \Big/ \sum_{j=1}^{k} x_j \frac{\partial P}{\partial x_j} \,. \tag{A.3}$$

Then $P(\boldsymbol{y}(\boldsymbol{x})) > P(\boldsymbol{x})$ holds unless $\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{x}$.

A.2 Differential equations

Let Y be an open set in $\mathbb{R}^k$, $f : Y \to \mathbb{R}^k$ a (sufficiently often) differentiable function, and let the state space X be contained in Y. We consider time-independent ordinary differential equations of the form

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \dot{\boldsymbol{x}} = f(\boldsymbol{x}) \,, \tag{A.4}$$

where $f(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}))$. If for all $\boldsymbol{x} \in X$ and all $t \in \mathbb{R}$, the solution $\boldsymbol{x}(t)$ with $\boldsymbol{x}(0) = \boldsymbol{x}$ is defined and lies in X, then (A.4) determines a *continuous dynamical system* in X. Many of the concepts and results encountered above for difference equations possess straightforward analogues for such differential equations.

To each $\boldsymbol{x} \in X$ corresponds the *orbit* $\{\boldsymbol{x}(t) : t \in \mathbb{R}\}$. Sometimes, one considers semi-dynamical systems defined only for $t \geq 0$ and *positive semi-orbits* $\{\boldsymbol{x}(t) : t \geq 0\}$. Let $\boldsymbol{x}(t)$ be a solution of (A.4) defined for all $t \geq 0$, satisfying the initial condition $\boldsymbol{x}(0) = \boldsymbol{x}$. The ω-limit of $\boldsymbol{x}$ is the set of all accumulation points of $\boldsymbol{x}(t)$ for $t \to \infty$, i.e.,

$$\omega(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{x}(t_k) \to \boldsymbol{y} \text{ for some sequence } t_k \to \infty\} \,. \tag{A.5}$$

The ω-limit may be empty. However, if the positive semi-orbit remains in some compact set, the ω-limit cannot be empty. Every point on the orbit of $\boldsymbol{x}$ has the same ω-limit. A set is called (*positively*) *invariant* if every solution which starts in it remains there for all $t \in \mathbb{R}$ (for all $t \geq 0$). α-limits and negatively invariant sets are defined in the same way, but for $t \to -\infty$. The following theorem is analogous to Theorem A.1.

Theorem A.6. *1. Every ω-limit is closed and invariant.*

2. If $\boldsymbol{x}(t)$ remains in a compact set for all $t \geq 0$, then $\omega(\boldsymbol{x})$ is nonempty, compact, connected, invariant, and is the smallest closed set that $\boldsymbol{x}(t)$ approaches as $t \to \infty$.

A point $\boldsymbol{x}$ is called an *equilibrium* if $\boldsymbol{x}(t) = \boldsymbol{x}$ for all $t \in \mathbb{R}$. These points are characterized by the condition $f(\boldsymbol{x}) = \boldsymbol{0}$. A point $\boldsymbol{x}$ is called a *periodic point* if $\boldsymbol{x}(\tau) = \boldsymbol{x}(0)$

for some $\tau > 0$, but $\boldsymbol{x}(t) \neq \boldsymbol{x}$ for all $t \in (0, \tau)$. The number τ is called the period. Such a motion describes a periodic oscillation. Equilibria and periodic orbits constitute their own ω-limit.

A continuously differentiable function $V : Y \to \mathbb{R}$ is called a *Lyapunov function* if V is nondecreasing (or nonincreasing) along orbits, i.e., if the time derivative $\dot{V}$ of the map $t \to V(\boldsymbol{x}(t))$ satisfies $\dot{V} \geq 0$ (or $\dot{V} \leq 0$). This derivative can be calculated without knowing the solutions, because

$$\dot{V}(\boldsymbol{x}) = \operatorname{grad} V(\boldsymbol{x})^\top \dot{\boldsymbol{x}} = \sum_{i=1}^{k} \frac{\partial V(\boldsymbol{x})}{\partial x_i} f_i(\boldsymbol{x}) \,, \tag{A.6}$$

where $\operatorname{grad} V(\boldsymbol{x})$ denotes the gradient vector. Now we formulate a generalization of Lyapunov's invariance principle.

Theorem A.7. *Let S and Y_1 be positively invariant subsets of Y with respect to (A.4) such that $S \subset Y_1 \subseteq Y$.*

1. *If $\dot{V} < 0$ on $Y_1 \setminus S$ (thus V is Lyapunov function on $Y_1 \setminus S$), then for all $\boldsymbol{x} \in Y_1$ the ω-limit $\omega(\boldsymbol{x})$ is contained in S.*

2. *If $\dot{V} \leq 0$ on Y_1, then every such ω-limit is contained in the maximal invariant subset of $\{\boldsymbol{y} \in Y_1 : \dot{V}(\boldsymbol{y}) = 0\}$.*

Proofs and more details may be found in LaSalle (1976) and Hofbauer and Sigmund (1998) (cf. also Bürger 1983b).

For differential equations, the notions of attractor, asymptotic stability, etc. are defined in the same way as for difference equations. Thus, Theorem A.7 provides an important tool for proving global stability of equilibria relative to some subset of the state space.

In many cases the local behavior of solutions near an equilibrium $\hat{\boldsymbol{x}}$ of (A.4) can be determined by studying the approximating linear differential equation

$$\dot{\boldsymbol{x}} = \mathsf{A}\boldsymbol{x} \,, \tag{A.7}$$

where $\mathsf{A} = D_{\boldsymbol{x}} f(\hat{\boldsymbol{x}})$ is the Jacobian matrix of f. In an informal way, the Theorem of Hartman and Grobman states that for any equilibrium $\hat{\boldsymbol{x}}$ of (A.4) that has no eigenvalues on the imaginary axis (such equilibria are called *hyperbolic*), the orbits of (A.4) near $\hat{\boldsymbol{x}}$ look like those of (A.7) near $\mathbf{0}$. In particular, if all eigenvalues of A have (strictly) negative real part, then $\hat{\boldsymbol{x}}$ is (locally) asymptotically stable; the equilibrium $\hat{\boldsymbol{x}}$ is a *saddle* if some eigenvalues are in the left half and some in the right half of the complex plane, but none on the imaginary axis. The orbits whose ω-limit is $\{\mathbf{0}\}$ form a linear submanifold of $\mathbb{R}^k$, called the *stable manifold*; those whose α-limit is $\{\mathbf{0}\}$ form the *unstable manifold*. The

local behavior of orbits of (A.4) near equilibria with eigenvalues on the imaginary axis depends on higher-order terms of the Taylor expansion of f.

In population genetics, as well as in several other fields of evolutionary biology, (systems of) differential equations occur, with the simplex S_k as their state space, that are of the form

$$\dot{x}_i = x_i[f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})]\,, \qquad i = 1, \ldots, k\,, \tag{A.8}$$

with $\bar{f}(\boldsymbol{x}) = \sum_{i=1}^{k} x_i f_i(\boldsymbol{x})$. These are called replicator equations (Hofbauer and Sigmund 1998). A typical example is the selection equation in continuous time. Equilibria of such equations are given by the conditions $x_i = 0$ or $f_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, k$. An equilibrium point $\hat{\boldsymbol{x}}$ of (A.8) is called *saturated*, or *externally stable*, if $f_i(\hat{\boldsymbol{x}}) \leq \bar{f}(\hat{\boldsymbol{x}})$ whenever $\hat{x}_i = 0$. Every equilibrium in the interior of S_k is trivially saturated. For an equilibrium on the boundary, the condition means that if a missing type is introduced at low frequency it will be lost. Typical examples of nonsaturated, or externally unstable, equilibria are the boundary equilibria in the one-locus two-allele selection model with overdominance. If mutation is added to a model like (A.8), then a saturated equilibrium at the boundary will be pushed into the interior of the state space. For a boundary equilibrium $\hat{\boldsymbol{x}}$, $f_i(\hat{\boldsymbol{x}}) - \bar{f}(\hat{\boldsymbol{x}})$ is an eigenvalue of the Jacobian if $x_i = 0$. It is called a transversal eigenvalue and measures the rate of approach to the face $x_i = 0$ near $\hat{\boldsymbol{x}}$. A boundary equilibrium is saturated if and only if all its transversal eigenvalues are nonpositive (see Hofbauer and Sigmund (1998) for more results).

A.3 Gradient systems

A differential equation of the form (A.4) is called a *gradient system* if there exists a function $V : Y \to \mathbb{R}$ with continuous partial second-order derivatives, such that

$$\dot{\boldsymbol{x}} = -\operatorname{grad} V(\boldsymbol{x})\,. \tag{A.9}$$

The function V is called the *potential*. Its derivative is

$$D_{\boldsymbol{x}} V(\boldsymbol{y}) = \operatorname{grad} V(\boldsymbol{x})^{\top} \boldsymbol{y}\,. \tag{A.10}$$

By (A.6) and (A.9), this implies that the time derivative of V along orbits is given by

$$\dot{V}(\boldsymbol{x}) = \operatorname{grad} V(\boldsymbol{x})^{\top} \dot{\boldsymbol{x}} = -\left|\operatorname{grad} V(\boldsymbol{x})\right|^2\,. \tag{A.11}$$

Hence, V is a Lyapunov function, and $\dot{V}(\boldsymbol{x}) = 0$ holds if and only if $\boldsymbol{x}$ is an equilibrium. Our next theorem summarizes the basic properties of gradient systems.

Theorem A.8. *1. At regular points* ($\operatorname{grad} V(\boldsymbol{x}) \neq 0$)*, the orbits cross the level surfaces* ($V(\boldsymbol{x}) = const.$) *orthogonally.*

2. *Nonregular points are equilibria.*

3. *The α- and ω-limits of any orbit are critical points (where $\operatorname{grad} V(\boldsymbol{x}) = 0$) of V.*

4. *Isolated minima of V are asymptotically stable.*

5. *The differential equation $\dot{\boldsymbol{x}} = -f(\boldsymbol{x})$ is a gradient system if and only if the integrability conditions $\partial f_i/\partial x_j = \partial f_j/\partial x_i$ hold for every i, j. In particular, the Jacobian matrix $D_{\boldsymbol{x}} f = (\partial^2 V/(\partial x_i \partial x_j))$ is symmetric and all eigenvalues at an equilibrium are real.*

B Perron–Frobenius theory of nonnegative matrices

In this appendix we summarize some important results from the spectral theory of nonnegative matrices. These were discovered by Perron and Frobenius around 1910 and are useful tools in proving existence, uniqueness, positivity, and stability of equilibrium solutions in mutation-selection models. For a more complete account of the spectral theory of nonnegative matrices, including proofs, the reader is referred to Gantmacher (1959), Schaefer (1974, Chapter I), or Seneta (1981). The latter reference contains, in particular, a detailed treatment of countably infinite matrices.

A $k \times k$ matrix $\mathsf{A} = (a_{ij})$ is called *nonnegative*, $\mathsf{A} \geq 0$, if $a_{ij} \geq 0$ for every i, j. It is called *positive*, $\mathsf{A} > 0$, if $a_{ij} > 0$ for every i, j. Similarly, a vector $\boldsymbol{x} = (x_1, \ldots, x_k)^\top$ is said to be nonnegative (positive) if $x_i \geq 0$ ($x_i > 0$) for every i.

The *spectral radius* $r = r(\mathsf{A})$ of an arbitrary matrix A is the radius of the smallest circle in the complex plain that contains all eigenvalues of A, i.e., $|\lambda| \leq r$ for all eigenvalues λ of A. It can be shown that $r = \lim_n \bigl\|\mathsf{A}^n\bigr\|^{1/n}$, where $\bigl\|\mathsf{A}\bigr\|$ is an arbitrary norm of the matrix A, e.g., $\bigl\|\mathsf{A}\bigr\| = \max_i \sum_{j=1}^k |a_{ij}|$. (Throughout this appendix, $\lim_n$ denotes the limit for $n \to \infty$.) Since the sequence $\bigl\|\mathsf{A}^n\bigr\|^{1/n}$ is monotone decreasing, $r \leq \bigl\|\mathsf{A}^n\bigr\|^{1/n}$ holds for every $n \geq 1$. Nonnegative matrices have the following important property:

Theorem B.1. *Let $\mathsf{A} \geq 0$. Then the spectral radius r of A is an eigenvalue and there is at least one nonnegative eigenvector $\boldsymbol{x} \geq 0$ ($\boldsymbol{x} \neq \mathbf{0}$), i.e., $\mathsf{A}\boldsymbol{x} = r\boldsymbol{x}$. In addition, if A has an eigenvalue λ with an associated positive eigenvector, then $\lambda = r$.*

We use the notation $\mathsf{A}^n = (a_{ij}^{(n)})$ for nth powers. A nonnegative matrix A is called *irreducible* if for every pair of indices (i, j) an integer $n = n(i, j) \geq 1$ exists such that $a_{ij}^{(n)} > 0$. Now we state the *Theorem of Perron–Frobenius*.

Theorem B.2. *If A is irreducible, then the following hold:*

1. *The spectral radius r is positive and a simple root of the characteristic equation.*
2. *To r there corresponds a positive right eigenvector $\boldsymbol{x} > 0$ such that $\mathsf{A}\boldsymbol{x} = r\boldsymbol{x}$, and $\boldsymbol{x}$ is unique except for multiplication by a positive constant.*
3. *No other eigenvalue of A is associated with a nonnegative eigenvector.*

This theorem is sufficient to prove our existence, uniqueness, and stability results for the haploid mutation-selection model in continuous time. As noted below (4.7), in discrete time a stronger condition than irreducibility is needed.

A nonnegative matrix A is called *primitive* if an integer $n \geq 1$ exists such that $\mathsf{A}^n > 0$. Obviously, every positive matrix is primitive, and every primitive matrix is irreducible.

Theorem B.3. *For an irreducible matrix* A *with spectral radius* r*, the following assertions are equivalent:*

1. A *is primitive.*
2. $|\lambda| < r$ *for all eigenvalues* $\lambda \neq r$ *of* A*.*
3. $\lim_n (r^{-1}\mathsf{A})^n$ *exists.*

Concerning property 3, it is readily shown that for an arbitrary matrix A, $\lim_n \mathsf{A}^n = 0$ is equivalent to $r(\mathsf{A}) < 1$, and that $r(\mathsf{A}) > 1$ always implies that $\lim_n \mathsf{A}^n$ does not exist. If $r(\mathsf{A}) = 1$, then $\lim_n \mathsf{A}^n$ exists if and only if $r(\mathsf{A}) = 1$ is a simple root of the minimal polynomial and all other eigenvalues satisfy $|\lambda| < 1$.

A stronger and more precise statement of Theorem B.3.3. is the following:

Theorem B.4. *Let* A *be primitive with spectral radius* r *and corresponding eigenvector* $\boldsymbol{x} > 0$*. Then there exists a decomposition* $\mathsf{A} = r\mathsf{P} + \mathsf{B}$*, where* P *is a projection on the eigenspace spanned by* $\boldsymbol{x}$ *(i.e., for every* $\boldsymbol{y} \in \mathbb{R}^k$ *there is a constant* c *such that* $\mathsf{P}\boldsymbol{y} = c\boldsymbol{x}$*, and* $\mathsf{P}\boldsymbol{x} = \boldsymbol{x}$*),* $\mathsf{PB} = \mathsf{BP} = 0$*, and* $r(\mathsf{B}) < 1$*. Consequently,*

$$\lim_n (r^{-1}\mathsf{A})^n \boldsymbol{y} = c\boldsymbol{x} + \lim_n (r^{-1}\mathsf{B})^n \boldsymbol{y} = c\boldsymbol{x} \tag{B.1}$$

holds for all $\boldsymbol{y} \in \mathbb{R}^k$*.*

Finally, the exponential

$$e^{\mathsf{A}} = \sum_{n=0}^{\infty} \frac{1}{n!} \mathsf{A}^n \tag{B.2}$$

of an irreducible matrix A is always positive and, hence, primitive. It follows that

$$\lim_{t\to\infty} e^{-rt} e^{\mathsf{A}t} \boldsymbol{y} = c\boldsymbol{x} \tag{B.3}$$

for some constant c depending on $\boldsymbol{y}$.

Remark B.5. A much more elaborate treatment of Sections 1 – 5 can be found in Chapters I and III of my book (Bürger 2000).

Preliminary bibliography

References

[1] Akin, E. 1993. The General Topology of Dynamical Systems. Providence, R.I.: Amer. Math. Soc.

[2] Atkinson, F.V., Watterson, G.A., and Moran, P.A.P. 1960. A matrix inequality. Quart. J. Math. 11, 137–140.

[3] Baake, E., and Wiehe, T. 1997. Bifurcations in haploid and diploid sequence space models. J. Math. Biol. 35, 321–343.

[4] Barton, N.H., Briggs, D.E.G., Eisen, J.A., Goldstein, D.B., and Patel, N.H. (2007). Evolution. Cold Spring Habor, New York: CSHL Press.

[5] Bennett, J., 1983. Fisher's other correspondence on natural selection and heredity. In: Bennett, J. (Ed.), Natural Selection, Heredity, and Eugenics. Including selected correspondence of R.A. Fisher with Leonhard Darwin and others. Clarendon Press, Oxford.

[6] Broom, M., Cannings, C., and Vickers, G.T. 1993. On the number of local maxima of a constrained quadratic form. Proc. R. Soc. Lond. A 443, 573–584.

[7] Bürger, R. 2000. The Mathematical Theory of Selection, Recombination, and Mutation. Chichester: Wiley.

[8] Bürger, R. 2020. Multilocus population-genetic theory. Theor. Popul. Biol. 133, 40 – 48.

[9] Bürger, R., and Hofbauer, J. 1994. Mutation load and mutation-selection-balance in quantitative genetic traits. J. Math. Biol. 32, 193–218.

[10] Castle, W.E. 1903. The laws of Galton and Mendel and some laws governing race improvement by selection. Proc. Amer. Acad. Arts Sci. 39, 233–242.

[11] Charlesworth B., Charlesworth, D. 2010. Elements of Evolutionary Genetics. Greenwood Village: Roberts & Co.

[12] Crow, J.F., and Kimura, M. 1970. An Introduction to Population Genetics Theory. New York: Harper and Row.

[13] Darwin, C. 1859. The Origin of Species. London: John Murray.

[14] Ewens, W. J. 1969. Mean fitness increases when fitnesses are additive. Nature 221, 1076.

[15] Ewens, W.J. 2004. Mathematical Population Genetics. Springer, Berlin.

[16] Ewens, W.J. 2004. Mathematical Population Genetics. 2nd edition. Springer, New York.

[17] Ewens, W.J., Lessard, S. 2015. On the interpretation and relevance of the fundamental theorem of natural selection. Theor. Popul. Biol. 104, 59-67.

[18] Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinb. 52, 399–433.

[19] Fisher, R.A. 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon Press.

[20] Geiringer, H. 1944. On the probability theory of linkage in Mendelian heredity. Ann. Math. Stat. 15, 25–57.

[21] Haldane, J.B.S. 1932. The Causes of Evolution. London: Longmans Green. (Reprinted with a new introduction and afterword by E.G. Leigh, Jr., by Princeton University Press, 1992.)

[22] Haldane, J.B.S. 1937. The effect of variation on fitness. Amer. Natur. 71, 337 – 349.

[23] Hardy, G.H. 1908. Mendelian proportions in a mixed population. Science 28, 49-50.

[24] Hartfield, M., and Keightley, P.D. 2012. Current hypotheses for the evolution of sex and recombination. Integrative Zoology 2012; 7: 192 – 209.

[25] Hofbauer, J. 1985. The selection mutation equation. J. Math. Biol. 23, 41–53.

[26] Hofbauer, J. 1990, An index theorem for dissipative semiflows. Rocky Mountain J. Math. 20, 1017–1031.

[27] Hofbauer, J., Sigmund, K. 1998. Evolutionary Games and Population Dynamics. Cambridge: University Press.

[28] Karlin, S. 1984. Mathematical models, problems, and controversies of evolutionary theory. Bull. Amer. Math. Soc. 10, 221–273.

[29] Karlin, S., and Lessard, S. 1984. On the optimal sex-ratio: A stability analysis based on a characterization for one-locus multiallele viability models. J. Math. Biol. 20, 15–38.

[30] Kingman, J.F.C. 1961a. An inequality in partial averages. Quart. J. Math. 12, 78-80.

[31] Kingman, J.F.C. 1961b. A mathematical problem in population genetics. Math. Proc. Camb. Phil. Soc. 57, 574–582.

[32] Kingman, J.F.C. 1977. On the properties of bilinear models for the balance between genetic mutation and selection. Math. Proc. Camb. Phil. Soc. 81, 443–453.

[33] Kingman, J.F.C. 1978. A simple model for the balance between selection and mutation. J. Appl. Prob. 15, 1–12.

[34] Krein, M.G., Rutman, M. A. 1950. Linear operators leaving invariant a cone in a Banach space, Amer. Math. Soc. Transl. 26, 1–128.

[35] Lande, R. 1975. The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. 26, 221–235

[36] LaSalle, J.P. 1977. Stability theory for difference equations. In: J.K. Hale (Ed.), Studies in Ordinary Differential Equations. Studies in Mathematics, vol. 14. Mathematical Association of America, Washington, DC, pp. 1–31.

[37] Lessard, S. 1997. Fisher's fundamental theorem of natural selection revisited. Theor. Pop. Biol. 52, 119–136.

[38] Li, C.C. 1955. The stability of an equilibrium and the average fitness of a population. Amer. Natur. 89, 281-295.

[39] Li, C.C. 1967. Fundamental theorem of natural selection. Nature 214, 505-506.

[40] Losert, V., and Akin, E. 1983. Dynamics of games and genes: discrete versus continuous time. J. Math. Biol. 17, 241–251.

[41] Lyubich Yu.I. 1971. Basic concepts and theorems of evolutionary genetics of free populations. Russ. Math. Surv. 26, 51-123.

[42] Lyubich, Yu.I., Maĭstrovskiĭ, G.D., and Ol'khovskiĭ, Yu.G. 1980. Selection-induced convergence to equilibrium in a single-locus autosomal population. Probl. Inform. Transm. 16, 66–75.

[43] Lyubich Yu.I. 1992. Mathematical Structures in Population Genetics. Berlin Heidelberg New York: Springer.

[44] Malécot, G. 1948. Les Mathématiques de l'Hérédité. Paris: Masson. Extended translation: The Mathematics of Heredity. San Francisco: W.H. Freeman (1969).

[45] Mendel, G.J. 1866. Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereines, Abhandlungen, Brünn 4, 3–47.

[46] Moran, P.A.P. 1976. Global stability of genetic systems governed by mutation and selection. Math. Proc. Camb. Phil. Soc. 80, 331–336.

[47] Mulholland, H.P., and Smith, C.A.B. 1959. An inequality arising in genetical theory. Amer. Math. Monthly 66, 673–683.

[48] Nagylaki, T. 1974. Continuous selective models with mutation and migration. Theor. Pop. Biol. 5, 284–295.

[49] Nagylaki, T. 1991. Error bounds for the fundamental and secondary theorems of natural selection. Proc. Natl. Acad. Sci. USA 88, 2402–2406.

[50] Nagylaki, T. 1992. Introduction to Theoretical Population Genetics. Berlin: Springer.

[51] Nagylaki, T., and Crow, J.F. Continuous selective models. Theor. Popul. Biol. 5, 257–283.

[52] Nagylaki, T., and Lou, Y. 2006. Multiallelic selection polymorophism. Theor. Popul. Biol. 69, 217–229.

[53] Norman, M.F. 1974. A central limit theorem for Markov processes that move by small steps. Ann. Prob. 2, 1065–1074.

[54] Passekov, V.P. 1978. On the relation between steady-state patterns of the gene frequencies' distribution and extension of the Fisher fundamental theorem for a single multi-allelic locus. Doklady MOIP, Obschaya Biologiya, 1978, pp. 135–138 (in Russian).

[55] Provine, W.B. 1971. The Origins of Theoretical Population Genetics. Chicago: Univ. Press.

[56] Price, G.R. 1970. Selection and covariance. Nature 227, 520–521.

[57] Robbins, R.B. 1918. Some applications of mathematics to breeding problems. III. Genetics 3, 375–389.

[58] Scheuer, P.A.G., and Mandel, S.P.H. 1959. An inequality in population genetics. Heredity 13, 519–524.

[59] Seneta, E. 1981. Non-negative Matrices, 2nd ed. London: Allen and Unwin.

[60] Shahshahani, S. 1979. A new mathematical framework for the study of linkage and selection. Memoirs Amer. Math. Soc. 211. Providence, R.I.: Amer. Math. Soc.

[61] Sharp, N.P., and Otto, S.P., 2016. Evolution of sex: Using experimental genomics to select among competing theories. BioEssays 38 (8), 751–757.

[62] Svirezhev, Yu.M. 1972. Optimality principles in population genetics. In: Studies in Theoretical Genetics, pp. 86–102. Novisibirsk: Inst. of Cytology and Genetics. (In Russian).

[63] Svirezhev, Yu.M., and Passekov V.P. 1990. Fundamentals of Mathematical Evolutionary Genetics. Dordrecht: Kluwer.

[64] Vickers, G.T., and Cannings, C. 1988. On the number of stable equilibria in a one-locus multi-allele system. J. Theor. Biol. 131, 273–277.

[65] Thompson, C.J., and McBride, J.L. 1974. On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. Math. Biosci. 21, 127–142.

[66] Wakeley, J. 2008. Coalescent Theory: An Introduction. Roberts & Company Publishers, Greenwood Village, Colorado.

[67] Walsh, B., and Lynch, M., 2018. Evolution and Selection of Quantitative Traits. Sinauer Associates, Oxford.

[68] Weinberg, W. 1908. Über den Nachweis der Vererbung beim Menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg 64, 368-382.

[69] Weinberg, W. 1909. Über Vererbungsgesetze beim Menschen. Zeitschrift für induktive Abstammungs- und Vererbungslehre 1, 377-392, 440-460; 2, 276-330.

[70] Wright, S. 1931. Evolution in Mendelian populations. Genetics 16, 97-159.

[71] Wright, S. 1935. Evolution in populations in approximate equilibrium. J. Genetics 30, 257–266.

[72] Yule, G.U. 1902. Mendel's laws and their probable relations to intra-racial heredity. New Phytol. 1, 192–207, 222–238.