

---

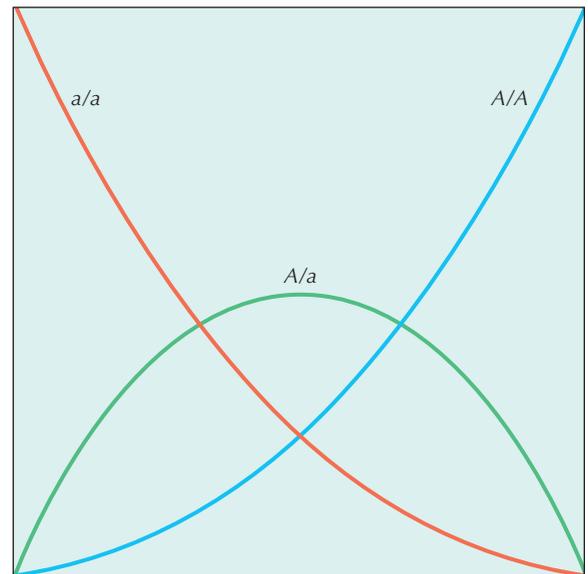
## Models of Evolution

**Q**UANTITATIVE MODELS PLAY A CRUCIAL role in evolutionary biology, especially in population genetics. Mathematical analysis has shown how different evolutionary processes interact, and statistical methods have made it possible to interpret observations and experiments. The availability of computing power and the flood of genomic data have made quantitative methods even more prominent in recent years.

Although we largely avoid explicit mathematics in *Evolution*, mathematical arguments lie behind much of what is discussed. In this chapter, the basic mathematical methods that are used in evolutionary biology are outlined. The chapters in the book can be understood in a qualitative way without this chapter, and that may be appropriate for an introductory course. Similarly, many of the problems (online) associated with Chapters 13–23 involve only elementary algebra and do not require the mathematical methods explained in this chapter. However, reading through this chapter and working through the problems that depend on it will give you a much stronger grasp of evolutionary modeling and will allow you to fully engage with the most recent research in the field.

Most students will have had some basic courses in statistics, focusing on methods for testing alternative hypotheses and making estimates from the data. Although we do emphasize the evidence that is the basis for our understanding of evolutionary biology, we do not do more than touch on methods for statistical inference. However, our understanding of the evolutionary processes does fundamentally rest on knowledge of probability: Indeed, much of modern probability theory has been motivated by evolutionary problems (see the Random Processes section of this chapter). This is true both when we follow the proportions of different genotypes within a population and, more obviously, when we follow random evolutionary processes (Chapter 15) in explicitly probabilistic models. (The importance of probability will become apparent when you work through the problems [online].) In this chapter, we discuss the definitions and meaning of probability and summarize the basic theory of probability.

This chapter begins with an explanation of how mathematical theory can be used to model reproducing populations. Next, we discuss how deterministic processes (i.e., those with no random element) can be modeled by following genotype frequencies. Then, random processes are examined, summarizing important probability distribu-

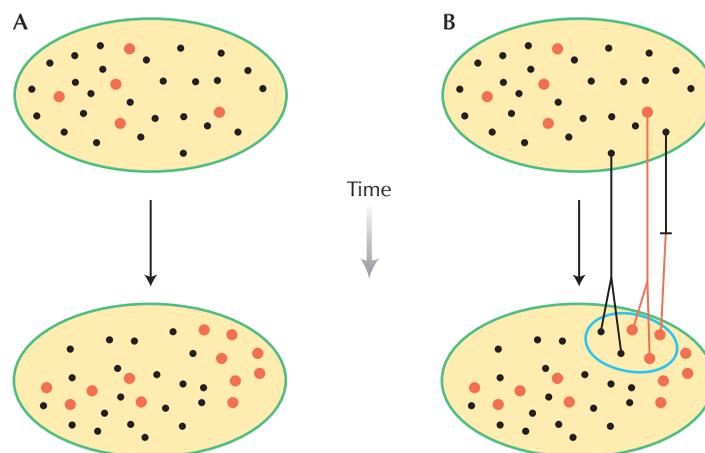


tions and describing two fundamentally random phenomena, **branching processes** and **random walks**. The chapter ends with an outline of the **diffusion approximation**, which has played a key role in the development of the **neutral theory** of molecular evolution (p. 59).

Throughout this chapter, only straightforward algebra and some basic calculus are used. The emphasis is on graphical techniques rather than manipulation of algebraic symbols. Thus, many of the problems (online) can be worked through using a pencil and graph paper. In general, it is more important to understand the qualitative behavior of a model than to get exact predictions: Indeed, models that are too complex may be as hard to understand as the phenomenon being studied (p. 382). This a particular problem with computer simulations: These can show complex and intriguing behavior, but, all too often, this behavior is left unexplained. This chapter should give you an appreciation of the key role that simple mathematical models play in evolutionary biology.

### MATHEMATICAL THEORY CAN BE USED IN TWO DISTINCT WAYS

To aid in understanding evolutionary biology, mathematical techniques can be used in two ways (Fig. 28.1). The first is to work out the consequences of the various evolutionary processes and to identify the key parameters that underlie evolutionary processes. For example, how do complex traits, which are influenced by many genes, evolve (Chapter 14)? How do random drift, gene flow, and selection change populations (Chapters 15–17)? How do these evolutionary forces interact with each other and with other processes (Chapter 18)? As an aid in answering questions like these, the main role of mathematical theory is to guide our understanding and to develop our intuition rather than to make detailed quantitative predictions. Often, verbal arguments can be clarified using simple “toy models,” which show whether an apparently plausible effect can operate, at least in principle. Much of our progress in understanding sexual selection (Chapter 20) and the evolution of genetic systems (Chapter 23) has come from this interplay between verbal argument and mathematical theory.



**FIGURE 28.1.** Theory is used in two ways in evolutionary biology. (A) The evolution of a population can be traced forward in time, giving predictions about the effects of the various evolutionary processes. (B) We can focus on a sample of individuals and trace its ancestry backward in time. This allows us to make inferences about the past processes that shaped the sampled genes. Genes are shown by *dots*, with color indicating allelic state (*red* or *black*). There is a single mutation (*black*→*red*) in the ancestry of the five sampled genes.

Thus, although only verbal arguments are presented in other chapters, these correspond to definite mathematical models.

Mathematical techniques can also identify key parameters underlying evolutionary processes. For example, for a wide range of models, the number of individuals that migrate in each generation determines the interaction between random drift and migration. This number of migrants can be written as  $Nm$ , where  $N$  is the number in the population and  $m$  is the proportion of migrants (see Box 16.2). Similarly, the parameter  $Ns$  (which is, roughly, the number of **selective deaths** per generation) determines the relative rates of random drift and selection. Sometimes, mathematical techniques can identify a key quantity that determines the outcome of a complex evolutionary process. For example, the total rate of production of deleterious mutations determines the loss of fitness due to mutation (p. 552), and the inherited variance in fitness determines the overall rate of adaptation (pp. 462 and 547–549). Once these key quantities have been identified, experiments can then be designed that focus on measuring them, and other complications (e.g., numbers of genes and their individual effects) can be ignored.

Mathematical theory is now being used in a second, and quite distinct, way. The deluge of genetic data that has been generated over the past few decades demands quantitative analysis. Thus, recent theory has focused on how to make inferences about evolving populations, based on samples taken from them. Some questions are quite specific: for example, using forensic samples to find the likely ethnic origin of a criminal, finding the age of the mutation responsible for an inherited disease, or mapping past population movements (pp. 434, 743–748, and 768–773). Others are broader: such as finding what fraction of the genome has a function (pp. 542–547) or measuring rates of recombination along chromosomes (Fig. 15.19).

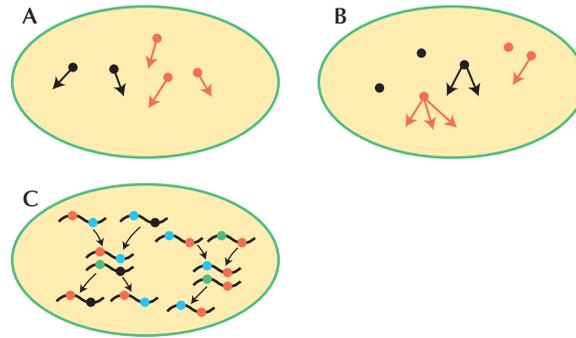
Answering many of these questions requires tracing the ancestry of genes back through time (Fig. 28.1b) rather than modeling the evolution of whole populations forward in time (Fig. 28.1a). Thus, making inferences from genetic data requires new kinds of population genetics and new statistical techniques. These have been especially important in the development of the neutral theory (p. 59) and, more recently, in analyzing DNA sequence data (e.g., pp. 435–437 and 760–765).

## DETERMINISTIC PROCESSES

### A Population Is Described by the Proportions of the Genotypes It Contains

It is populations that evolve and it is populations that we must model. Populations are collections of entities that can be of several alternative types; thus, it is necessary to follow the proportions of these different types. For example, when dealing with a collection of diploid individuals, and focusing on variation at a single gene with two alleles, the proportions of the three possible diploid genotypes QQ, PQ, and PP must be followed. If, instead, the focus is on the collection of genes, then the frequencies of the two types of gene (labeled, for example, Q and P) are followed.

Before going on, two points need to be made concerning terminology. First, different kinds of gene are referred to as **alleles**. Thus, the proportions of alleles Q and P can be written as  $q$  and  $p$ , which are known as **allele frequencies**, and, clearly,  $q + p = 1$ . (Genetic terminology is discussed in more detail in Box 13.1.) Second, perhaps the most important step in building a model is to decide which **variables** describe the population under study. When a population is modeled, these variables are either the allele frequencies or the **genotype frequencies** of the different types or genealogies that describe the ancestry of samples of genes. Variables must be distinguished from **parameters**, which are quantities that determine how the



**FIGURE 28.2.** (A) Physics deals with populations of atoms that remain intact and do not change state. (B) Biology deals with populations of genes or individuals that reproduce and die. The simplest case is where offspring are identical to their parents—as with species, asexual organisms, or genes at a single locus. (C) With sexual reproduction, pairs of individuals reproduce. Genes remain intact, but combinations of genes are reshuffled by sex and recombination.

population will evolve, such as selection coefficients, recombination rates, and mutation rates. As a population evolves, the parameters stay fixed, but the variables change.

Physics and chemistry also deal with populations, but these populations are atoms and molecules instead of genes and individuals. However, atoms are conserved intact, and although molecules may be transformed by chemical reactions, their constituent atoms are not affected (Fig. 28.2A). In biology, the situation is fundamentally different: individuals, and the genes they carry, die and reproduce. This leads to more complex—but also more interesting—phenomena.

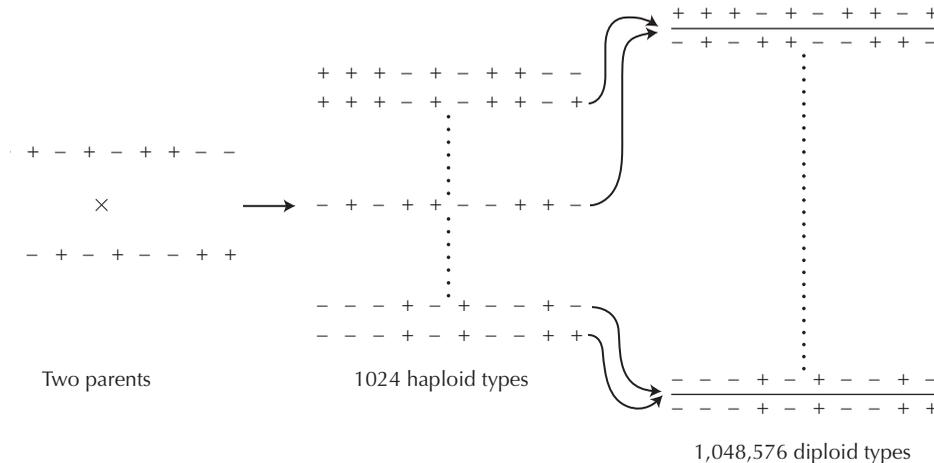
In the simplest case, offspring are identical to their parents (Fig. 28.2B). In ecological models, the assumption is that females of each species produce daughters of the same species, and so the numbers of each species in the whole ecosystem are followed. (Often, males have a negligible effect on population size and so can be ignored.) The same applies to asexual populations (e.g., bacteria), where offspring are genetically identical to their parents, or where the different alleles of a single gene are followed (Fig. 28.2B). Even in this simplest case, however, the reproduction of each individual is affected by the other individuals in the population, leading to a rich variety of phenomena (pp. 470–472 and 505–508). Different species or different asexual genotypes compete with each other, and the different alleles at a single genetic locus combine to form different diploid genotypes.

Sexual reproduction is harder to model, because two individuals come together to produce each offspring, which differs from either parent (Fig. 28.2C). We can no longer follow individual genes but must instead keep track of the proportions of all the different combinations of alleles that each individual carries. Although each gene stays intact *per se*, the set of genes that it is associated with changes from generation to generation.

Whether reproduction is sexual or asexual, a population can be modeled by keeping track of the proportion of each type of offspring. This is much harder for a sexual population, because so many different types can be produced by recombination (Fig. 28.3). However, the basic approach to modeling is the same as when there are just two types. In this chapter, we will work through the simplest case, that of two alleles of a single gene. The basic principles generalize to more complicated cases, where a large number of types must be followed (see Problem 19.15 [online] and Problem 23.7 [online]).

### Reproducing Populations Tend to Grow Exponentially

Evolution usually occurs so slowly that it can be treated as occurring continuously through time, as a continuous change in genotype frequencies or in the distribution



**FIGURE 28.3.** With sexual reproduction, a mating between two types that differ at ten genes can produce  $2^{10} = 1024$  different haploid types and  $2^{20} = 1,048,576$  different diploid genotypes.

of a quantitative trait. Sometimes, evolution may be truly continuous. For example, suppose that in each small time interval, organisms have a constant chance of dying and a constant chance of giving birth, regardless of age. Then, only their numbers through time,  $n(t)$ , need to be followed. Their rate of increase—that is, the slope of the graph of numbers against time—is written as

$$\frac{dn}{dt} = rn.$$

This is a simple **differential equation** that gives the rate of change of the population size in terms of its current numbers. The solution to this equation can be seen in qualitative terms from the graph of Figure 28.4, for  $r = 1$ . When numbers are small, the rate of increase is small, but as numbers increase, so does the rate, giving the characteristic pattern of exponential change. This solution is written as

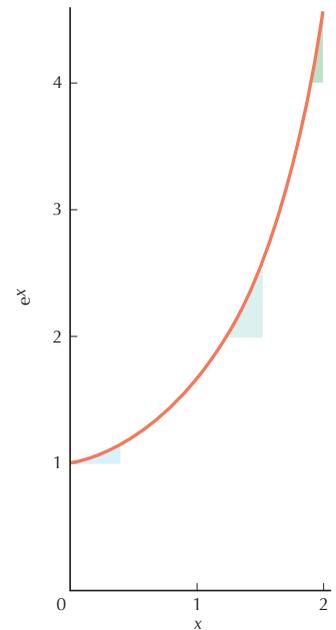
$$n(t) = n(0) e^{rt} \text{ or } n(0) \exp(rt),$$

where  $e = 2.71827\dots$ . This pattern can also be thought of as an approximation to a discrete model of geometric growth, in which numbers increase by a factor  $(1 + r)$  in every time unit. After  $t$  generations, the population has increased by a factor  $(1 + r)^t$ , which tends to  $\exp(rt)$  for small  $r$ . (For example,  $1.1^{10} = 2.594$ , and  $1.01^{100} = 2.705$ , approaching  $e = 2.718\dots$ .)

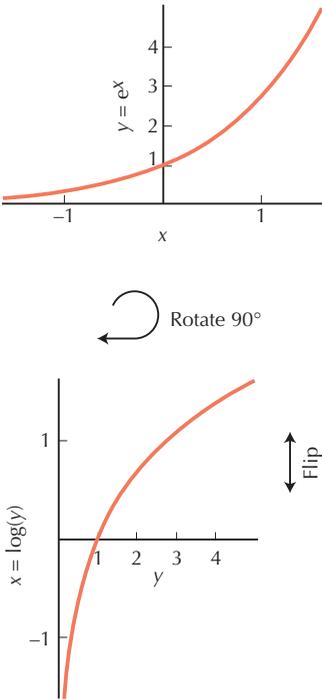
The exponential relationship often needs to be inverted, in order to find the time taken for the population to grow by some factor (Fig. 28.5). This inverse relationship is the **natural logarithm**, defined by  $\log(\exp(x)) = x$ . Thus,

$$\log\left(\frac{n(t)}{n(0)}\right) = rt.$$

Using this formula, the time it takes for a population to grow by a factor  $e = 2.718$  is given by  $rt = \log(e) = 1$ , which is a time  $t = 1/r$ . Similarly, it takes a time  $\log(10)/r \sim 2.30/r$  to increase tenfold, and twice that to increase 100-fold ( $\log(100)/r \sim 4.60/r$ ). (In this book, natural logarithms are always written as  $\log(x)$ . They can be written as  $\log_e(x)$  or  $\ln(x)$  to distinguish them from logarithms to the base 10,  $\log_{10}(x)$ , which were commonly used for arithmetical calculations. Now that electronic calculators are so widespread, logarithms to the base 10 are rarely seen.)



**FIGURE 28.4.** The exponential function  $e^x$  has a slope equal to its value. In other words, it satisfies the differential equation  $dn/dx = n$ . This is illustrated in the graph, which shows slopes of 1, 2, 4 when  $e^x = 1, 2, 4$ , respectively. The shaded triangles illustrate the gradients at points where  $e^x = 1, 2, 4$ . Recall that the slope  $dn/dx$  is the ratio between the vertical and horizontal edges of the triangles.

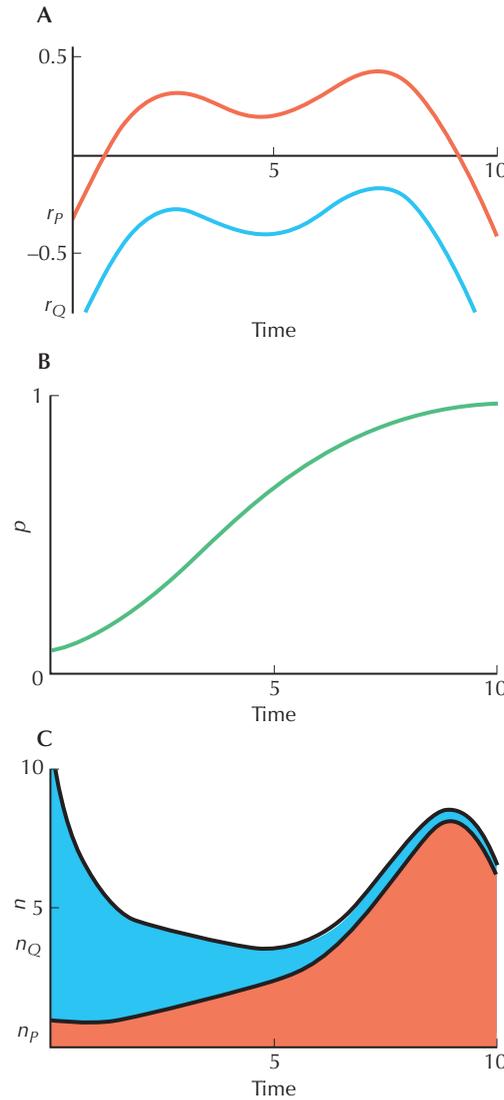


**FIGURE 28.5.** The inverse of the exponential function ( $y = e^x$ ) is the natural logarithm ( $x = \log(y)$ ).

If a population contains two types, Q and P, each with its own characteristic growth rate  $r_Q$  and  $r_P$ , then the numbers of each type will grow exponentially. As explained in Box 28.1, the proportions change according to

$$\frac{dp}{dt} = spq,$$

where the **selection coefficient**  $s$  is the difference in growth rates,  $r_P - r_Q$ . (The consequences of this formula were explored in detail in Box 17.1.) Thus, in continuous time, the appropriate measure of fitness is the growth rate  $r$ , and the selection coefficient is the difference in fitness between genotypes. In any actual population, growth rates will change over time. If the population is to remain more or less stable, growth rates must decrease as the population grows and increase as the population shrinks. However, provided that the growth rates of each genotype change in the same way, the differences between them stay the same, and so changes in allele frequency are not affected by changes in population size (Fig. 28.6).



**FIGURE 28.6.** Provided that the difference in fitness between alleles,  $r_P - r_Q$ , stays the same (as in A), the allele frequency  $P$  follows the standard sigmoid curve over time (B) and is not affected by changing population size,  $n = n_P + n_Q$  (C). (Numbers of alleles P and Q,  $n_P$  and  $n_Q$ , are shown in red and blue, respectively.)

### Box 28.1 Selection in Discrete and Continuous Time

Whether a population reproduces in discrete generations or continuously through time, the rate of change in allele frequency is  $\Delta p = spq$ , where  $s$  is the **selection coefficient**, which is the difference between the relative fitnesses of the two alleles.

**Discrete Time:** Imagine a haploid population, with alleles Q, P at frequencies  $q, p$  (where  $q + p = 1$ ). Each gene of type Q produces  $W_Q$  copies, and each gene of type P produces  $W_P$ . In the next generation, the numbers of each will be in the ratios  $qW_Q:pW_P$ . Dividing by the total gives the new frequencies:

$$\begin{array}{rcc}
 & q & p \\
 \text{selection} & \downarrow & \downarrow \\
 & qW_Q & pW_P \\
 \text{divide by} & \downarrow & \downarrow \\
 \text{total} & q^* = \frac{qW_Q}{\bar{W}} & p^* = \frac{pW_P}{\bar{W}}
 \end{array}
 \quad \bar{W} = qW_Q + pW_P.$$

The change in allele frequency is

$$\begin{aligned}
 \Delta p = p^* - p &= \frac{pW_P}{\bar{W}} - p = \frac{pW_P - p\bar{W}}{\bar{W}} \\
 &= p \frac{W_P - pW_P - qW_Q}{\bar{W}} \\
 &= pq \frac{W_P - W_Q}{\bar{W}}.
 \end{aligned}$$

This can be written as

$$\Delta p = spq, \text{ where } s = \frac{W_P - W_Q}{\bar{W}}.$$

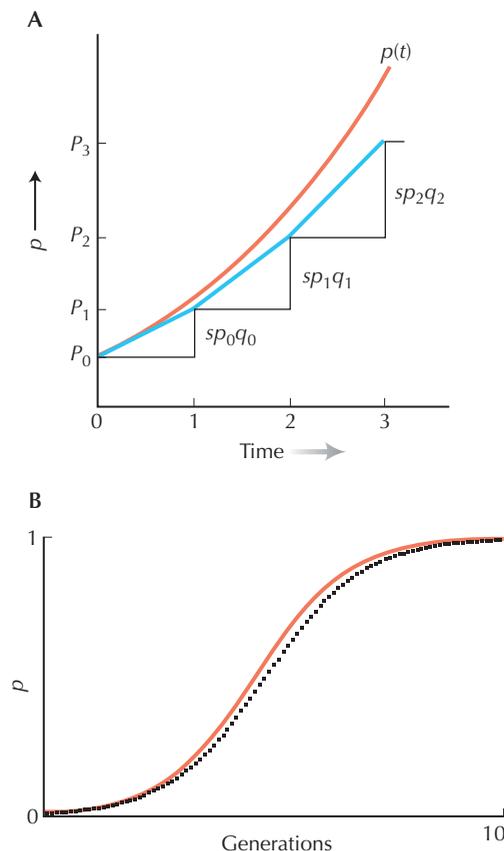
**Continuous time:** Suppose that for types Q, P, their numbers  $n_Q, n_P$  are each growing exponentially at rates  $r_Q, r_P$ . How does the allele frequency,  $p = n_P/(n_Q + n_P)$  change? This is found by differentiating with respect to  $t$  and applying the rules for differentiating products and ratios:

$$\begin{aligned}
 \frac{dn_Q}{dt} &= r_Q n_Q, \quad \frac{dn_P}{dt} = r_P n_P, \\
 \frac{dp}{dt} &= \frac{d}{dt} \left( \frac{n_P}{n_Q + n_P} \right) \\
 &= \frac{dn_P}{dt} \left( \frac{1}{n_Q + n_P} \right) - \left( \frac{dn_Q}{dt} + \frac{dn_P}{dt} \right) \left( \frac{n_P}{(n_Q + n_P)^2} \right) \\
 &= \left( n_Q \frac{dn_P}{dt} - n_P \frac{dn_Q}{dt} \right) \frac{1}{(n_Q + n_P)^2} = (r_P - r_Q) \frac{n_Q n_P}{(n_Q + n_P)^2} \\
 &= (r_P - r_Q) pq \\
 &= spq, \text{ where } s = (r_P - r_Q).
 \end{aligned}$$

## Gradual Evolution Can Be Described by Differential Equations

The basic differential equation  $dn/dt = rn$  has been presented as an exact model of a homogeneous population that reproduces continuously in time. However, it is also a close approximation to populations that reproduce in discrete generations, as long as selection is not too strong. The change in allele frequency from one generation to the next is  $\Delta p = spq$ , where the selection coefficient  $s$  is again the difference in relative fitness between the two alleles (Box 28.1). Approximating this discrete process as a continuous change in allele frequency  $p(t)$  can be thought of as drawing a smooth curve through a series of straight-line segments (Fig. 28.7). More complicated situations are described in Chapter 17; for example, populations may contain a mixture of ages, each with different birth and death rates. However, as long as change is gradual, the population will settle into a pattern of steady growth that can be modeled by a simple differential equation (Box 17.1). It is almost always more convenient to work with differential equations than with discrete recursions, and so most population genetics theory is based on differential equations. Box 28.2 shows how equations can be integrated to find out how long it takes for selection to change allele frequency. Problem 17.9 (online) shows how this method can be used for a variety of models.

Describing gradual evolution by differential equations can be extended to more complicated situations involving several variables, such as following the frequencies of three or more alleles of one gene or following allele frequencies at multiple genes. In such cases, a set of coupled differential equations are generated, in which the rate of change of each variable depends on all the others:



**FIGURE 28.7.** A differential equation is a good approximation to a discrete recursion when change is gradual. (A) The series of triangles shows the solution to the recursion  $p_{t+1} = pt + sp_tq_t$  for generations 0, 1, 2, ... . The continuous red curve is the solution to  $dp/dt = sp(t)q(t)$ . This has the same slope at time  $t = 0$ , when  $p(0) = p_0$ . (Compare the slope of the red curve with the diagonal line, which has slope  $sp_0q_0$ . However, because the differential equation allows for a continuous acceleration (i.e., an upward curve) as allele frequency increases, it increases slightly faster than the discrete recursion (black steps).) (B) The discrete recursion (black dots) with  $W_p/W_Q = 1.1$  compared with the continuous solution (red curve) to  $dp/dt = spq$  with  $s = 0.1$ ; allele frequency is initially 0.01. Agreement is close at first, but errors build up over time; nevertheless, the error is never more than 4%.

### Box 28.2 Solving Differential Equations

Differential equations such as  $dp/dt = spq$  can be solved by thinking of time as depending on allele frequency,  $t(p)$ , rather than of allele frequency as a function of time,  $p(t)$ . In other words, we can ask how long it takes a population to go from one allele frequency to another. This is written as

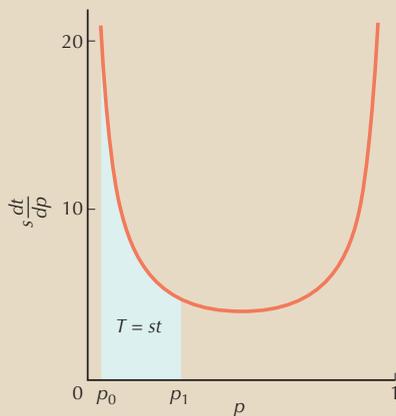
$$\frac{dt}{dp} = \frac{1}{sp(1-p)} = \frac{1}{s} \left( \frac{1}{p} + \frac{1}{1-p} \right).$$

Using the fundamental relation between differentiation and integration,

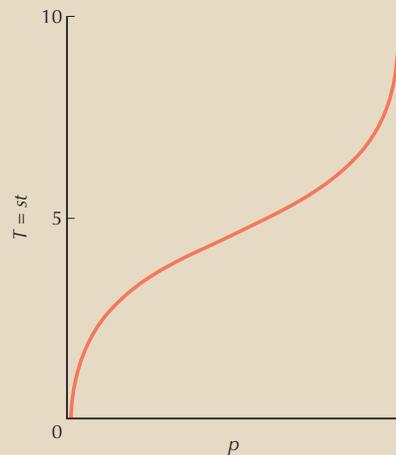
$$\begin{aligned} t(p_1) - t(p_0) &= \int_{p_0}^{p_1} \frac{dt}{dp} dp = \frac{1}{s} \int_{p_0}^{p_1} \left( \frac{1}{p} + \frac{1}{1-p} \right) dp \\ &= \frac{1}{s} \log \left( \frac{p_1}{1-p_1} \frac{1-p_0}{p_0} \right). \end{aligned}$$

This gives the time taken for the allele frequency to change from  $p_0$  to  $p_1$ . As emphasized in the main text, this time is inversely proportional to the selection coefficient.

This derivation can be seen graphically with the aid of Figure 28.8, which plots  $s(dt/dp) = 1/pq$ , the rate of increase of time against allele frequency. The area under the curve is the total time elapsed,  $T = st$ . By definition, this equals the integral of  $1/pq$ , which is plotted in Figure 28.9; the gradient of this graph equals  $1/pq$ . This graph gives the (scaled) time that has elapsed. Turning this graph around gives the more familiar graph of allele frequency against time,  $p(T)$ .



**FIGURE 28.8.** Plot of  $s(dt/dp) = 1/pq$ , the rate of time against allele frequency. The blue area under the curve is the total time taken to go from  $p_0$  to  $p_1$ .



**FIGURE 28.9.** Plot of the integral of  $1/pq$ , which is equal to  $s$  times the elapsed time (i.e.,  $T = st$ ).

$$\frac{dp_1}{dt} = f_1(p_1, p_2, \dots),$$

$$\frac{dp_2}{dt} = f_2(p_1, p_2, \dots),$$

⋮

⋮

⋮

where  $f_1, f_2$  represent some arbitrary relationship between the present state of the population (described by  $p_1, p_2, \dots$ ) and its rate of change. Even quite simple systems of this sort can show surprisingly complex behavior, including steady limit cycles or chaotic fluctuations.

The most important point to take from this section is a qualitative one: The timescale of evolutionary change is inversely proportional to the selection coefficient  $s$ . In the simplest case, where  $dp/dt = spq$  and  $s$  is a constant, a new measure of time,  $T = st$ , can be defined. Then,

$$\frac{dp}{dT} = \frac{dp}{d(st)} = \frac{1}{s} \frac{dp}{dt} = \frac{1}{s} (spq) = pq.$$

The allele frequency depends only on  $T$ ; thus, the same equation applies for any strength of selection provided that the timescale is adjusted accordingly.

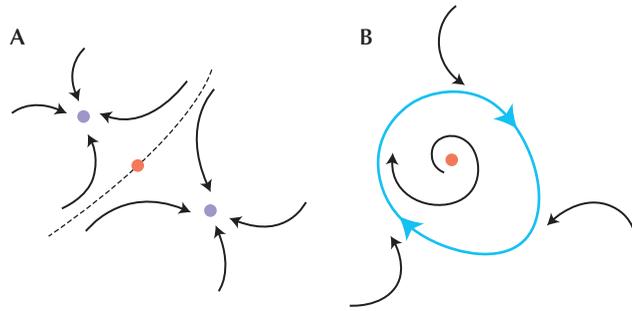
This can also be done in the more general case, where there is a set of equations with rates all proportional to some coefficient  $s$  ( $dp_1/dt = sf_1, dp_2/dt = sf_2, \dots$ , so that  $dp_1/dT = f_1, dp_2/dT = f_2, \dots$ ). Thus, all that matters is the scaled time  $T$ . For example, the pattern of change would look the same if the strength of selection were halved throughout but would be stretched over a timescale twice as long. Elsewhere in the book, when processes such as migration, mutation, and recombination are considered, we see that change occurs over a timescale set by the rates of the processes involved, and that the pattern itself only depends on the ratios between the rates of the different processes.

## Populations Tend to Evolve toward Stable Equilibria

It is hardly ever possible to find an exact solution to the differential equations that arise in evolutionary models. Today, it is straightforward to solve differential equations using a computer, provided that numerical values for the parameters are chosen. However, this is insufficient to understand the system fully, because often there will be too many parameters to explore the full range of behaviors. In any case, the goal usually is to do more than just make numerical predictions. We want to understand which processes dominate the outcome: Which terms matter? How do they interact? What is their biological significance? In this section, we will see how we set about understanding models in this way.

The first step is to identify which parameter combinations matter. For example, we have just seen that, in models of selection, only the scaled time  $T = st$  matters, not  $s$  or  $t$  separately. Often, this can drastically reduce the number of parameters that need to be explored. The next step is to identify equilibria—points where the population will remain—and find whether they are **locally stable**. In other words, if the population is perturbed slightly away from an equilibrium, will it move farther away or converge back to the equilibrium? There may be several stable equilibria, however, so where the population ends up depends on where it starts. In that case, each stable equilibrium is said to have a *domain of attraction* (Fig. 28.10A). Conversely, there may be *no* stable equilibrium, so that the population will never be at rest (Fig. 28.10B). However, identifying the equilibria and their stability takes us a long way toward understanding the full dynamics of the population.

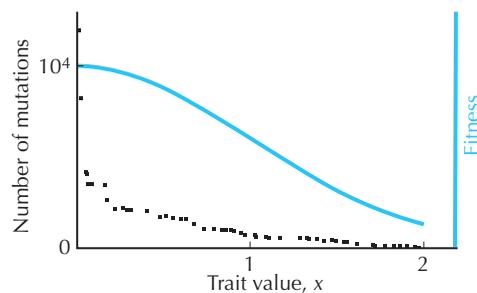
In population genetics, in the absence of mutation or immigration, no new alleles can be introduced. In that case, there will always be equilibria at which just one of the possible alleles makes up the entire population. At such an equilibrium, the allele is said to be **fixed** in the population. The crucial question, then, is whether the population is stable toward invasion from new alleles. This depends on whether the new allele tends to increase from a low frequency, which it will do if it has higher fitness than



**FIGURE 28.10.** (A) If there is more than one stable equilibrium, then where the population ends up depends on its initial state. The diagram shows two stable equilibria (*purple dots*) separated by an unstable equilibrium (*red dot*). The dotted line separates the **domains of attraction** of the two stable equilibria. (B) There may be no stable equilibria, in which the population will change continually. This diagram shows a stable **limit cycle** (*blue*), which encloses an unstable equilibrium (*red dot*).

the resident allele. Thus, analysis is especially simple if, most of the time, the population is fixed for a single genotype. In that case, the fate of each new mutation is examined to determine whether the mutation can invade to replace the previous allele.

Matters are more complicated if, when an allele invades, it sets up a stable polymorphism with the resident allele. Whether new alleles can invade this polymorphic state can still be determined, because such invasion depends only on whether or not a new allele has higher fitness than that of the residents. A substantial body of theory, known as **adaptive dynamics**, is used to analyze situations like this by making the additional assumption that invading alleles are *similar* to the resident (Fig. 28.11). This approach is best thought of as a way of exploring the dynamics when a range of alleles is available rather than as an actual model of evolution. (For one thing, when new alleles are similar in fitness to the residents, evolution will be very slow [Fig. 28.11].) In Chapter 20, we looked at other ways of finding what kinds of phenotypes will evolve. In particular, evolutionary **game theory** focuses on whether a resident phenotype is stable toward invasion by alternatives. Here, the analysis avoids making assumptions about the genetics by simply comparing the fitnesses of different phenotypes. All of these methods are guides to the direction of evolution rather than ways of making exact predictions.



**FIGURE 28.11.** **Adaptive dynamics** is a method for analyzing evolutionary models. This example represents the evolution of an asexual population, which varies in some continuous trait  $x$ . At low density, individuals have highest fitness if they are close to  $x = 0$  (*blue curve*). Initially, the population consists of poorly adapted individuals with trait value  $x = 2$  (*bottom right*). New alleles are introduced by mutation and differ slightly from the resident type. They can invade if they bring the phenotype closer to the optimum at  $x = 0$ . The *black dots* show how each successive mutation brings the phenotype closer to the optimum. Because mutations are assumed to be similar to the resident population and to substitute one by one, adaptation is slow: This diagram shows the outcome of more than  $10^4$  trial mutations.

## Stability Is Determined by the Leading Eigenvalue

How do we find the stability of a polymorphic equilibrium in which two alleles coexist? The simplest example is that of a single diploid gene, where the heterozygote is fitter than either homozygote. If, for example, the fitnesses of QQ, PQ, PP are 1:1+s:1, then the allele that is less common finds itself in heterozygotes more often and so tends to increase. In Box 17.1, it was shown that

$$\frac{dp}{dt} = -spq(p - q), \text{ where } p + q = 1.$$

There are equilibria where an allele is fixed ( $p = 0, 1$ ). To see whether the equilibrium at  $p = 0$  is stable toward the introduction of an allele, assume that  $p$  is very small, so that  $q$  is close to 1. Then,  $dp/dt \sim +sp$ , and a rare P allele will increase exponentially at a rate  $s$  (i.e.,  $p(t) = p(0)\exp(st)$ , for  $p \ll 1$ ). Similarly, a rare Q allele will increase as  $\exp(st)$ . There is also a polymorphic equilibrium where  $p = q = \frac{1}{2}$ , so that  $p - q = 0$ . To show that this is stable, suppose that allele frequency is perturbed slightly to  $p = \frac{1}{2} + \epsilon$ , where  $\epsilon$  is very small. Then,

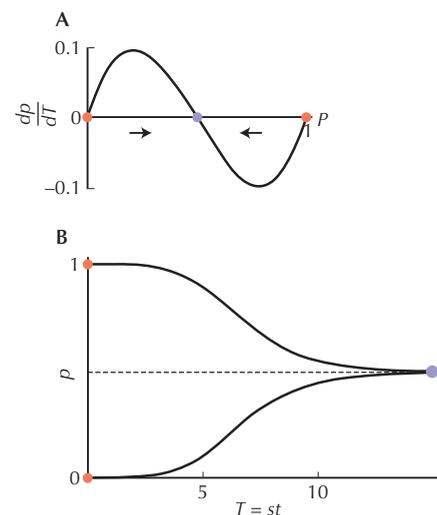
$$\frac{d\epsilon}{dt} = -s\left(\frac{1}{2} + \epsilon\right)\left(\frac{1}{2} - \epsilon\right)(2\epsilon) \sim -\frac{s}{2}\epsilon,$$

where we have ignored small terms such as  $\epsilon^2$ . Perturbations will decrease exponentially at a rate  $-s/2$ , so that  $\epsilon(t) = \epsilon(0)\exp(-st/2)$ . The polymorphism is, therefore, stable. Equations such as this, in which the variables change at rates proportional to their values, are termed **linear**. Because they do not include more complicated functions (such as  $\epsilon^2$  or  $\exp(\epsilon)$ ), they behave much more simply than the **nonlinear** equations that describe the full dynamics (here  $dp/dt = -spq(p - q)$ ).

In this simple, one-variable example, it can be easily seen, without going through the stability analysis, that a polymorphism must be stable (Fig. 28.12). With more variables, however, the outcome is not at all obvious, and a formal stability analysis becomes essential.

There are three key points (which also apply with slight modification in more complex situations) to take from this first example.

- When near to an equilibrium, the equations almost always simplify to a linear form:  $d\epsilon/dt = \lambda\epsilon$ . The rate of change,  $\lambda$ , is known as the **eigenvalue**.



**FIGURE 28.12.** If heterozygotes have a selective advantage  $s$  over either homozygote, a stable polymorphism is maintained. (A) The rate of change of allele frequency is  $dp/dT = -spq(p - q)$ . There are unstable equilibria at  $p = 0, 1$  (red dots) and a stable equilibrium at  $p = 0.5$  (purple dot); arrows show the direction of allele frequency change. (B) Allele frequencies converge on the stable equilibrium (dashed line at  $p = 0.5$ ). Time is scaled as  $T = st$ .

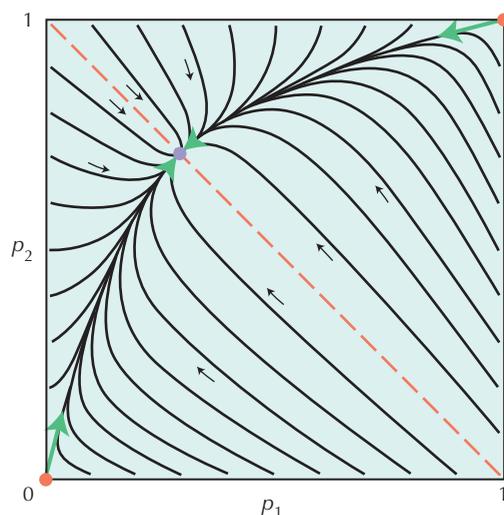
- Therefore, perturbations increase or decrease exponentially, as  $\exp(\lambda t)$ .
- Equilibria are unstable if  $\lambda > 0$ , so that perturbations grow, and are stable if  $\lambda < 0$ .

We now show how these points extend to an example involving two variables.

### Stability Analysis Extends to Several Variables

Suppose that there are two subpopulations, which exchange a proportion  $m$  of their genes per unit time. (Such subpopulations are referred to as **demes**; see p. 441.) An allele is favored in one deme by selection of strength  $s$  but reduces fitness by the same amount in the other deme. The entire system is described by two variables—namely, the allele frequencies in the two demes,  $p_1, p_2$ . In Box 28.3, it is shown that there are two equilibria at which one or the other allele is fixed throughout the whole population, and one equilibrium at which both demes are polymorphic. At each equilibrium, a pair of linear equations can be written that describes slight perturbations to the allele frequencies  $\varepsilon_1, \varepsilon_2$ . Solutions are sought in which these perturbations grow or shrink steadily but stay in the same ratio (e.g.,  $\varepsilon_1 = e_1 \exp(\lambda t)$  and  $\varepsilon_2 = e_2 \exp(\lambda t)$ ). *With two variables, there are two solutions of this form, with two values of  $\lambda$ .* Crucially, any perturbation can be written as a mixture of these two characteristic solutions. In the long run, whichever solution grows fastest (or decays the slowest) will dominate. This is just the solution with the largest value of  $\lambda$ , which is known as the **leading eigenvalue**. This characteristic behavior can be seen in Figure 28.13.

The patterns seen in this example of migration and selection in two demes do not quite exhaust the possibilities. Sometimes, *no* solution of the form  $\varepsilon_1 = e_1 \exp(\lambda t)$  can be found. In such cases, solutions have the form  $\varepsilon_1 = e_1 \exp(\lambda t) \sin(\omega t + \theta_1)$ , which correspond to oscillations with frequency  $\omega$  and grow or shrink at a rate  $\lambda$ . The larger value of  $\lambda$  still dominates, because it grows faster, and so the stability of the equilibrium is still determined by whether the larger  $\lambda$  is positive or negative. Even when the model as a whole involves complicated nonlinear interactions, near to its equilibrium it behaves in an approximately linear way. Thus, the standard linear analysis of the behavior near to equilibria tells us a lot about the overall behavior.



**FIGURE 28.13.** A stable polymorphism (purple dot) can be maintained when selection favors allele P in deme 2 but acts against it in deme 1. There are also two equilibria, corresponding to fixation of allele Q or allele P in both demes (red dots). The diagram shows how populations converge on the stable equilibrium (black lines). The green arrows show the **leading eigenvectors** at the equilibria: Populations tend to move away from the unstable equilibria, and toward the stable equilibrium, along these special directions. ( $M = m/s = 0.5$ .)

## Box 28.3 Migration and Selection in Two Demes

Two subpopulations (or demes) exchange migrants at a rate  $m$  per unit time. In deme 1, allele P is disadvantageous, with selection coefficient  $-s$ , and in deme 2, allele P is favorable, with selection  $+s$  (Fig. 28.14). The allele frequencies  $p_1$ ,  $p_2$  change at rates

$$\frac{dp_1}{dt} = -sp_1q_1 + m(p_2 - p_1),$$

$$\frac{dp_2}{dt} = +sp_2q_2 - m(p_2 - p_1).$$

These equations are an exact description of a haploid population reproducing continuously in time but are also a good approximation to a variety of discrete time models of diploids (see Box 16.1).

The problem can be simplified by dividing both sides by  $s$ , and scaling time as  $T = st$ . Then, it is evident that the solution depends only on the ratio between migration and selection,  $M = m/s$ :

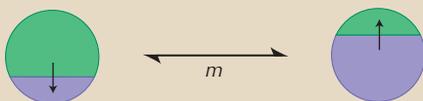
$$\frac{dp_1}{dT} = -p_1q_1 + M(p_2 - p_1),$$

$$\frac{dp_2}{dT} = +p_2q_2 - M(p_2 - p_1).$$

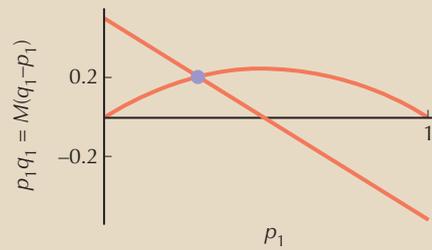
Clearly, there are equilibria where only one allele is present (i.e.,  $p_1 = p_2 = 0$  or  $p_1 = p_2 = 1$ ). There is also always a polymorphic equilibrium in which selection balances migration. Specifically, selection tends to fix allele Q in deme 1 and P in deme 2, whereas migration tends to make the allele frequencies equal (Fig. 28.13). By symmetry, at this equilibrium,  $q_2 = p_1$ . (In other words, the equilibrium must lie on the diagonal running from upper left to lower right in Fig. 28.13, indicated by a dotted line.) Setting the two equations to 0 and solving reveals that the polymorphic equilibrium occurs where  $p_1q_1 = M(p_2 - p_1) = M(q_1 - p_1)$ . This can be solved graphically (Fig. 28.15) or algebraically as

$$p_1^* = \frac{1}{2} + M - \sqrt{\frac{1}{4} + M^2}, \quad p_2^* = 1 - p_1^*,$$

where  $p_1^*$  and  $p_2^*$  are the equilibrium allele frequencies. Only the symmetric model, in which the selection coefficients in each deme are equal and opposite, is analyzed here. If they are different, and if migration is fast enough relative to selection, then whichever allele is favored overall will fix everywhere. However, if the selection coefficients are precisely opposite as above, there will be a polymorphic equilibrium even when migration is fast.



**FIGURE 28.14.** Allele Q (green) is favored in deme 1, and allele P (purple) is favored in deme 2. Arrows indicate the direction of selection. Genes are exchanged between the demes at a rate  $m$ .



**FIGURE 28.15.** The equation  $p_1q_1 = M(q_1 - p_1)$ , which gives the polymorphic equilibrium, can be solved by finding where the graphs for the left and right sides cross. Here,  $M = 0.5$ , and the equilibrium is at  $p_1 = 0.293$ .

To find the stability of the three equilibria, first suppose that  $p_1$  and  $p_2$  are small (lower left in Fig. 28.11). The equations then simplify to

$$\frac{dp_1}{dT} = -(1 + M)p_1 + Mp_2,$$

$$\frac{dp_2}{dT} = Mp_1 + (1 - M)p_2.$$

We look for a solution that grows or shrinks steadily, so that  $p_1 = e_1 \exp(\lambda T)$ ,  $p_2 = e_2 \exp(\lambda T)$ . Because  $dp_1/dT = \lambda e_1 \exp(\lambda T)$ , and similarly for  $p_2$ , the factors of  $\exp(\lambda T)$  cancel throughout and

$$\lambda e_1 = -(1 + M)e_1 + Me_2,$$

$$\lambda e_2 = Me_1 + (1 - M)e_2.$$

There are two possible solutions:  $\lambda = -M \pm \sqrt{1 + M^2}$ . The larger of these must be positive, because  $\sqrt{1 + M^2} > M$ . So, this equilibrium is unstable, because allele P can invade at a rate  $\lambda = -M + \sqrt{1 + M^2} > 0$ . (The green arrow at the bottom left of Figure 28.13 shows this characteristic solution.) The allele invades deme 2 faster than it invades deme 1, because it is favored there; thus, the arrow points upward. A similar argument shows that the equilibrium with P fixed is unstable to invasion by Q (upper right in Fig. 28.13).

The polymorphic equilibrium can be analyzed in a similar way, by letting  $\epsilon_1 = p_1 - p_1^*$  and  $\epsilon_2 = p_2 - p_2^*$  be small deviations from equilibrium. Discarding small terms such as  $\epsilon^2$  and  $\epsilon^3$  gives

$$\frac{d\epsilon_1}{dT} = (p_1^* - q_1^* - M)\epsilon_1 + M\epsilon_2,$$

$$\frac{d\epsilon_2}{dT} = +M\epsilon_1 + (q_2^* - p_2^* - M)\epsilon_2.$$

By searching for solutions of the forms  $p_1 = e_1 \exp(\lambda T)$  and  $p_2 = e_2 \exp(\lambda T)$ , again an equation for  $\lambda$  is obtained, which has two solutions. The larger is  $\lambda = -s(q_1^* - p_1^*)$ ; because  $q_1^* > p_1^*$ , the polymorphic equilibrium is stable. As they converge toward this stable equilibrium at a rate  $\lambda$ , populations tend to follow this solution (the **leading eigenvector**), as indicated by the pair of green arrows in Figure 28.13 (upper left).

## RANDOM PROCESSES

## Our Understanding of Random Events Is Relatively Recent

Most mathematics is long established. The fundamentals of geometry were worked out at least three millennia ago by the Babylonians, for practical surveying. The Greeks developed both geometry and logic into a rigorous formal system by around 500 B.C. Algebra, the representation of numbers by symbols, was established by Arab mathematicians during the 9th and 10th centuries A.D., and calculus was discovered independently by Gottfried Leibniz and Isaac Newton in the mid-17th century. Yet, the most elementary notions of probability were established much more recently. Probability is a subtle concept; it can be interpreted in a variety of ways and is generally less well understood than most of the mathematics in common use. Some key ideas in basic probability are summarized in Box 28.4.

Probability did not become an integral part of physics until the 1890s, when Ludwig Boltzmann showed how gross thermodynamic properties such as temperature and entropy could be explained in terms of the statistical behavior of large populations of molecules. Even then, random fluctuations themselves were not analyzed until 1905, when Albert Einstein showed that Brownian motion, which is the random movement of small particles, is due to collisions with small numbers of molecules (Fig. 28.16).

Arguably, most developments in probability and statistics have been driven by biological—especially, genetic—problems. Francis Galton introduced the ideas of re-

## Box 28.4 Basic Probability

**Notation:**

$\text{Prob}(A)$ , $\text{Prob}(B)$	probability of events $A$ , $B$ .
$f_i$	probability that $i = 0, 1, 2, \dots$ .
$f(x)$	probability density of a continuous variable $x$ .

**Probabilities sum to 1:**

$$\sum_{i=0}^{\infty} f_i = 1 \quad \text{or} \quad \int_{-\infty}^{\infty} f(x) dx = 1.$$

**Probabilities of exclusive events sum:**

$\text{Prob}(A \text{ or } B) = \text{Prob}(A) + \text{Prob}(B)$  if  $A$  and  $B$  cannot both occur.

**Probabilities of independent events multiply:**

$\text{Prob}(A \text{ and } B) = \text{Prob}(A)\text{Prob}(B)$  if the chance of  $A$  occurring is *independent* of whether  $B$  occurs, and vice versa.

**Expectation:**

$$E[y] = \sum_{i=0}^{\infty} y_i f_i \quad \text{or} \quad \int_{-\infty}^{\infty} y(x) f(x) dx$$

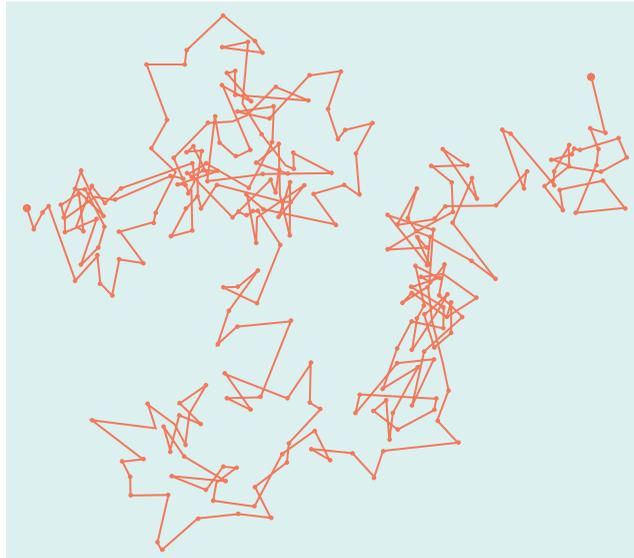
is the **expected value** of  $y$ ; it is an average over the probability distribution  $f$ .

**Mean and variance:**

The **mean** of  $y$  is  $E[y]$ , sometimes written  $\bar{y}$ .

The **variance** of  $y$  is  $E[(y - \bar{y})^2]$  or  $\text{var}(y)$ .

The **standard deviation** of  $y$  is  $\sqrt{\text{var}(y)}$ .



**FIGURE 28.16.** Brownian motion was discovered by Robert Brown in 1827 while he was observing pollen particles by microscopy. Albert Einstein in a famous 1905 paper (and three subsequent papers) presented the laws governing Brownian motion as a way to confirm the existence of atoms.

gression and correlation in order to understand inheritance, and Karl Pearson developed these as a tool for analysis of quantitative variation and natural selection (p. 21). The **analysis of variance**—key to separating and quantifying the causes of variation—was first used by R.A. Fisher in 1918 to describe the genetic variation that affects quantitative traits (p. 393).

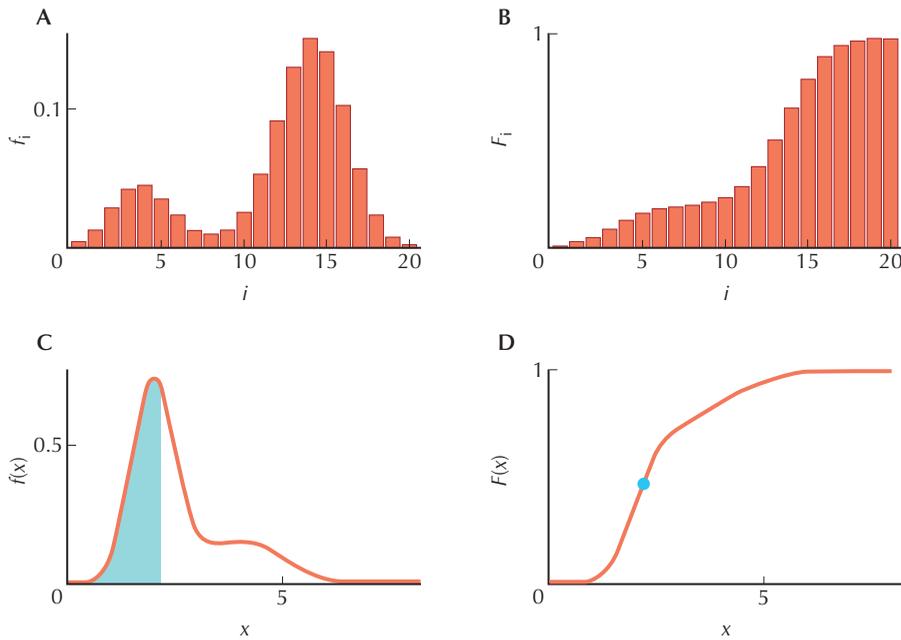
Even today, new concepts in probability theory, and new statistical methods, have been stimulated by biological problems. These include Motoo Kimura’s use of the diffusion approximation in his **neutral theory** of molecular evolution (pp. 59–60 and 425–427). More recently, coalescent models have provided us with a different way of describing the same process of neutral diffusion (pp. 421–425), in terms of the ancestry of genetic lineages. The fundamental importance of random processes in evolution was explained in Chapter 15, and examples of inference from genetic data are given in Chapter 19. Here, basic concepts of probability are reviewed, and some important random processes are described.

### A Random Process Is Described by Its Probability Distribution

A **probability distribution** lists the chance of every possible outcome. These may be discrete (e.g., numbers of mutations), or there may be a continuous range of possibilities (e.g., allele frequencies lying between 0 and 1). For a discrete probability distribution, the total must sum to 1. For a continuous distribution  $f(x)$ , the chance of any particular value is infinitesimal, but the area under the curve gives the probability that the outcome will lie within some range (Fig. 28.17). Using the notation of integral calculus,

$$\int_a^b f(x)dx$$

is the chance that  $x$  will lie between  $a$  and  $b$ .



**FIGURE 28.17.** (A) An example of a discrete probability distribution  $f_i$  on the range  $i = 0, \dots, 20$ . (B) This can also be represented as a **cumulative distribution**  $F_i$ , which is the probability of drawing a value less than or equal to  $i$ . (C) The area under a **probability density**  $f(x)$  gives the probability that a continuous variable  $x$  lies within some range of values. For example, the *blue shaded area* gives the chance that  $x < 2.2$  as

$$\int_{-\infty}^{2.2} f(x) dx = 0.47.$$

(D) The cumulative distribution  $F(x)$  gives the probability that the value is smaller than  $x$ ; the point at  $x = 2.2$ ,  $F(x) = 0.47$  corresponds to the shaded area in C.

The distribution of a continuous variable is sometimes called a **probability density**. Again, the total area must sum to 1; that is,

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

(Fig. 28.17).

The overall location of a distribution can be described by its **mean**. This is simply the average value of the random variable  $x$ , which is sometimes written as  $\bar{x}$  and sometimes as the **expectation**,  $E[x]$  (see Box 28.4). The spread of the distribution is described by the **variance**, which is the average of squared deviations from the mean ( $\text{var}(x) = E[(x - \bar{x})^2]$ ). The variance has units of the square of the variable (e.g., meters<sup>2</sup> if  $x$  is a length). Thus, the spread around the mean is often expressed in terms of the **standard deviation**, equal to the square root of the variance, and with the same units as the variable itself.

### Discrete Distributions

Box 28.5 summarizes the properties of the most important distributions. The simplest cases have only two possible outcomes, such as a coin toss giving heads or tails or one of two alleles being passed on at meiosis. This distribution has often been used when

looking at the chance of either of two alleles being sampled from a population, with probability  $q$  or  $p$ . If the two outcomes score as 0 or 1, then the mean score is  $p$ , and the variance in score is  $pq$ .

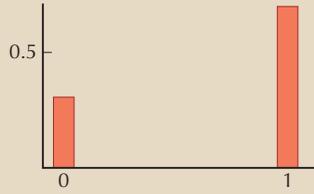
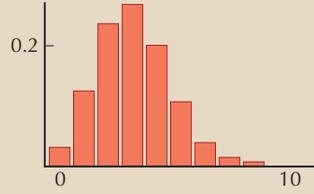
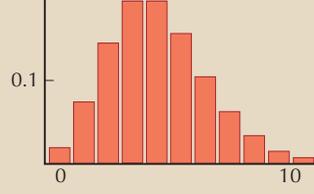
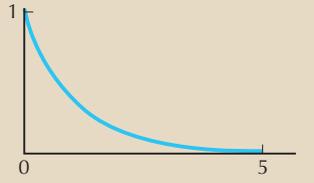
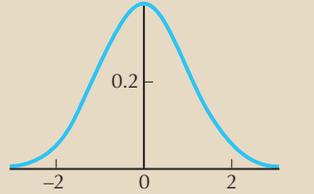
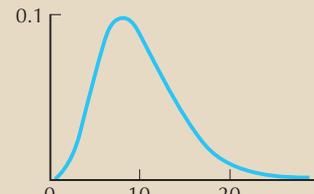
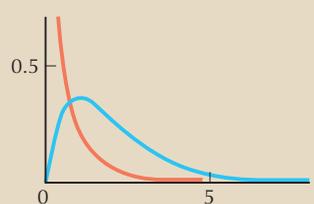
A natural extension of this two-valued distribution is the **binomial distribution**. This gives the distribution of the number of copies of one type out of a total of  $n$  sampled items. It is the distribution of the sum of a series of independent draws from the two-valued distribution, each with probabilities  $q, p$ . (This is the basis of the **Wright–Fisher model**; see p. 416.) It must be assumed that either the population is very large or each sampled gene is put back before the next draw (“sampling with replacement”). Otherwise, successive draws would not be independent (sampling a P would make sampling more Ps less likely). The mean and variance of a sum of independent events are just the sums of the means and variances of the component means and variances. Therefore, a binomial distribution has mean and variance  $n$  times those of the two-valued distribution just discussed,  $np$  and  $npq$ , respectively. The typical deviation from the mean (measured by the standard deviation) is  $\sqrt{npq}$ , which increases more slowly than  $n$ . Therefore, the binomial distribution clusters more closely around the mean as the numbers sampled ( $n$ ) increase. The binomial distribution approaches a normal distribution with mean  $np$  and variance  $npq$  as  $n$  gets large.

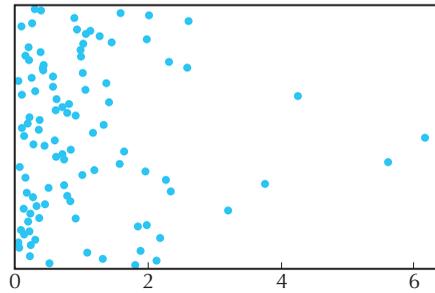
The **Poisson distribution** gives the number of events when there are a very large number of opportunities for events, but each is individually rare. The classic example is the number of Prussian officers killed each year by horse kicks, described by Ladislaus Bortkiewicz in his 1898 book *The Law of Small Numbers*. Over the course of 20 years, Bortkiewicz collected data on deaths from horse kicks occurring in 14 army corps, over the course of 20 years, and found that the overall distribution of the deaths followed a Poisson distribution. Another example of the Poisson distribution is the number of radioactive decays per unit time from a mass containing very many atoms, each with a very small chance of decaying. For example, a microgram of radium 226 contains  $2.6 \times 10^{15}$  atoms and is expected to produce 37,000 decays every second; the actual number follows a Poisson distribution with this expectation. Similarly, the number of mutations per genome per generation follows a Poisson distribution, because it is the aggregate of mutations that may occur at a very large number of bases, each with a very low probability (p. 531). The Poisson distribution is the limit of a binomial distribution when  $n$  is very large and  $p$  is very small. It has mean  $np = \lambda$ , and the variance is equal to this mean (because  $npq \sim np = \lambda$  when  $p$  is small). Like the binomial, it converges to a Gaussian distribution when  $\lambda$  is large. We very often need to know the chance that *no* events will occur. Under a binomial distribution, this is  $(1 - p)^n$ . As  $n$  becomes large, this tends to  $\exp(-np) = \exp(-\lambda)$ , which is the probability of no events occurring under a Poisson distribution. For example, the chance that there are no mutations in a genome when the expected number of mutations is  $\lambda = 5$  is  $\exp(-5) \sim 0.0067$ .

### Box 28.5 Common Probability Distributions

Table 28.1 defines commonly used probability distributions. The parameters used in the plots on the right-hand column of the table are two-valued,  $p = 0.7$ ; binomial,  $n = 10$ ,  $p = 0.3$ ; Poisson,  $\lambda = 4$ ; exponential,  $\lambda = 1$ ; normal,  $\bar{x} = 0$ ,  $\sigma^2 = 1$ ; chi-square,  $n = 10$ ; and Gamma:  $\beta = 2$ ,  $\alpha = 0.5$  (red),  $\alpha = 2$  (blue). The factorial  $1 \times 2 \times 3 \times \dots \times n$  is written as  $n!$ .  $\Gamma(x)$  is the gamma function;  $\Gamma(n + 1) = n!$ . Note that the chi-square distribution is a special case of the Gamma distribution.

**TABLE 28.1.** Common probability distributions

Distribution	Mean	Variance	
Discrete			
Two-valued $f_0 = q, f_1 = p$	$p$	$pq$	
Binomial $\frac{n!}{i!(n-i)!} q^{n-i} p^i$	$np$	$npq$	
Poisson $e^{-\lambda} \frac{\lambda^i}{i!}$	$\lambda$	$\lambda$	
Continuous			
Exponential $\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	
Gaussian (or normal) $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right)$	$\bar{x}$	$\sigma^2$	
Chi-square $\frac{1}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{(n/2)-1} e^{-x/2}$	$n$	$2n$	
Gamma $\frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta}$	$\alpha\beta$	$\alpha\beta^2$	



**FIGURE 28.18.** Exponentially distributed variables have a wide range. In this sample of 100, values range between 0.008 and 6.2, even though the expectation is 1.

### Continuous Distributions

We turn now to distributions of continuous variables. The time between independent events, such as radioactive decays or the time between mutations in a genetic lineage, follows an **exponential distribution** (Fig. 28.18). This distribution has the special property that the expected time to the next event is the same, regardless of when the previous event occurred. (This follows from the assumption that events occur independently of each other.) The exponential distribution has a very wide range; specifically, 1% of the time, events occur closer together than 1% of the average interval, and 1% of the time, they take more than 4.6 times the average. This distribution is closely related to a Poisson distribution. If the rate of independent events is  $\alpha$ , the number within a total time  $t$  is Poisson distributed with mean  $\alpha t$ .

In contrast, a **normal** or **Gaussian distribution** is tightly clustered around the mean. (There is a less than 4.6% chance of deviating by more than 2 s.d., less than 0.27% of deviating by more than 3 s.d., less than  $6.4 \times 10^{-5}$  of deviating by more than 4 s.d., and so on.) We saw in Figure 14.2 that many biological traits follow a normal distribution. Later in this chapter, we will see why this distribution appears so often. A closely related distribution is the **chi-square distribution**, which is the distribution of the sum of a number  $n$  of squared Gaussian variables,  $z_1^2 + z_2^2 + \dots + z_n^2$  and is written as  $\chi_n^2$ . This determines the distribution of the variance of a sample of  $n$  values.

There are many other continuous distributions. One that appears quite often is the **Gamma distribution**, which has two parameters ( $\alpha$ ,  $\beta$ ). It is useful in modeling because it gives a way of varying the distribution from a tight cluster ( $\alpha$  large) to a broad spread ( $\alpha$  small). The time taken for  $k$  independent events to occur, each with rate  $\mu$ , is gamma distributed, with parameters  $\alpha = k$ ,  $\beta = 1/\alpha$ .

Only distributions of a single variable (either discrete or continuous) have been discussed. But distributions of multiple variables will also be encountered, which may be correlated with each other, for example, the distribution of age and body weight. The most important distribution here is the multivariate Gaussian, which has the useful property that the distribution of each variable considered separately, and of any linear combination of variables (e.g.,  $2z_1 + 3z_2$ , say), is also Gaussian (Box 28.5). This distribution is especially important for describing variation in quantitative traits (pp. 386–387).

### Probability Can Be Thought of in Several Ways

We have explained how probabilities can be described and manipulated. But what exactly is a “probability”? The simplest case is when outcomes are equally probable. For example, the chance that a fair coin will land heads up or tails up is always 1/2, and the chance that a fair die will land on any of its six faces is 1/6. Probability calculations often rest on this simple appeal to symmetry. For example, a fair meiosis is similar to a fair coin toss, with each gene having the same chance of being passed on; and if the

total mutation rate is 1 per genome per generation, then the probability of mutation is  $1/(3 \times 10^9)$  per generation at each of  $3 \times 10^9$  bases. However, matters are rarely so simple: Meiotic transmission can be distorted (pp. 587–589), and mutation rates vary along the genome (p. 347). Often, we would like to talk about the probability of events that are not part of any set of symmetric possibilities. What is the chance that the average July temperature will be hotter than 20°C? What are the chances that some particular species will go extinct? What is the chance that some particular disease allele will spread?

One possibility is to define probability as the frequency of an event in a long series of trials or among a set of replicates carried out under identical conditions. This is clear enough if we think of a coin toss or meiotic segregation, because we could imagine repeating the random event any number of times, under identical conditions. However, this is rarely possible in practice (or even in principle). To find the distribution of July temperatures, we could record temperatures over a long series of summers, although there might be good reason to think that the conditions under which the readings were taken are not equivalent, because, for example, the Earth's orbit, Sun's brightness, and atmospheric composition fluctuate systematically. The goal is to use this information to find the probability distribution of temperature in one *particular* summer. Similarly, to answer the other two questions, it is necessary to know the probability of survival of some particular species or allele, which may not be equivalent to any others.

Another problem with using the long-run frequency to define probability is that, given some particular sequence of events, a way to judge the plausibility of the hypothesis is needed. For example, if the sequence is 480 heads and 520 tails, we still need to be certain that the coin is fair (i.e., that the probability of heads = 1/2). There are other possibilities: for example, the probability of heads might be 0.48, or the frequency of heads might increase over the course of the sequence. In real-life situations, we are frequently faced with distinguishing alternative hypotheses on the basis of only one set of data. Invoking hypothetical replicates does not avoid the difficulty, because judgments must still be made on the basis of a finite (albeit larger) dataset.

An apparently quite different view is that probability expresses our degree of belief. This allows us to talk about probabilities not only of events but also of hypotheses. For example, what is the probability that there is life on Mars, or that introns evolved from transposable elements? It is hardly practical to define probability as a psychological measure of actual degree of belief, because that would vary from person to person and from time to time. Instead, probability is seen as the “rational degree of belief” and can be defined as the odds that a rational person would use to place a bet, so as to maximize his or her expected winnings. But put this way, it is either equivalent to the definition based on long-run frequency just discussed (because the expected winnings depend on the long-run frequency) or it is still subjective (because it depends on the information available to each person). These different interpretations of probability influence the way statistical inferences are made.

We (i.e., the authors) favor the view that probability is an objective property, which can be associated with unique events. We calculate probabilities from models. For example, a model of population growth that includes random reproduction gives us a probability of extinction. A model of the oceans and atmosphere gives us the probability distribution of temperature or rainfall in the future. With this view, what is crucial is how hypotheses are distinguished from one another—that is, distinguishing between different models that predict our observations with different probabilities. This issue is explored in more depth in the Web Notes (online). We take the view that the simplest approach is to prefer hypotheses that predict the actual observations with the highest probability.

### Probabilities Can Be Assigned to Unique Events

To give a concrete example of the types of probability arguments that arise in practice, we take as an example predictions of the effects of climate change on specific

events. The summer of 2003 was exceptionally hot across the Northern Hemisphere and especially so in Western Europe (Fig. 28.19A). As a result, an additional 22,000–35,000 people died during August 2003, a figure estimated from a sharp peak in mortality rates (Fig. 28.19B); around Paris, there was a major crisis as mortuaries overflowed. In addition, considerable economic damage resulted from crop failures and water shortages. Now, climate is highly variable, and we cannot attribute this particular event to the general increase in global temperatures over the past century, and nor can we say with certainty that it was caused by the greenhouse gases produced by human activities that are largely responsible for that trend. However, we can make fairly precise statements about the probability that such high temperatures would occur under various scenarios.

Models of ocean and atmosphere can be validated against past climate records, so that a narrow range of models is consistent with physical principles and with actual climate. These models predict a distribution of summer temperatures. The summer of 2003 was unlikely, but not exceptionally so. Given current rates of global warming, such hot summers will become common by about 2050. As the distribution shifts toward higher average temperatures, and its variability increases, the chances of extreme events greatly increase (Fig. 28.19C,D). The climate models can be run without including the influence of greenhouse gases and show that the probability of extremely high temperatures would then be much lower. Overall, about half the risk of a summer as extreme as that of 2003 can be attributed to human pollution. It is calculations such as these—estimating the probability of unique events—that lay the legal basis for the responsibility of individual polluters. The point here is that we can and do deal with the probabilities of single, unique events.

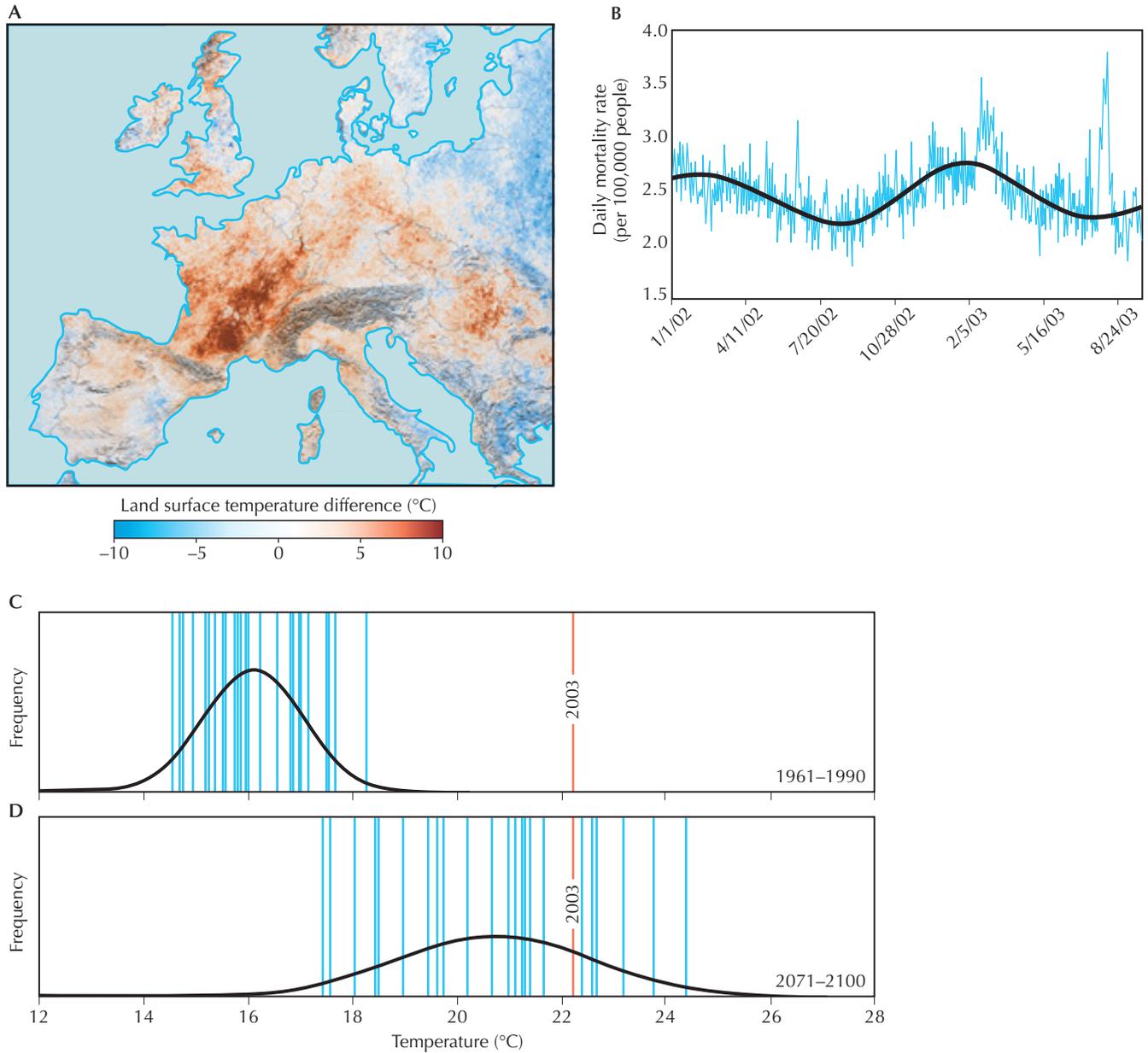
### The Increase of Independently Reproducing Individuals Is a Branching Process

This chapter is concerned with models of reproducing populations. The simplest such model is one in which individuals reproduce *independently* of each other. In other words, each individual produces 0, 1, 2, ... offspring, with a distribution that is independent of how many others there are and how they reproduce. Mathematically, this is known as a **branching process**, and it applies to the propagation of genes, surnames, epidemics, and species, as long as the individuals involved are rare enough that they do not influence each other. This problem was discussed by Galton, who wished to know the probability that a surname would die out by chance. However, the basic mathematics applies to a wide range of problems.

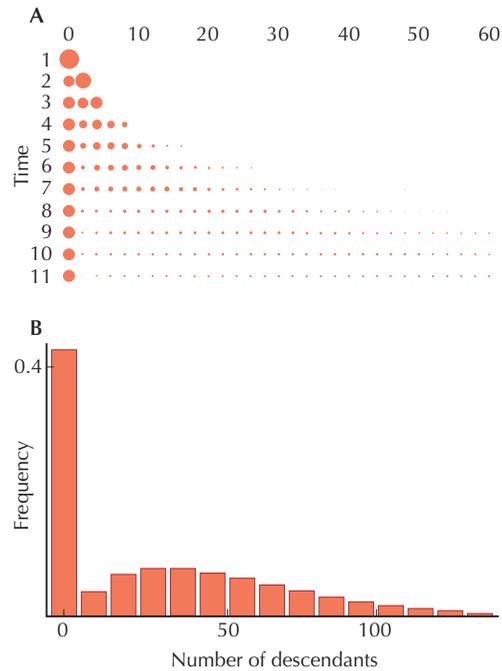
It is simplest to assume that individuals reproduce only once, and in synchrony, so that time can be counted in discrete generations. Each individual leaves 0, 1, 2, ... offspring with probability  $f_0, f_1, f_2, \dots$ . Clearly,  $f_0 + f_1 + f_2 + \dots = 1$ , and the mean offspring number (i.e., the mean fitness) is

$$\bar{k} = \sum_{k=0}^{\infty} k f_k.$$

On average, the population is expected to grow by a factor  $\bar{k}$  every generation, and by  $\bar{k}^t$  over  $t$  generations. Indeed, the size of a very large population would follow the smooth and deterministic pattern of geometric increase as modeled above (see Reproducing Populations Tend to Grow Exponentially section of this chapter). However, this average behavior is highly misleading when numbers are small. The number of offspring of any one individual has a very wide distribution, with a high chance of complete extinction (Fig. 28.20).



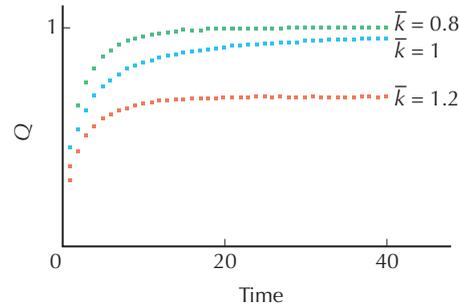
**FIGURE 28.19.** (A) Land surface temperatures for summer 2003, relative to summers 2000–2004. (B) Daily mortality rate in the German state of Baden-Württemberg; the red line shows the seasonal cycle, with higher mortality in winter. The graph shows the extra mortality due to an influenza outbreak in February–March 2003 and the sharp peak due to the heat wave of August 2003. That heat wave caused an additional 900–1300 deaths, out of 10.7 million people. (C) Distribution of summer temperatures in northern Switzerland, under a climate model representing conditions in 1961–1990. The exceptionally hot summer of 2003 is shown by a red bar. (D) The same, but for predicted conditions in 2071–2100, assuming current rates of increase in greenhouse gases. (A, Redrawn from Allen M.R. and Lord R. 2004. The blame game. *Nature* **432**: 551–552. B, Redrawn from Box 1 figure in Schar C. and Jendritsky G. 2004. Hot news from summer 2003. *Nature* **432**: 559–560. C, Modified from Fig. 3a,b in Schar C. et al. 2004. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**: 332–336.)



**FIGURE 28.20.** A branching process leads to a broadly spread distribution. In each generation, there is a 30% chance that an individual dies and a 70% chance that it divides, leaving two offspring. The expected number of offspring is 1.4, and so after ten generations, an individual is expected to leave  $1.4^{10} \sim 30$  descendants. However, the distribution becomes increasingly widely spread. Either there are no descendants (probability  $0.3/0.7 = 43\%$ ) or there are likely to be a large number. (A) The distribution of numbers of descendants over time. The areas of the circles are proportional to probability. (The diagram is cut off at the right, at 60 descendants.) (B) The distribution of the number of descendants after ten generations.

What is the probability that after  $t$  generations, one individual will leave no descendants ( $Q_t$ ) or some descendants ( $P_t$ ; necessarily  $Q_t + P_t = 1$ )? Initially, this is  $Q_0 = 0$ , because one individual must be present at the start ( $t = 0$ ). The chance of loss after one generation is the chance that there are no offspring in the first generation, that is,  $Q_1 = f_0$ . To find the solution for later times, sum over all possibilities in the first generation. There is a chance  $f_0$  of no offspring, in which case loss by time  $t$  is certain. There is also a chance  $f_1$  of one offspring, in which case loss by time  $t$  is the same as the chance that this one offspring leaves no offspring after  $t - 1$  generations. More generally, if there are  $k$  offspring, the chance that all  $k$  will leave no offspring after the remaining  $t - 1$  generations is just  $Q_{t-1}^k$ . Thus, a recursion allows us to calculate the chance of loss after any number of generations:

$$\begin{aligned}
 Q_0 &= 0, \\
 Q_1 &= f_0, \\
 Q_2 &= f_0 + f_1 Q_1 + f_2 Q_1^2 + f_3 Q_1^3, \\
 Q_3 &= f_0 + f_1 Q_1 + f_2 Q_2^2 + f_3 Q_3^3, \\
 &\vdots \\
 &\vdots \\
 Q_t &= \sum_{k=0}^{\infty} f_k Q_{t-1}^k.
 \end{aligned}$$



**FIGURE 28.21.** The probability that a gene will leave no descendants,  $Q$ , increases toward 1 if its average fitness is not greater than 1 ( $\bar{k} = 0.8, 1$ ). If the fitness is greater than 1, then the probability of loss converges to a value less than 1. For fitness  $\bar{k} = 1.2$ ,  $Q$  tends to 0.68, so that there is a 32% chance that the gene leaves some descendants.

This solution is shown in Figure 28.21, assuming a Poisson distribution of offspring number. If the mean fitness is less than or equal to 1 ( $\bar{k} \leq 1$ ), extinction is certain. However, if the mean fitness is greater than 1, there is some chance that the lineage will establish itself, to survive and grow indefinitely. To find the chance of ultimate survival, find when  $Q_t$  converges to a constant value, so that  $Q_t = Q_{t-1} = Q$ . This gives an equation for  $Q$  that can be solved either using a computer or graphically (Fig. 28.22). With binary fission, for example, an individual either dies or leaves two offspring (i.e.,  $f_0 + f_2 = 1$ ). Thus,

$$Q = f_0 + f_2 Q^2 \quad \text{so that } Q = \frac{f_0}{f_2},$$

where the solution is only valid ( $Q < 1$ ) if  $f_0 < f_2$ , which must be true if the population is to grow at all. If the distribution of offspring number is Poisson, with mean  $\lambda > 1$ , then

$$Q = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} Q_{t-1}^k = \exp(-\lambda(1-Q)),$$

where the fact that

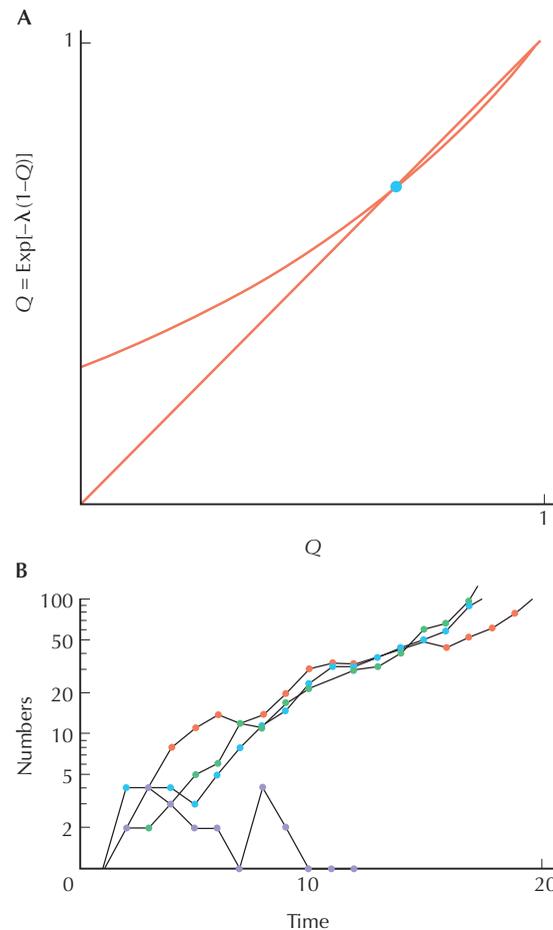
$$\sum_{k=0}^{\infty} \frac{\lambda Q^k}{k!} = \exp(\lambda Q)$$

is used. The solution to this equation is shown graphically in Figure 28.22. For a growth rate of  $\lambda = 1.2$ , the probability of ultimate loss is  $Q \sim 0.686$ .

Usually, our concern is with the fate of alleles that have a slight selective advantage ( $\bar{k} = 1 + s$ , where  $s \ll 1$ ). There is a simple approximation that does not depend on the (usually unknown) shape of the distribution of offspring number,  $f_k$ :

$$P = \frac{2s}{\text{var}(k)}.$$

This remarkably simple result shows that when an allele has a small selective advantage  $s$ , its chance of surviving is twice that advantage divided by the variance in numbers of offspring,  $\text{var}(k)$ . (With a Poisson number of offspring,  $\text{var}(k)$ , the variance of offspring number is equal to its mean, which is  $\bar{k} = 1$  for a stable population; thus,  $P = 2s$ .) The consequences of this relation are considered on pages 489–490.



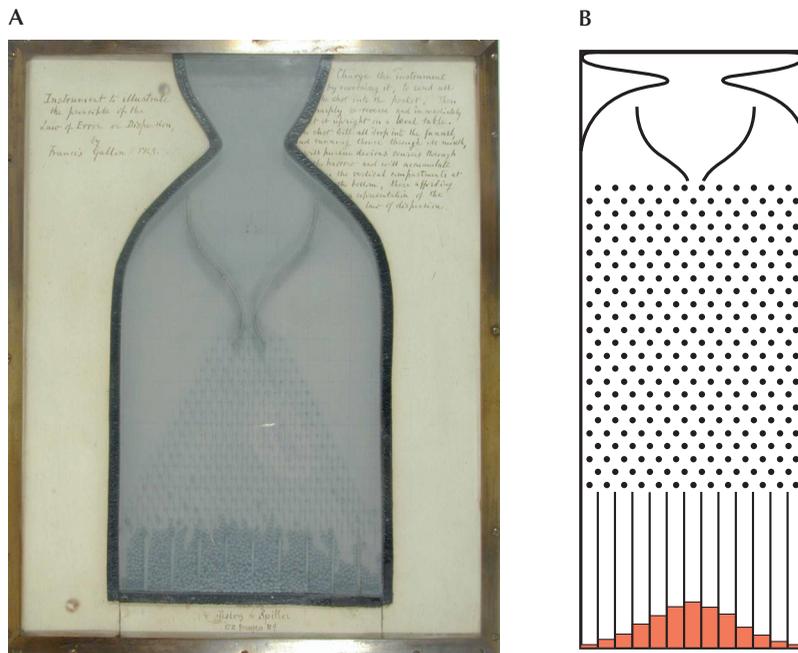
**FIGURE 28.22.** (A) The probability of ultimate extinction can be found graphically. The *diagonal line* plots  $Q$ , and the *curve* plots  $\text{Exp}[-\lambda(1 - Q)]$  for  $\lambda = 1.2$ . The solution to  $Q = \text{Exp}[-\lambda(1 - Q)]$  is the point where the curves cross, at  $Q = 0.686$ . (B) In ten replicates of the process, six went extinct within four generations. The remaining four are plotted here; three survived indefinitely. (Numbers are plotted on a log scale, so that exponential growth appears as a straight line.)

### The Sum of Many Independent Events Follows a Normal Distribution

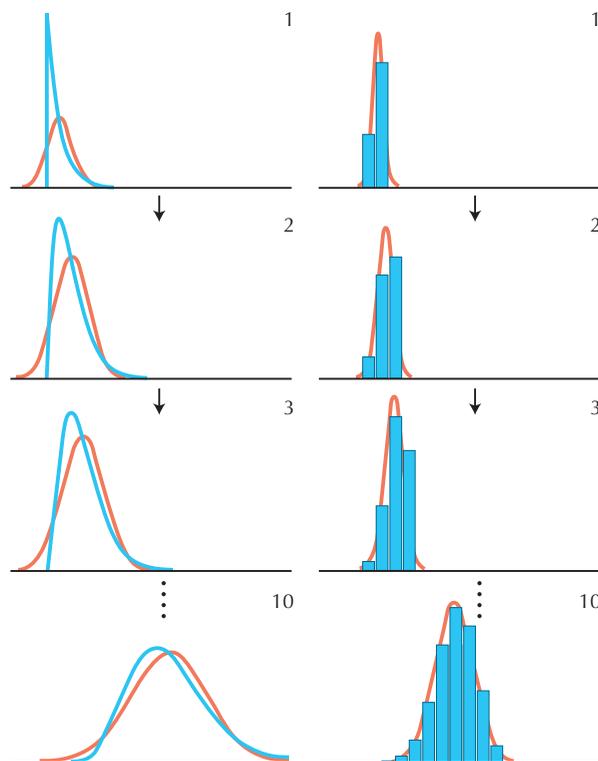
The phenotype of an organism is the result of contributions from many genes and many environmental influences; the location of an animal is the consequence of the many separate movements that it makes through its life; and the frequency of an allele in a population is the result of many generations of evolution, each contributing a more or less random change. To understand each of these examples quantitatively, the combined effects of many separate contributions must be considered. Often, the overall value can be treated as the *sum* of many *independent* effects. If no effect is too broadly distributed, then the sum follows an approximately Gaussian distribution. This is known as the **Central Limit Theorem** and is illustrated in Figure 28.23 by an ingenious device designed by Galton.

Figure 28.24 illustrates this convergence to a Gaussian distribution and shows the distribution of each individual random effect. Although the distributions are initially quite different, and far from Gaussian, the distribution of their sum becomes smoother and converges rapidly to a Gaussian distribution. Necessarily, the distribution of the sum of two or more Gaussians is itself a Gaussian.

The mean of the overall distribution is the sum of the separate means, and the variance of the overall distribution is the sum of the separate variances. Because the shape of a Gaussian distribution is determined solely by its mean and variance, all



**FIGURE 28.23.** (A) Galton's "pinball" machine, also called Galton's quincunx, was devised to illustrate the convergence to a normal distribution. Written on the face of the device by Galton is "Instrument to illustrate the principle of the Law of Error on Dispersion." (B) Rendering of Galton's quincunx based on an illustration in his book *Natural Inheritance*. (A, Courtesy of the Galton Collection, University College London. Photo GALT063. <http://www.ucl.ac.uk/silva/museums/galton/collections/highlights/statistics/GALT063>. B, Modified from Fig. 7, p. 63 in Galton F. 1889. *Natural Inheritance*, MacMillan, New York and London.)



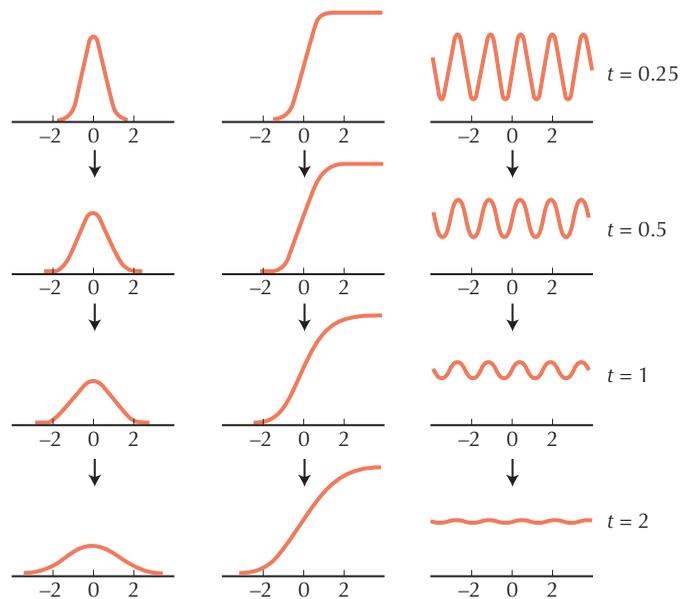
**FIGURE 28.24.** The distribution of the sum of several independent random variables converges to a normal (i.e., Gaussian) distribution. (Top left panel) Exponential distribution (blue) together with a Gaussian distribution with the same mean and variance (red). (Left column) The blue curves show the distribution of the sum of 2, 3, ..., 10 exponentially distributed variables. (Right column) The same, but for a two-valued discrete distribution. In both cases, the distribution of ten independent variables fits well to a Gaussian curve.

other features of the component distributions are irrelevant. This is an important simplification, because it means that the process can be described in terms of just two variables—namely, the mean and variance of the component distributions. This is used elsewhere in the book to simplify our description of quantitative variation (p. 385) and that of the flow of genes through populations (Fig. 16.3).

### Random Walks Can Be Approximated by Diffusion

The sum of a large number of independent random variables tends to a Gaussian distribution. We can think of this convergence as occurring over time. For example, consider an individual moving in a random walk so that its position changes by a random amount every generation. At any time, it will be at a random position drawn from a distribution that converges to a Gaussian. (To be more concrete, think instead of a population of randomly moving individuals and follow the distribution of all of them.) In this section, we show how the probability distribution changes approximately continuously in time, and can be described by a **diffusion equation**. This parallels our discussion of deterministic models (see the Deterministic Processes section in this chapter), where it was shown that a variety of different models of evolution can be approximated by a differential equation in continuous time. Similarly, a range of random models can all be approximated by a diffusion of the probability distribution. This gives a powerful method for understanding the interaction between different evolutionary processes.

Consider a variable  $x$ , which has distribution  $\psi(x, t)$  at time  $t$ . If the changes in each time step are small and have constant mean  $M$  and variance  $V$ , then the distribution spreads out according to a differential equation that gives the rate of change of the distribution,  $d\psi/dt$  (Box 28.6). If the population starts out concentrated at a single point (or the position of a single individual that starts at that point is considered) then the solution is a Gaussian distribution with mean  $M_t$  and variance  $V_t$  (Fig. 28.25, left column). Other initial distributions lead to different solutions. A sharp step decays in a sigmoidal curve (Fig. 28.25, middle column), whereas sinusoidal oscillations decay



**FIGURE 28.25.** Solutions to the basic diffusion equation,  $d\psi/dt = \sigma^2 d^2\psi/dx^2$ . (*Left column*) If the population starts concentrated at a point, it spreads out in a Gaussian distribution with variance  $\sigma^2 t$ . (*Middle column*) If the population starts out living only on the left side ( $x > 0$ ), the step smooths out over time. (*Right column*) Fluctuations in density ( $\sim \sin(\omega t)$ ) decay in amplitude, as  $\exp(-\sigma^2 \omega^2 t/2)$ . The rate of diffusion is set to  $\sigma^2 = 1$ .

## Box 28.6 The Diffusion Approximation

Suppose that at time  $t$ , a variable  $x$  has distribution  $\psi_t(x)$ ;  $x$  might represent the position of a randomly moving individual or the allele frequency in a randomly evolving population. In the next generation, a small random quantity  $y$  is added, giving a new and broader distribution,  $\psi_{t+1}(x)$ . The probability of finding  $x$  at time  $t + 1$  is the chance that the value was  $x - y$  at time  $t$ , and that the additional perturbation was  $y$ . Averaging over the distribution of perturbations,  $f(y)$ :

$$\psi_{t+1}(x) = \int \psi_t(x - y) f(y) dy.$$

Now, assume that the distribution  $\psi_t$  is smooth, and much broader than the distribution of perturbations,  $f(y)$ . Then,  $\psi_t(x - y)$  can be approximated as

$$\psi_t(x) - y \frac{d\psi_t}{dx} + \frac{y^2}{2} \frac{d^2\psi_t}{dx^2} \dots$$

(This is known as a Taylor's series approximation [Fig. 28.26].) It can be shown that as long as  $f(y)$  is not too broad, only the terms shown here are needed, involving the gradient ( $d\psi_t/dx$ ) and curvature ( $d^2\psi_t/dx^2$ ), just as, when the Central Limit Theorem holds, only the mean and variance are needed to describe the distribution. Substituting into the integral above,

$$\begin{aligned} \psi_{t+1}(x) &= \int \left( \psi_t(x) - y \frac{d\psi_t}{dx} + \frac{y^2}{2} \frac{d^2\psi_t}{dx^2} \right) f(y) dy \\ &= \psi_t(x) - \frac{d\psi_t}{dx} \int y f(y) dy + \frac{d^2\psi_t}{dx^2} \int \frac{y^2}{2} f(y) dy. \end{aligned}$$

Finally assume that changes from one generation to the next are small, so that  $\psi_{t+1} - \psi_t$  can be approximated by the continuous rate of change  $d\psi/dt$ . If the mean perturbation is written as

$$\int y f(y) dy = M,$$

and the mean square perturbation (which is approximately equal to the variance for small  $M$ ) as

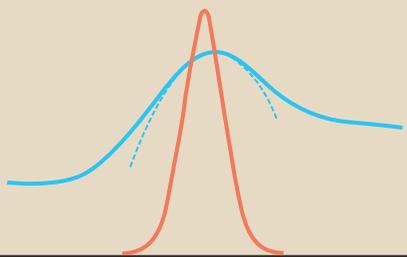
$$\int y^2 f(y) dy = V,$$

then

$$\frac{d\psi}{dt} = -M \frac{d\psi}{dx} + \frac{V}{2} \frac{d^2\psi}{dx^2}.$$

If the distribution is initially concentrated as a sharp spike at  $x = 0$ , then the solution to this diffusion equation is a Gaussian with mean  $M_t$  and variance  $V_t$  (Fig. 28.25, left column).

In this derivation, it is assumed that the distribution  $f(y)$  and hence  $M$  and  $V$  are constant. In general, they will depend on  $x$  and can be written as  $M(x)$ ,  $V(x)$ . Then, the formula given in the main text applies.



**FIGURE 28.26.** The diffusion approximation is derived using a Taylor's series (Box 28.6). The red curve shows the distribution of fluctuations,  $f(y)$ , and the blue curve shows the probability distribution  $\psi(x)$ . If the distribution of fluctuations is narrow, then the distribution can be approximated by a quadratic curve

$$\psi_t(x) - y \frac{d\psi_t}{dx} + \frac{y^2}{2} \frac{d^2\psi_t}{dx^2} \dots$$

shown by the dotted line.

at a steady rate, proportional to the square of their spatial frequency (Fig. 28.25, right column). The effect of diffusion is to smooth out distributions over time.

When the mean and variance ( $M$ ,  $V$ ) of each small fluctuation are fixed, the distribution follows a Gaussian (Box 28.6). Thus, the simple case is really a different way of presenting the Central Limit in terms of differential equations. In general, however, the mean and variance of changes in each generation will depend on the state of the system, in which case the distribution will almost always not be Gaussian. For example, changes in allele frequency must decline to 0 as allele frequencies tend to 0 or 1, because they must remain within these bounds. The successive changes are then no longer independent of each other, and the Central Limit Theorem no longer applies. However, the change in the distribution can still be described by a diffusion equation, which is a straightforward extension of that derived in Box 28.6:

$$\frac{d\psi}{dt} = \frac{d}{dx} \left( -M(x)\psi(x) + \frac{1}{2} \frac{d}{dx} (V(x)\psi(x)) \right).$$

### Box 28.7 Selection and Migration in a Small Population

The diffusion equation is a powerful tool for finding how different evolutionary processes combine. For example, imagine that an allele  $P$  is favored by selection  $s$  in a small population of  $N$  diploid individuals. However, polymorphism is maintained by immigration at rate  $m$  from a larger population with allele frequencies  $q^*$  and  $p^*$ . In this model, the mean change of allele frequency per generation is  $M(p) = spq + m(p^* - p)$  and the variance of fluctuations in allele frequency caused by random drift is  $V(p) = pq/2N$  (Box 15.1). Substituting these into the diffusion equation in the main text, the rate of change of the distribution of allele frequency is

$$\frac{d\psi}{dt} = \frac{d}{dp} \left( -(spq + m(p^* - p))\psi(p) + \frac{1}{2} \frac{d}{dp} \left( \frac{pq}{2N} \psi(p) \right) \right).$$

The probability distribution will reach an equilibrium where  $d\psi/dt = 0$ . (If there are many subpopulations, then this will be the distribution of allele frequency across the entire set.) This equilibrium can be found by setting

$$0 = -(spq + m(p^* - p))\psi(p) + \frac{1}{2} \frac{d}{dp} \left( \frac{pq}{2N} \psi(p) \right).$$

This can be rearranged to give a formula in terms of  $pq\psi$ :

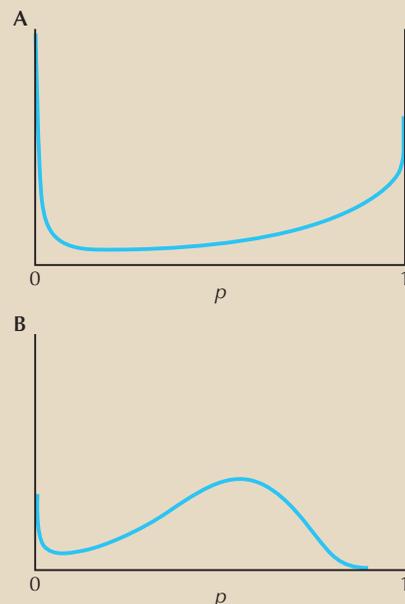
$$\frac{1}{4N} \frac{d}{dp} (pq\psi(p)) = \left( s + m \frac{p^* - p}{pq} \right) (pq\psi(p)),$$

which can be integrated to give a solution

$$\psi = C p^{4Nm p^* - 1} q^{4Nm q^* - 1} \exp(4Nsp),$$

where the constant  $C$  is chosen so that the probability sums to 1. (This solution can be verified by substituting it back into the diffusion equation.)

The most important feature of this distribution is that it depends only on the scaled parameters  $Nm$  and  $Ns$ , which give the strength of migration and selection relative to drift. When  $Ns$  is large, selection can maintain the favored allele  $P$  at high frequency, so that the distribution is peaked at the right. When  $Nm$  is small, the population tends to be fixed for one or the other allele (Fig. 28.27A), whereas when it is large, the distribution is pulled toward the frequency in the migrant pool ( $p \sim p^*$ ) (Fig. 28.27B).



**FIGURE 28.27.** The distribution of allele frequencies under migration, selection, and drift;  $p^* = 0.05$ . (A)  $Nm = 0.25$ ,  $Ns = 1$ . (B)  $Nm = 2$ ,  $Ns = 4$ .

This differential equation is extremely important in population genetics, because it gives a simple way to analyze the effects of different evolutionary processes on the distribution of allele frequencies (e.g., see Box 28.7). The **neutral theory** of molecular evolution (p. 59) was largely based on this method for analyzing the effects of random fluctuations; additional examples are on pages 495–496 and 641.

## SUMMARY

Mathematical theory is widely used in evolutionary biology, both to understand how populations evolve forward in time and to make inferences from samples of genes traced back through time. Models of deterministic processes, such as selection, mutation, recombination, and migration, follow the proportions of different genotypes through time. In sexually reproducing populations, this is difficult with more than a few genes, because of the huge number of possible gene combinations. Reproducing populations tend to grow exponentially; a variety of models, evolving either in discrete generations or continuously, can be approximated by a simple differential equation. Differential equations involving a single allele frequency can be solved, but scaling arguments can be used in more complex cases to show that the outcome depends only on parameter combinations such as the ratio of migration to selection,  $m/s$ . Models can be understood by identifying

their equilibria and by finding whether these equilibria are stable. Near to equilibria, the model can be approximated by simple linear equations, and its behavior depends on the magnitude of the leading eigenvalue.

The concept of probability is crucial both for following the proportions of different genotypes in deterministic models and for understanding the random evolutionary process of genetic drift. The probability distribution describes the chance of any possible outcome and may be defined for discrete possibilities (e.g., the numbers of individuals) or for continuous variables (spatial location or allele frequency in a large population, say). We describe two important models for how probability distributions change through time: branching processes, which apply to populations of genes or individuals that reproduce independently of each other, and the diffusion approximation, which applies when random fluctuations are small.

## FURTHER READING

Hacking I. 1975. *The emergence of probability*. Cambridge University Press, Cambridge.

A nice history of ideas in probability and statistics; useful to thoroughly understand the basic concepts of probability.

Haldane J.B.S. 1932. *The causes of evolution*. Longmans, New York.

This classic account includes an appendix that gives a concise sum-

mary of the basic models of population genetics.

Otto S.P. and Day T. 2007. *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton University Press, Princeton, New Jersey.

An excellent introduction to both deterministic and stochastic models in ecology and evolution.

