CHAPTER TWO

Alternative Views of the Scientific Method and of Modeling

Science is a process for learning about nature in which competing ideas about how the world works are measured against observations (Feynman 1965, 1985). Because our descriptions of the world are almost always incomplete and our measurements involve uncertainty and inaccuracy, we require methods for assessing the concordance of the competing ideas and the observations. These methods generally constitute the field of statistics (Stigler 1986). Our purpose in writing this book is to provide ecologists with additional tools to make this process more efficient. Most of the material provided in subsequent chapters deals with formal tools for evaluating the confrontation between ideas and data, but before we delve into the methods we step back and consider the scientific process itself. No scientist can be truly "neutral." We all operate within a fundamental philosophical worldview, and the types of statistical tools we employ and the types of experiments we do depend on that philosophy. Here we present four such philosophies.

There is a commonly accepted model for the scientific process (and from it arose a well-developed body of statistics that is taught in nearly every university in North America). The basic view can be thought of as a learning tree of critical experiments, which was described by Platt (1964) as "strong inference," and consists of the following steps:

1. Devising alternative hypotheses
2. Devising a crucial experiment (or several of them) with

   alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses
3. Carrying out the experiment so as to get a clean result
4. Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain, and so on (Platt 1964, 347)

Platt likens this to climbing a tree, where each fork of the tree corresponds to an experimental outcome, and we base the direction of the climb on the outcomes so far. It is especially interesting for us as ecologists that Platt associates a "second great intellectual revolution" with the "method of multiple hypotheses," and attributes some of the most original thinking in this area to the geologist T. C. Chamberlain who published at the end of the last century. In particular, Chamberlain stressed that we are guaranteed to get into trouble when we consider only a single hypothesis rather than multiple hypotheses. This is especially interesting because the similarities between the geological and ecological sciences are in some ways much greater than the similarities between the other physical and the ecological sciences. In both ecology and geology, experiments may be difficult to perform and so we must rely on observation, inference, good thinking, and models to guide our understanding of the world. In fact, ecology may be much more of an "earth science" than a "biological science" (Roughgarden et al. 1994). We include a reprint of Chamberlain's classic paper—first published in the 1890s—as the Appendix.

ALTERNATIVE VIEWS OF THE SCIENTIFIC METHOD

Platt's view is to a very great extent the logical extension of the work of Karl Popper (1979), who revolutionized the philosophy of science in the twentieth century by arguing that hypotheses cannot be proved, but only disproved

TABLE 2.1. Four philosophies of science.

Philosopher	Key word or phrase	Type of confrontation
Popper	Falsification of hypotheses	Single hypothesis is disproved by confrontation with the data.
Kuhn	Paradigms, normal science, scientific revolution	Single hypothesis used until there is so much contradictory information that it is "overthrown" by a "better" hypothesis.
Polanyi	Republic of science	Multiple views of the world allowed according to the different opinions of scientists. Confrontation between these views and the data judged on (i) plausibility, (ii) value, (iii) interest.
Lakatos	Scientific research program	Confrontation of multiple hypotheses with data as arbitrator.

(Table 2.1). The essence of Popper's method is to challenge a hypothesis repeatedly with critical experiments. If the hypothesis stands up to repeated experiments, it is not validated, but rather acquires a degree of respect, so that in practice it is treated as if it were true. Most "modern" scientific journals adopt this approach, even though there are difficulties in using it even under the best circumstances (e.g., Lindh 1993).

Coinciding with Popper's philosophical development was the statistical work of Ronald Fisher, Karl Pearson, Jerzy Neyman, and others, who developed much of the modern statis-

tical theory associated with "hypothesis testing" (e.g., Kendall and Stuart 1979, 175 ff.). In hypothesis testing, we focus on a single hypothesis (called the "null hypothesis") and calculate the probability that the data would have been observed if the null hypothesis were true. If this probability is small enough (usually 0.01 or 0.05), then we "reject" the null hypothesis. To complete the calculation, we must also compute the statistical power associated with the test (Peterman 1990a,b; Greenwood 1993; Thompson and Neill 1993). The power is the probability that if the null hypothesis were actually false and we were given the same data, we would reject it.

For example, we might begin with the idea that larger flocks of birds forage more effectively than smaller flocks. The null hypothesis could be that there is no relationship between flock size and foraging efficiency. A typical application of hypothesis testing would be to use linear regression to test the null hypothesis by calculating the probability that the slope of a graph of flock size versus feeding efficiency is non-zero. If the probability that the data could have arisen from the null hypothesis (slope = 0) is greater than 0.05 (or 0.01), the null hypothesis is not rejected at the "5% level" (or the 1% level). In the case considered here, if the null hypothesis could not be rejected at the 5% or 1% level and the power were sufficiently high, then the real ecological hypothesis—larger flocks forage more efficiently—would effectively be rejected.

After testing the hypothesis that larger flocks forage more efficiently, we would continue to climb Platt's decision tree to another set of experiments, depending on whether the effect of flock size on foraging efficiency was or was not statistically significant. The key elements of this view of science are (1) the confrontation between a single hypothesis and the data, (2) the central idea of the critical experiment, and (3) falsification as the only "truth." Popper supplied the philosophy and Fisher, Pearson, and colleagues supplied the

statistics. At best, this view of science is exceptionally narrow and actually does not fit many ecological situations. At worst, it can be downright dangerous, if, for instance, we accept the null hypothesis as true and the experiment had low power (also see Bernays and Wege 1987). Before we explain our own perspective, we want to provide an overview of some other views of science.

Thomas Kuhn (1962) introduced the ideas of "normal science," "scientific paradigms," and "scientific revolutions." According to Kuhn, scientists normally operate within specific paradigms, which are broad descriptions of the way nature works. Normal science involves collection of data within the context of the existing paradigm. Normal science does not confront the existing paradigm, rather, it embellishes it. The paradigm dictates what type of experiments to perform, what data to collect, and how to interpret the data. In Kuhn's view, real change occurs only when (i) a large body of contradictory data accumulates and the existing paradigm cannot explain the data, and (ii) there is an alternative paradigm that can explain the discrepancies between the old paradigm and the observations. Kuhn argues that there is rarely, if ever, a critical experiment at the level of the paradigm. Instead, a particular anomaly will be explained as a measurement problem. It is the collection of contradictory experiments that leads to the revolution.

The Kuhnian perspective is that the type of experimental trees and critical experiments described by Platt may occur, but only within an individual paradigm, and that they are the standard procedures of normal science. The example we gave earlier of examining the relationship between flock size and foraging efficiency would be considered normal science within a broad paradigm of natural selection acting on behavior.

Michael Polanyi (1969) describes a "republic of science" consisting of a community of independent thinkers cooperating in a relatively free spirit. To Polanyi, this represents a

simplified version of a free society in which scientists develop by "training" with a "master" so that the practice of science is analogous to apprenticing with a master artisan and learning the skills of the artisan by close observation and participation. Scientists are chosen through this apprenticing system; the individuals constitute the "republic" of citizens taught through the master-apprentice chain. It is this system that prevents science from becoming moribund or rigid, since the apprentice both learns high standards from the master and develops his or her own judgment for scientific matters. There are three main criteria for judgment (Polanyi 1969, 53 ff.): (1) plausibility, (2) scientific value (consisting of accuracy, intrinsic interest, and importance), and (3) originality. The criteria of plausibility and scientific value will encourage conformity, whereas the value given to originality encourages creative thinking and dissent. This forms the essential tension in any scientific field, and the three criteria considered by Polanyi are appropriate ones that we can use for confronting models with data. Polanyi implicitly argues that the intellectual confrontation is not between a model and data, but between models (i.e., different descriptions of how the world works) and data (observations and measurements).

There is an overlap between the ideas of Polanyi and Kuhn. The apprentice system is the essence of Kuhn's normal science: apprentices learn from their masters what type of experiments to perform, and then, to a large extent, continue to work on this type of problem for the rest of their careers. It is the unusual scientist who breaks away from the material of the apprenticeship and enters a new field. We have noticed how common it is in ecology for someone to do a Ph.D. in a specific area, often with a particular taxonomic group, and then continue for most of a career to study the same topic. One of our colleagues in a chemistry department said that it was the same in his field: more than 70% of his colleagues worked with the same types of reac-

tions they studied for their Ph.D.'s. This is the apprentice system and normal science. It is unlikely to lead to innovation or breakthroughs.

Imre Lakatos (1978) describes "scientific research programs" (SRPs) that consist of a set of methodological rules that guide research by indicating paths to avoid and paths to pursue. The "hard core" is the key element of the SRP, which generates a set of surrounding hypotheses that make specific predictions. Lakatos refers to these surrounding hypotheses as a "belt" that protects the hard core. The individual elements of the belt can be tested, and rejected, but one can rarely, if ever, directly challenge the hard core.

Lakatos points out that many hypotheses (e.g., Newton's laws and the theory of gravity) have been highly regarded and used despite their acknowledged inconsistency with some aspect of the data. Organic chemists worked for years with models that they knew were wrong but for which alternatives were lacking. Lakatos argues that the value of an SRP is its ability to make new predictions and provide a simple and elegant explanation of what is known. An SRP can only be replaced by another SRP: One cannot reject a hypothesis unless there is something better on hand to replace it. Mitchell and Valone (1990) argued that optimization in biology should be viewed as an SRP (also see Orzack and Sober 1994).

Thus, in the Lakatosian view, the contest must always be between competing hypotheses and the data. An individual hypothesis may well be inconsistent with the data, but unless there is another hypothesis that is more consistent with the data, you will not discard the first hypothesis because you have to keep working. The recognition of the importance of more than one model is slowly appearing. For example, Chen et al. (1992) compare a number of functions used to describe the growth of fish. If we only consider one growth function, we shall surely use it to make predictions, regardless of its efficacy, but comparing different growth functions

allows choice in the description of how nature works. Similarly, Schnute and Groot (1992) confront ten different models of animal orientation with data, Ribbens et al. (1994) compare different models for seedling recruitment in forests, and Kramer (1994) compares six different models for the onset of growth in the European beech.

To a great extent, Popper's view of falsification, Kuhn's normal science, Polanyi's republic, and Lakatos's testing of the "belt" of auxiliary hypotheses are different descriptions of the same scientific activity. It is rare that the major ideas, such as evolution by natural selection or the theory of relativity, are truly tested. In fact, most of the work of the ecological detective will be at a considerably more mundane level. Indeed, it is safe to say that we are writing this book as a handbook for the practice of normal science. (Although, of course, we hope that something more exciting comes from it.)

As briefly described in the previous chapter, the field of likelihood/Bayesian statistics is well suited for the analysis of the contest between competing hypotheses and data. The essence of likelihood/Bayesian analysis is the calculation of the chance of the data given a particular hypothesis, and (for Bayesian methods) from that, "posterior distributions" that describe the probability assigned to each possible hypothesis after data are collected. We describe the mechanics of Bayesian statistics in succeeding chapters. Here we briefly contrast the approaches of classical and likelihood/Bayesian statistics. We shall show in succeeding chapters that likelihood methods are a special case of Bayesian ones, so that from now on we simply refer to them as Bayesian methods.

In classical statistics, we test each hypothesis against the data in a mock confrontation with a "null hypothesis." In Bayesian statistics, we test the hypotheses together against each other, using the data to evaluate the degree of belief that should be accorded each of the hypotheses. The result of a classical analysis is rejection or nonrejection of the lone

hypothesis, whereas the result of a Bayesian analysis is "degrees of belief" associated with the different hypotheses.

Two of the three pillars of the classical viewpoint, falsification and the confrontation between a single hypothesis and data, are directly opposed by the Bayesian viewpoint. In the classical approach hypotheses are falsified (but never proved), but in the Bayesian viewpoint degrees of belief are increasing or decreasing. "Falsification" exists only as low degrees of belief and "proof" is strong belief. The two views also are diametrically opposed on whether the confrontation is between a hypothesis and the data, or between competing hypotheses and data. According to Lakatos, we cannot reject a hypothesis unless something better awaits, and Bayesian computation requires more than one hypothesis. In the viewpoint of Popper and classical statistics, we can reject a hypothesis by itself in single combat with data. But then what?

There is much more compatibility between the differing viewpoints on the question of critical experiments. To a Bayesian, a critical experiment is one that will greatly change the degrees of belief in competing hypotheses. Indeed, there is no point in conducting an experiment that will not change the degrees of belief. To a Bayesian, the ideal Popperian critical experiment is one that will change the degrees of belief to almost 1.0 for one hypothesis and almost 0.0 for the others, depending upon the outcome of the experiment. The best experiments are those that discriminate most clearly, although the Popperian/classical view would not require that there be competing hypotheses. We find the Lakatosian/Bayesian view more compelling: that the contest is between competing hypotheses and data, not between a single hypothesis and the data.

We must also consider the issue of statistical significance versus biological significance. Too many people operate on the premise that if statistical significance cannot be shown, the work cannot be published. Yet even elementary statistics courses teach us that statistical significance often has little,

if any, relation to biological significance. Two curves can be statistically significantly different even if they differ by less than one percent, given a large enough sample size or small enough measurement error. Conversely, given small sample sizes or high variability, even the most different of biological relationships can fail to be statistically significant. And yet, especially when experiments are difficult or management actions needed, we may not have the luxury of obtaining statistical significance before needing to act on our hypotheses.

STATISTICAL INFERENCE IN EXPERIMENTAL TREES

Now let us return to Platt's experimental tree and consider it from the different perspectives. The basic structure of an experimental tree is compatible with the varying viewpoints if they are suitably modified. Lakatos would insist that each experiment be a contest between competing hypotheses, whereas Popper would accept experiments testing a hypothesis with no competitor. More importantly, Lakatos would not accept that the "hard core" of an SRP could be experimentally tested in this way. Popper would see the experiments as testing the key hypothesis, since a good hypothesis is one that is amenable to direct experimental falsification.

Platt's experimental tree is based on the premises of (i) very clear and distinct hypotheses and (ii) nonambiguous outcomes. Examining the nature of the statistical tests that could be used in working through an experimental tree shows the problems of the method of hypothesis testing. Imagine you are at experiment A and are asking if larger flocks forage more efficiently. Suppose that if the null hypothesis cannot be rejected, experiment B is appropriate, whereas if the null hypothesis is rejected (therefore large flocks do forage more efficiently), experiment C will be next. What significance level should one choose to decide which branch of the tree to follow? Should experiment C be

next, even if the estimated increase in foraging efficiency for larger flocks is biologically trivial, although statistically significant?

In our view, an experimenter would more profitably operate as follows. At the conclusion of experiment A there are really seven options, not two:

1. go on to experiment B,
2. go on to experiment C,
3. repeat experiment A,
4. perform both B and C,
5. perform both A and C,
6. perform both A and B, or
7. perform A, B, and C!

Indeed, if the experiments are inexpensive to set up and run but require considerable waiting time for the outcome, it would be best to do A, B, and C simultaneously.

Progress through an experimental tree thus depends on several factors including (1) the cost of each experiment, (2) the time required to do each experiment, and (3) the relative degree of belief in competing hypotheses. At any stage in the tree, a good scientist will compare the cost and time required to do each experiment to the degree of belief in competing hypotheses and from these calculate the optimal next experiment(s).

UNIQUE ASPECTS OF ECOLOGICAL DATA

Platt envisioned very clean experiments in which one hypothesis would be clearly discredited. Indeed, a key thrust of Platt's argument is that the fields that made the most rapid progress were those fields that routinely thought about and designed such experiments. Clearly, a field will make more rapid progress if such clear, critical experiments can be designed and conducted, and ecologists should seek to work on systems that are amenable to such analysis. Whenever

possible, conduct an experiment (Hairston, 1989, 1994; Underwood, 1991). However, many ecological studies are motivated by problems where such clear experimentation and "hard data" are often not possible (Fagerström 1987) or lead to other difficulties, as the recent "Frontiers in Biology" in *Science* (269:313–61, 1995) and associated correspondence (269:561–64, 1201–3) demonstrate.

For example, consider the problems in understanding the dynamics of populations of blue whales. There is no possibility for experimental manipulation (for decades at least), there is no possibility for replication, since there are so few individuals and they may constitute a single population, and the time scale of their dynamics is very slow. We cannot design a Platt-type experimental tree for manipulation of blue whales—but we could design such an experimental tree for many hypotheses and use observation, rather than experiment, to differentiate between the hypotheses.

Blue whales are an extreme example, but the following attributes of ecological systems often make experimentation difficult:

- Long time scales: Many ecological systems have time scales of years or decades
- Poor replication: Many ecological systems are difficult to replicate, and replicates are rarely, if ever, perfect
- Inability to control: One can rarely, if ever, control all aspects of an ecological experiment

Because of these factors it is often harder to get clear, unambiguous results in ecological experiments (cf. Shrader-Frechette and McCoy 1992). Platt described an experimental approach that did not really need statistics, because each experiment produced a clear result. This is not often the case in ecological work.

Of course, new students should seek systems that do not have these problems, and we encourage you (especially graduate students) to find systems that operate on short time

scales and can be easily replicated and easily controlled. It sometimes happens that we are able to apply knowledge from small-scale experimental systems to larger-scale "real world" systems, but it is likely that at least some of the work of the ecological detective will be on ecological systems that may present all three of these difficulties.

DISTINGUISHING BETWEEN MODELS AND HYPOTHESES

We begin by trying to sort out "theory," "hypothesis," and "model." The etymology of theory is Greek, *theoria*, meaning "a looking at, contemplation, speculation," and we understand theory to mean "a systematic statement of principles involved" or "a formulation of apparent relationships or underlying principles of certain observed phenomena which has been verified to some degree." The theory of evolution by natural selection, without doubt the most important theory in modern biology, is still mainly nonmathematical. The same is true of the theory of Crick and Watson that DNA is a double helix (Crick 1988). The etymology of hypothesis is also Greek, *hypotithenai*, meaning "to place under."

A hypothesis is "an unproved theory, proposition, supposition, etc., tentatively accepted to explain certain facts or to provide a basis for further investigation." Webster's dictionary (Neufeldt and Guralnik 1991) separates theory and hypothesis as follows: "**theory**, as compared here, implies considerable evidence in support of a formulated general principle, explaining the operation of certain phenomena; **hypothesis** implies an inadequacy of support of an explanation that is tentatively inferred, often as a basis for further experimentation."

The etymology of model is from Latin *modus*, meaning the way in which things are done. A model is an archetype, "a stylized representation or a generalized description used in analyzing or explaining something." Thus, models are tools for the evaluation of hypotheses (our best understanding of how the world works), but they are not hypotheses (cf. Caswell 1988; Hall 1988; Onstad 1988; Ulanowicz 1988).

Most hypotheses could be represented by a number of models. The hypothesis that birds forage more efficiently in flocks than individually could be represented by several models relating consumption rate C and flock size S:

$$
\begin{aligned}
C &= aS && \text{Model A: Consumption proportional to flock size,} \\
C &= \frac{AS}{1 + bS} && \text{Model B: Consumption saturates as flock size increases,} \\
C &= aSe^{-bS} && \text{Model C: Consumption increases and then decreases with increasing flock size,}
\end{aligned} \tag{2.1}
$$

where a and b are parameters of the models. Each model is a more explicit statement of the hypothesis that "birds forage more efficiently in larger flocks" (Figure 2.1). The "null hypothesis" is the model that forage efficiency is independent of flock size, or $C = a$. In the Popperian confrontation models A, B, and C would individually be "tested" against the null hypothesis. In a Lakatosian world the confrontation would be between the four competing models (A, B, C, and the "null").

One can think of hypotheses and models in a hierarchic fashion with models simply being a more specific version of a hypothesis. Furthermore, particular parameter values of the models are even more specific hypotheses. Indeed, in later chapters that deal with probability, likelihood, and Bayes' theorem, we use the word "hypothesis" to refer to particular parameter values of specific mathematical models. The use of "hypothesis" with reference to probabilities is unfortunate, though necessitated by the general statistical usage, but do not confuse the distinction between a hypothesis as a general statement about the natural world and the

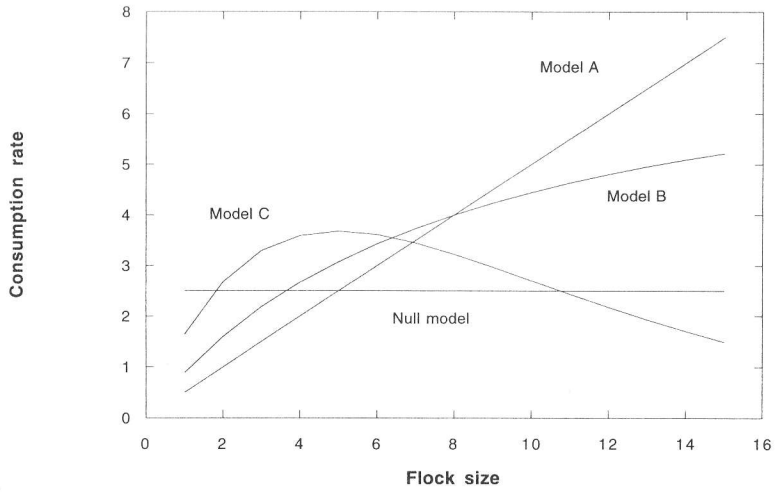

FIGURE 2.1. Three models of how foraging efficiency might be affected by flock size. The flat line is the null hypothesis that flock size does not affect foraging efficiency.

variety of mathematical models that can be used to represent the hypothesis.

We use models to evaluate hypotheses in terms of their ability both to explain existing data and predict other aspects of nature. We use models to combine what we know with our best guesses about what we do not know. The equations of a model represent a very specific expression of the hypothesis. For example, a hypothesis might be that "predation has a significant effect on the average abundance of the population of *X*." Models of this hypothesis would describe the interaction between the organism and its predators in the context of specific mathematical forms (one of which—the null model—could include no predation). Were such models confronted with abundance data, we might find that models including predation explained the abundance of *X* no better than a model without predation. We would then have some evidence that the hypothesis is incorrect. In this "Lakatosian" view of hypotheses and models, the individual

models are the surrounding belt that defends the core hypothesis. We chip away at the individual model and eventually, as we exhaust the possibilities of different mathematical representations of predation, decrease belief in the underlying hypothesis of the importance of predation and increase belief in the alternative hypotheses. Wise (1993) provides an example of how this program is followed in understanding the roles of spiders in ecological systems.

Models have a number of different purposes in the general evaluation of scientific hypotheses. First, models help us clarify verbal descriptions of nature and of mechanisms. Formulation of a model often forces the researcher to think about processes that he or she had previously ignored. The formulation leads to identification of parameters that must be measured and often helps crystallize thinking about the processes involved.

Second, models often help us understand which are the important parameters and processes and which ones are not important. For example, in the formulation of a model we often see that combinations of parameters, rather than the individual parameters themselves, determine the behavior of the system (see Mangel and Clark, 1988, epilogue). Models thus allow us to rank the importance of different factors about the phenomenon in a quantitative manner.

Third, since a model is not a hypothesis we must admit from the outset that there is no "fully correct" model. Instead, there are sequences of models, some of which may be better than others as tools for understanding the natural world. Different models of the same phenomenon can be quite useful, as we shall see in several of the case studies presented later. Different models allow us to assess the validity of different assumptions and, in some cases, of fully different hypotheses. The development of different models usually represents a progression in the understanding of the natural system. This is especially important; one must focus on the system of interest and be willing to forego the model

BOX 2.1

SEPARATING HYPOTHESES AND MODELS: A SCENARIO FROM CLASSICAL PHYSICS

Here we expand upon an example used by Mangel and Clark (1988). This example requires elementary physics. Envision a mass M attached to a spring which is then attached to a ceiling. We pull the ball away from the ceiling and let it go; the ball starts to oscillate. Our goal is to understand what is happening. We begin with the usual hypothesis of Newton's second law of motion: $F = Ma$ (force equals mass times acceleration). If $X(t)$ denotes the displacement of the mass from the original resting spot at time t, then the simplest model for the restoring force is that it is proportional to the displacement

$$\text{Model 1:} \quad M = \frac{d^2X}{dt^2} = -KX. \tag{B2.1}$$

Here d^2X/dt^2 is the acceleration. The solution of this differential equation (which you may have once studied in physics or calculus) leads to two important predictions. First, this simple model predicts that the spring will oscillate forever. Second, the frequency of oscillations depends on the combination $\sqrt{K/M}$ and not on K or M independently; this is something that we could not have determined without the model.

However, there are problems. Real springs ultimately slow down and stop oscillating. Do we conclude that the hypothesis $F = Ma$ is wrong or that the model is missing something? For example, we have ignored frictional forces which tend to slow things down according to the size of their velocity. Hence, we might modify model 1 to obtain

$$\text{Model 2:} \quad M\frac{d^2X}{dt^2} = -KX - K_1V, \tag{B2.2}$$

where $V = dX/dt$ is the velocity and we have added another parameter K_1 that relates the frictional force and velocity.

BOX 2.1 CONT.

Once again, by solving the equation we could learn that the answer does not depend on K_1 itself but on the ratio K_1/M, and that model 2 predicts that the spring will slow down. Consequently, this is a clear improvement in the model without any change of hypothesis.

However, real springs slow down and stop in a finite time, but the spring described by model 2 will only stop as time becomes infinite. Once more, we conclude that there is a problem with the model and might introduce

$$\text{Model 3:} \quad M\frac{d^2X}{dt^2} = -KX - K_1V - K_2V^3, \tag{B2.3}$$

where we have added yet another parameter, which now relates the friction force to the cube of the velocity. The solution of Equation B2.3 requires advanced methods and is usually not treated in introductory courses. Note, however, that these three models are "nested": we obtain model 2 or model 1 from model 3 by setting certain parameters equal to 0.

Thus, with the single hypothesis $F = Ma$, we have at least three different models and could confront these models with the observations. Surely we believe that none of these is "correct," but that they are increasingly better descriptions of reality within the hypothesis.

Now suppose that the mass is a ball containing sand and that there is a hole in the bottom so that the sand falls out as the oscillations occur. In this case, our hypothesis is no longer correct, because $F = Ma$ assumes that the mass is constant. In more advanced physics courses, one learns that the appropriate hypothesis for the case in which the mass is a changing function of time, $M(t)$, is $F = (d/dt)$ (momentum), where momentum $= M(t)V$. This is an alternate hypothesis, which requires another series of models like the ones we just discussed.

when a better one arises (that is, don't fall in love with your model). Complicated models with more parameters and mechanisms will usually give better fits to data than simpler models, but if our models are as complicated as nature itself, then we may as well not bother with the model and focus only on the natural situation. Simpler models often provide insight that is more valuable and influential in guiding thought than accurate numerical fits. In fact, although the output of most models is numerical, the most influential models are the ones in which the numerical output is not needed to guide the qualitative understanding.

In summary, models allow us to tie together different bodies of data and aid in the identification of salient, necessary, and sufficient features of a system. The use of models while planning an experiment may help identify variables that will be confounded in the analysis of the results. Finally, models allow us to explore the parameter space and analyze multidimensional systems in ways that are virtually impossible from a purely empirical perspective.

Recognition of the model as a scientific tool has a number of important implications. First, one must try to validate assumptions before starting, or at least keep track of the untested assumptions. For example, the generally rancorous discussion concerning optimality theory in biology over the last twenty years was caused, in no small part, because both sides failed to recognize the nature of the assumptions and failed to clearly identify what was being tested and what was not being tested (e.g., Stephens and Krebs 1986; Mitchell and Valone 1990; Orzack 1993; Orzack and Sober 1994). The typical scenario often went like this: A model of an "optimally foraging animal" was constructed and compared with data. The data and model never matched completely, so opponents claimed that "optimal foraging" was disproved, while proponents modified the model and tried again to obtain agreement between the model and the data. And the argument continues.

The idea that models should be used as a principal tool in confronting hypotheses with data as arbitrator leads into a natural discussion of "model validation." It is a long-held and common view that in ecological studies, models should be "validated" by some kind of comparison of predictions of the model and the data that motivated it (e.g., Naylor and Finger 1967; Mankin et al. 1975; Shaeffer 1980; Leggett and Williams 1981; Feldman et al. 1984; Santer and Wigley 1990; Wigley and Santer 1990). Adopting a Popperian view, if the model is inconsistent with any of the data, then it (and the associated hypothesis) should be rejected. The model would be tested repeatedly, subjecting it to new challenges in the form of new empirical data. A model that withstood repeated challenges could be considered as "valid" only in the sense that it was not rejected. In contrast, adopting the Lakatosian view, all models will be found inconsistent with some of the data, and the question is which models are most consistent and which ones meet the challenges of new experiments and new data better. Thus, models are not validated; alternative models are options with different degrees of belief (see Oreskes et al. 1994 for an excellent discussion of this topic for models in the earth sciences). If one model clearly fits the existing data best and has proven ability to explain new data, we might have a very high degree of belief. It is not validated—but is better than the competitors. The favorite model of the current moment will likely be replaced by another model in the future. Levins (1966, 430–31) wonderfully states the situation:

> A mathematical model is neither an hypothesis nor a theory. Unlike scientific hypotheses, a model is not verifiable directly by an experiment. For all models are both true and false. . . . The validation of a model is not that it is "true" but that it generates good testable hypotheses relevant to important problems. A model may be discarded in favor of a more powerful one, but it usually is simply out-

> grown when the live issues are not any longer those for which it was designed. . . . The multiplicity of models is imposed by the contradictory demands of a complex, heterogeneous nature and a mind that can only cope with a few variables at a time . . . individual models, while they are essential for understanding reality, should not be confused with that reality itself.

TYPES AND USES OF MODELS

The ecological literature is filled with different kinds of models, which can be used for different kinds of investigations (Loehle 1983). One way to classify models is according to dichotomies. Here we specify some of these differences, and in the applications chapters you will see the different kinds of models in action.

Deterministic and Stochastic Models

Deterministic models have no components that are inherently uncertain, i.e., no parameters in the model are characterized by probability distributions. In stochastic models, on the other hand, some of the parameters are uncertain and characterized by probability distributions. For fixed starting values, a deterministic model will always produce the same results, but the stochastic model will produce many different results depending on the actual values the random variables take.

Statistical and Scientific Models

A scientific model begins with a description of how nature might work, and proceeds from this description to a set of predictions relating the independent and dependent variables. A statistical model foregoes any attempt to explain why the variables interact the way they do, and simply attempts to describe the relationship, with the assumption that the relationship extends past the measured values. Regression models are the standard form of such descriptions, and Peters (1991) argued that the only predictive models in ecology should be statistical ones; we consider this an overly narrow viewpoint.

Static and Dynamic Models

Static models predict a response to input variables that does not change over time. Dynamic models involve responses that change over time. In this regard, dynamic models become more complicated because they often involve the link of the response between one period and the next.

Quantitative and Qualitative Models

Quantitative models lead to detailed, numerical predictions about responses, whereas qualitative models lead to general descriptions about the responses. The ideal use of models is to develop quantitative models from which qualitative insights can be gained. It is often reasonable to test quantitative predictions that are based on simple models, using estimated or averaged parameters, with the intention of assessing how well the simple description of nature works. Qualitative models, on the other hand, can be used more broadly to describe regions in which one response is expected and regions in which a different response is expected. For example, when studying whether an insect of a given age and physiological state will oviposit on a host of a specified type, we might use a model (Mangel 1987) to divide the "age/state" plane into one region in which oviposition will occur and one in which it will not occur. A quantitative model would attempt to determine the precise location of the boundary, whereas a qualitative model would recognize that such a boundary exists and then ask how the responses would change in response to other parameters. Such predictions are quite testable (Roitberg et al. 1992, 1993).

Models for Understanding, Prediction, and Decision

We must recognize that in addition to different kinds of models there are different uses of models. We may model a natural system to broadly test our understanding of the mechanisms in the system. However, models usually lead to numerical predictions. In that case, we want to abstract qualitative, intuitive understanding from the broad pattern of the numerical predictions.

A model may be used for purposes of prediction. Such predictions can be both qualitative (e.g., "the system will/will not respond to this effect") and quantitative (e.g., "the level of the response will be . . ."). A model is most effective, of course, if it provides both understanding (of known patterns) and prediction (about situations not yet encountered).

Finally, we can use the model as part of a decision-making process. In this case, the model provides a means for evaluating the potential effects of various kinds of decisions. It is in this realm that models have the most to offer in terms of practical application, but also where the greatest potential danger lies.

NESTED MODELS

Very often, we want to develop different models for the description of the same phenomenon. A particularly useful way of doing this is by adding complexity so that the "next model" contains the "previous model" as a special case, usually when some parameter (or parameters) is fixed. A family of models is called nested if the simpler models are special cases of the more complex models (see McCullagh and Nelder 1989 for a general discussion).

As a specific example, suppose we had a set of observations of population abundance Y at a series of spatial sites, indexed by i, and a number of independent variables measured at the same sites, such as water availability, ground

cover, tree cover, insect abundance, etc. We denote these variables with X_{i1}, X_{i2}, X_{i3}, etc. (where X_{ij} is the value of the j^{th} measured variable at site i). One model relating these variables is

$$\log(Y_i) = p_0 + p_1X_{i1} + p_2X_{i2} + p_3X_{i3} + E_i, \qquad (2.2)$$

where the E_i represents a source of uncertainty and the parameters p_i are determined during the confrontation with the data. A model such as Equation 2.2 is called a log-linear model, because the logarithm of the dependent variable Y_i is assumed to be a linear function of the independent variables $\{X_i\}$. The model Equation 2.2 is one of a family of models that includes

$$\log(Y_i) = p_0 + p_1X_{i1} + p_2X_{i2} + E_i,$$

$$\log(Y_i) = p_0 + p_1X_{i1} + E_i,$$

$$\log(Y_i) = p_0 + p_1X_{i1} + p_3X_{i3} + E_i,$$

$$\log(Y_i) = p_0 + p_2X_{i2} + p_3X_{i3} + E_i,$$

$$\log(Y_i) = p_0 + p_2X_{i2} + E_i,$$

$$\log(Y_i) = p_0 + p_3X_{i3} + E_i, \qquad (2.3)$$

as some of the special cases. All the models in Equation (2.3) are special cases of the full model when different parameters are set to zero; this family of models is said to be nested. The same is true for some of the models in Equation 2.1 (reader, which ones?).

Many ecological models can be treated as nested models. The Leslie life history model (Caswell 1989), used frequently for age- or size-structured populations, is

$$\begin{aligned} N_{a+1,t+1} &= s_aN_{a,t}, \qquad \text{for } a > 1, \\ N_{1,t} &= \sum_a m_aN_{a,b} \end{aligned} \qquad (2.4)$$

where $N_{a,t}$ is the number of animals of age a at time t, s_a is the fraction of animals of age a surviving to age $a + 1$, and m_a is the reproduction by animals of age a. A special, and therefore nested, case of the Leslie model is one without age structure, which can be obtained by assuming that the survival at each age is the same (so set $s_a = s$) and that reproduction at each age is the same (so set $m_a = m$). Then if N_t is the total population size and $B_t = mN_t$,

$$N_{t+1} = sN_t + B_t. \tag{2.5}$$

The alternative to nested models is to consider models that are structurally different, where we cannot change a parameter to obtain one model from the other. In dealing with non-nested models we can no longer simply ask if we obtain better fits to the data by making the model more complex, but we must see how well the alternative models fit the data.

MODEL COMPLEXITY

Perhaps the most difficult decision in model building is "How complex should the model be?" With microcomputers and modern software it is easy to build models quickly, to run the models, and generate lots of output. It takes only a few minutes to add additional variables to the model and if we continue for a few hours, we could have a model with dozens or hundreds of variables. What is the best-sized model? There are usually two major factors influencing the answer to this question. On one hand, we can always imagine that the model would be better ("more realistic") if we added another component to it—something we have observed in nature and hate to leave out. On the other hand, if we have a smaller model, the computer will run faster, fewer parameters will be needed, and the output will be easier to understand. Most neophytes are tempted to build very large models, and we urge you to resist this temptation. Of

course, the best-sized model depends on the purpose of the model. Given this objective, the basic rule about model size is

Let the data tell you.

There are quantitative methods for determining the optimal size of a particular model (Ludwig and Walters 1985; Linhart and Zucchini 1986; Walters 1986; Punt 1988; Gauch 1993). If the model is too simple, we risk leaving out significant components of the system. If the model is too complex, we will not have sufficient information in the data to distinguish between the possible parameter values of the model.

For example, many ecological analyses of population dynamics rely on the Leslie matrix with age-specific survival and fecundity. If we wish to make projections of the population size and have estimated survival and fecundity for only a few individuals, we have the choice of several models. The simplest model (e.g., Equation 2.5) would average the survival and fecundity over all ages; the most complex model (e.g., Equation 2.4) would estimate the survival and fecundity at each age from the data. If the species is long lived and the number of individuals for whom survival and fecundities has been measured is small, estimates of the age-specific survival and fecundity are likely to be poor, and it would be better either to use a single value for all ages or at least to average survival and fecundity over age groups. The number of ages aggregated should depend on the amount of data available and the number of age classes considered.

Linhart and Zucchini (1986) provide a formal framework for considering different levels of model complexity in the reliability of model predictions. Their approach distinguishes between prediction error due to approximation, which decreases as model complexity increases, and prediction error due to estimation, which increases as model complexity increases. For any model and amount of data, the total predic-

tion error will decrease and then increase as model complexity increases—with respect to reliability of prediction, there is an optimal level of model complexity.

Linhart and Zucchini's approach is consistent with almost all quantitative work in this area that suggested the optimal model size is much smaller than intuition dictates. Ludwig and Walters (1985) obtained better predictions about management actions from a non-age-structured model, even when the data were derived, by simulation, from an age-structured model. That is, the "wrong" model can do better than the "right" model in prediction if parameters must be estimated. Similarly, Punt (1988) found very simple models of fisheries management, which often ignored substantial amounts of data, outperformed more complex models when parameters had to be estimated and decisions made.

When the objective is something other than prediction accuracy, the complexity of the optimal mode may be quite different. In Chapter 10, we show a fisheries example where a complex model fits the available data no better than a simpler model. However, the uncertainty in the sustainable harvest is quite low for the simple model, but high for the complex model. In this case the simple model under-represents the uncertainty, and we believe that a more complex model provides a better representation of the uncertainty.

The complexity of the optimal model will depend on the use of the model and on the data. Part of the work of the ecological detective is to iterate between alternative models, to understand their strengths and weaknesses, and to recognize that the most appropriate model will change from application to application.

CHAPTER THREE

Probability and Probability Models: Know Your Data

DESCRIPTIONS OF RANDOMNESS

The data we encounter in ecological settings involve different kinds of randomness. Many ecological models describe only the average, or modal, value of a parameter, but when we compare models to data, we need methods for determining the probability of individual observations, given a specific model and a value for the mean or mode of the parameter. This requires that we describe the randomness in the data. Similarly, when we build a model and want to generate a distribution of some characteristic, we first need a way to quantify the probability distribution associated with this characteristic. This involves understanding both the nature of your data and the appropriate probabilistic descriptions.

We assume that readers of this book are familiar with the normal or Gaussian distribution (the familiar "bell-shaped curve"). However, many of the distributions in nature are not normal. The purpose of this chapter is to introduce ideas about probability, describe a wide range of useful probability distributions (and consider biological processes that give rise to these distributions), and provide you with the tools you need to use these distributions in your work. We begin with advice on data and then review the concepts of probability. After that, we describe a number of different probability distributions and some of their ecological applications. We close with a description and illustration of the "Monte Carlo" method for generating data and testing models.