CHAPTER

1

Introduction

OUTLINE

1.1 ECOLOGY: THE STUDY OF DISTRIBUTION AND ABUNDANCE AND OF THE MECHANISMS DRIVING THEIR CHANGE

Ecology is concerned with the number (abundance, N) of living things—how many individuals there are and how their number evolves over time, where they are and where they go to. Important questions concern their interactions with the abiotic and biotic environment, including each other, and what mechanisms drive these numbers and their dynamics. This classic view of ecology is reflected by the titles of two seminal textbooks: *The Distribution and Abundance of Animals* (Andrewartha and Birch,

Bayesian Population Analysis using WinBUGS
DOI: 10.1016/B978-0-12-387020-9.00001-8

1954) and *Ecology: The Experimental Analysis of Distribution and Abundance* (Krebs, 2001).

More generally, ecology can be described as the science that studies how states of biological systems interact with their environment and how this results in the temporal dynamics and spatial patterns of organisms that we observe. Figure 1.1 shows how state S evolves over time. The arrows connecting states between successive time periods denote the rate parameters that govern changes of state. State S may denote an individual state such as "alive" or the state of a collection of individuals, that is, population, such as "occurrence" or "local abundance, N". For the individual state "alive", the arrows may represent the coin-flip-like survival process. For the abundance state (N), the arrows may represent the demographic rates of survival, fecundity, immigration, and emigration. It is those rates on which the ecological mechanisms act to determine how a population is distributed in space or evolves over time.

A pervasive theme in ecology is that of hierarchical scales of organization—genes are nested within individuals, individuals within populations, populations within metapopulations or communities, and communities within metacommunities. Interestingly, this view of ecology is again reflected by the title of an influential ecology textbook: *Ecology: Individuals, Populations, and Communities* (Begon et al., 1986). These scales have biologically quite different meanings, and the practitioners of the associated branches of ecology often have very little in common with one another. And yet, it is fascinating to recognize that we can move among these scales simply by a redefinition of counted units (i.e., what we call an "individual") and that they can be characterized by what is essentially the same set of quantitative demographic descriptors (Table 1.1).

At Scale 1, the unit is the classical individual living in a population (Table 1.1). It can move between states such as "alive" and "dead" or "newly recruited" and "not newly recruited", thereby defining demographic rates such as survival and recruitment, respectively. Scale 1 represents the classic population concept. The interest is usually in understanding how biotic and abiotic factors impact vital rates (e.g., Newton, 1998) and

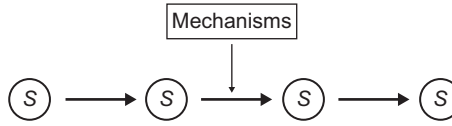

FIGURE 1.1 The classic view of ecology as the science dealing with how the state S (for instance, local population size N) of a living system evolves over time. These changes are governed by rate parameters (e.g., survival, fecundity, and dispersal in case of the abundance state). The interactions between the living system and its environment represent the ecological mechanisms that create the temporal dynamics we observe. An analogous scheme could be drawn also for spatial patterns.

TABLE 1.1 Four Ecological Scales of Organization; the First Three of which Are Dealt with in This Book (see Royle and Dorazio, 2008, for the Fourth)

Scale of Organization = Type of Population	Description	
	Static (State Variable)	**Dynamic (Vital Rates)**
(1) One Site, One Species: Classic Population = Population of Individuals	Abundance N	Survival Probability (ϕ) Recruitment Rate (γ)
(2) Multiple Sites, One Species: Metapopulation = Population of (Local) Populations	N_s z_s Occupancy $\psi(=\Pr(N>0))$	Extinction Probability $(1-\phi_s)$ Colonization Rate (γ_s) Dispersal Rates
(3) One Site, Multiple Species: Community = Population of Species	N_k z_k Species Richness	Extinction Probability $(1-\phi_k)$ Colonization Rate (γ_k),
(4) Multiple Sites, Multiple Species: Metacommunity = Population of Communities	$N_{k,s}$ $z_{k,s}$ Species Richness	Extinction Probability $(1-\phi_{k,s})$ Colonization Rate ($\gamma_{k,s}$) Dispersal Rates

Notes: All four can be represented as "populations" by simply redefining what represents an "individual" making up that population. These scales are hierarchical in the sense that each lower scale is included in a higher one. This means, for instance, that the quantitative descriptors of a classic population (Scale 1) may also be applied to the components of a metapopulation (Scale 2). N, abundance; z, presence/absence indicator; k, index for species; s, index for site.

how changes in vital rates translate into changes in numbers, that is, of population size (e.g., Sibly and Hone, 2002). Moving up one level, but still considering the individual unit, we have a collection of sites in which individuals can live. The movement probability among the associated populations (dispersal) is now an additional vital rate. The state variable is the size of the different populations.

At Scale 2, we view a single local population (or more generally, an occupied spatial unit) among a collection of potentially occupied spatial units as the item, and thereby obtain a metapopulation (Hanski, 1998). The basic, static descriptor of a metapopulation is the set of N_s values, that is, classic abundance at each spatial unit s. A less information rich, yet easier to measure version is the occupancy state $z = I(N > 0)$, where $I()$ denotes the indicator function that evaluates to 1 for an occupied unit and zero for an unoccupied one. The population average of z_s is called "incidence" in the metapopulation literature (e.g., Hanski, 1994, 1998) or occupancy probability, ψ (e.g., MacKenzie, 2006). Occupancy and abundance are directly related to each other via $\psi = \Pr(N_s > 0)$, that is, occupancy probability is simply the probability that abundance at a site is greater than zero (Royle and Nichols, 2003). So, clearly, there is a sense in which "distribution and abundance"

in the book titles cited above is redundant; the characterization of a metapopulation by local abundance is fully sufficient and directly yields a description in terms of occupancy (Royle et al., 2005, 2007b).

Metapopulation ecology has been a part of ecology's mainstream for several decades now (Levins, 1969; Hanski, 1994, 1998; Hanski and Gaggiotti, 2004) and has been extremely influential in conservation biology, for instance, by highlighting the importance of random extinctions of local populations even at sites with suitable habitat, and consequently, by stressing the importance of connectivity among subpopulations as a means of avoiding permanent extinction of patches. In a similar vein, metapopulation biology provides the understanding for why currently unoccupied habitat patches may be as important for the long-term survival of a species as currently occupied ones (Talley et al., 2007). The dynamic descriptors of a metapopulation are analogous to those of a classic population, except that "individuals" (=occupied sites, local populations) can be reborn, that is, go extinct and yet later the site may be recolonized. Metapopulation-like dynamic models of occurrence proved insightful in epidemiology and disease ecology and have been used to model the spread of a disease (e.g., West Nile virus, Marra et al., 2004) or invasive species (Wikle, 2003; Hooten et al., 2007; Bled et al., 2011b).

An alternative way to quantify the total occurrence of an organism in some area is simply the sum of occurrences (i.e., $\sum_s z_s$); this represents a "population size" of occupied spatial units. Both the ratio ψ and the sum of z_s characterize the *range* or *distribution* of an organism. Ranges are the focus of macroecology and biogeography (Brown and Maurer, 1989; Gaston and Blackburn, 2000). Many ecological studies aim to predict species occurrence (i.e., z_s) from habitat or other local site attributes (e.g., Scott et al., 2002), either for fundamental reasons, for example, to study a species' niche (Guisan and Thuiller, 2005), or for applied reasons, for example, to predict the location of previously undetected occurrences, or to determine the most suitable sites for reintroduction projects. In essence, these models focus on the extent of a metapopulation, and the latest of them try to incorporate biological interactions (such as the possibility for an unoccupied site to become recolonized from an occupied site nearby; Guisan and Thuiller, 2005), thus bringing them increasingly closer to a classical and more mechanistic, metapopulation model of a species distribution.

Another increasingly common example of an occupancy study is a distribution atlas (Hagemeijer and Blair, 1998; Schmid et al., 1998; see review in Gibbons et al., 2007) that documents distribution ranges, for instance, by the presence or absence of a species in each cell of a grid. The data collected during such atlas studies, when repeated over time in the same area, have become an important raw material for studies documenting effects of climate change on species ranges (Thomas and Lennon, 1999; Huntley et al., 2007). Finally, occupancy is an important state variable for biodiversity monitoring,

for example, in the Swiss biodiversity monitoring program BDM (Weber et al., 2004, also see www.biodiversitymonitoring.ch), in amphibian monitoring (Pellet and Schmidt, 2005), and as one of the important and most widely used criteria by which the IUCN Red list status of a species is assessed (www.iucnredlist.org/about/red-list-overview#redlist_criteria).

Moving up another level among the ecological scales of organization, a community can be conceived of as a "population" of species at a single site (Table 1.1, Scale 3). A community can be described at a point in time by the species–abundance distribution, N_k (Engen et al., 2008). A simpler community description is the sum of individual species' occurrences, that is, species richness ($\sum_k z_k$). Species richness and its dynamic components are the central focus of research in many branches of ecology such as biogeography (Jetz and Rahbek, 2002), as well as conservation science, for instance, when looking for hotspots of species richness to direct conservation funds (Orme et al., 2005). Indeed, species richness is the most widely used measure of biodiversity (Purvis and Hector, 2000) and is frequently used in monitoring programs (e.g., Weber et al., 2004; Pearman and Weber, 2007).

At the highest level of ecological scales of organization (Table 1.1, Scale 4), a metacommunity is a set of population of multiple species at many sites. Metacommunities have recently taken center stage in ecology with the neutral theory of biodiversity (Hubbell, 2001; Gotelli and McGill, 2006). In terms of its quantitative description, a metacommunity can be dealt with fairly analogously to a community (Table 1.1).

Of course, not every ecologist focuses directly on the population descriptors of Table 1.1. For instance, evolutionary, behavioral, or physiological ecologists deal with the interactions among individuals and with the environment that may become the mechanisms determining the size (N) and dynamics of a population (Sibly and Calow, 1986; Stearns, 1992; Krebs and Davies, 1993; Sutherland and Dolman, 1994). However, N remains important implicitly: because in order to be ecologically relevant, any evolutionary, behavioral, or physiological mechanism must ultimately have at least the potential to affect N.

The modeling of these hierarchical scales may be conducted very naturally in a hierarchical manner, that is, a metapopulation can be modeled in terms of patch occupancy z_s or in terms of the local population size N_s. Similarly, its dynamics can be expressed by the survival and colonization probabilities of patches or by the survival and recruitment probabilities of the individuals occupying these patches and by their dispersal among the patches. Analogous alternative descriptions in terms of the state and the dynamics are possible for communities and metacommunities. One important descriptor of the dynamics of all four levels is the sustained rate of change of the system, or trend. Trend is a consequence of survival, recruitment, and dispersal probabilities and thus a derived quantity rather

than a driver of the system. Nevertheless, it is the simplest and most parsimonious description of system dynamics and of tremendous practical importance in many applications of population ecology, such as conservation biology and wildlife management (Balmford et al., 2003).

In summary, the three key state variables used to describe populations, metapopulations, communities, and metacommunities are abundance, occurrence (distribution), and species richness (Royle and Dorazio, 2008). From a pure modeling point of view, all three simply represent variants of a "population" that can be described by its size. In addition, there are the parameters that govern the dynamics of these state variables: survival/extinction, fecundity, colonization, and dispersal (immigration and emigration). Collectively, we call the study of these demographic quantities *population analysis*. Population analysis permeates a large part of ecology and of its applications such as conservation biology or fisheries and wildlife management. Indeed, it could be argued that population ecology, which we see as somewhat synonymous with population analysis, is a central pillar of the entire discipline of ecology.

1.2 GENESIS OF ECOLOGICAL OBSERVATIONS

A widely ignored consideration regarding all these varieties of the state variable, along with their dynamic rates, is that they are usually not directly observable; rather, "individuals" of all kinds can be overlooked; their *detection probability* is not perfect (i.e., $p < 1$; Schmidt, 2005). Therefore, a more accurate view of ecology is depicted in Fig. 1.2. This *hierarchical* view considers all observations in ecology as a result of two coupled processes: an ecological process, which usually is the focus of our interest, and an observation process, which is conditional on (i.e., whose result depends on) the result of the ecological process (Royle and Dorazio, 2006, 2008). In most ecological studies, the state of the system and its dynamics are latent. Therefore, they must be inferred from the observations O by modeling the main features of the observation process.

The ecological process itself is influenced by mechanisms that may be deterministic (e.g., habitat) or stochastic (e.g., demographic or environmental stochasticity) and that together determine the state of a system, for example, the population size, N. However, our observations of the system are also the result of an observation process, which may again be influenced by a variety of factors, among them deterministic (e.g., habitat-dependent observation errors) and stochastic mechanisms. Our observations in ecology are thus always a combination resulting from ecological *and* observation mechanisms.

For instance, assume that more birds are counted (i.e., *observed*) in habitat A than B. This can mean that there really *are* more birds in A than B, but

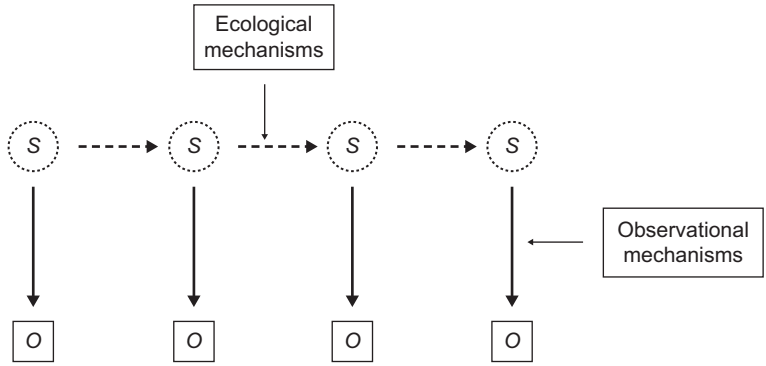

FIGURE 1.2 A hierarchical view of ecology that acknowledges the fact that state S is fully or partly unobservable (latent). Ecological mechanisms (representing the ecological process) affect the dynamics of the latent state, while observational mechanisms constitute the observation process. The observation process maps the latent state S (for instance, local population size N) on the observation O (for instance, a count). To emphasize the latent nature of the state process, arrows connecting states and circles around S are dashed.

it can also mean that birds are simply more *visible* in A than B or indeed any combination of the two. Similarly, not only the mean but also the variability of the observations is made up of these two components: one coming from the ecological process and the other from the observation process (e.g., nondetection error, sampling error; Royle and Dorazio, 2008, pp. 11–13). In most branches of ecology, we are thus faced with a situation where we have incomplete knowledge about an ecological system under study, and we must use error-prone observations to infer its characteristics, such as state variables or dynamic rates and the kind and strength of their interactions with the environment. In short, in the study of ecological systems, we must account for the fact that detection probability (p) of all three kinds of "individuals" shown in Table 1.1 is usually less than 1.

Direct inference based on the raw observations in ecology, and disregard of the observation process, may be risky. If nondetection error (rather than, say, false positives or double counting) represents the main feature of the observation process (i.e., $p < 1$), population size, distribution, or species richness will all be underestimated (Schmidt, 2005). Similarly, estimates of dynamic population descriptors will be biased. For instance, survival probabilities will be underestimated (Nichols and Pollock, 1983; Martin et al., 1995; Gimenez et al., 2008); extinction and turnover rates in metapopulations, communities, and metacommunities will be overestimated (Nichols et al., 1998b; Moilanen, 2002); and the perceived strength of a relationship between survival, abundance, or occurrence and environmental covariates will be underestimated (Tyre et al., 2003; MacKenzie et al., 2006); also see Section 13.2. It has not been sufficiently widely

recognized that what is typically called a distribution map in ecology (see, e.g., Scott et al., 2002), may in fact simply be a map of the difficulty with which an organism is found (Kéry et al., 2010; Kéry, 2011b). For instance, any spatially varying mechanism such as local density that causes a species to be more likely to be detected at some sites than at others will leave its imprint on a map of putative species distribution.

When imperfect detection is not accounted for in the modeling of ecological systems, the observed variation in the system (e.g., variance in population size) will typically be greater than the true variation in the ecological system (e.g., Link and Nichols, 1994). Some sort of variance decomposition must then be employed to separate true system variability from variability that is due to observation error (Franklin et al., 2000; De Valpine and Hastings, 2002). Such a partitioning of the observed variance is particularly important for population viability analyses (Lindley, 2003), investigations of density dependence (Dennis et al., 2006; Lebreton, 2009), and the setting of harvest regulations (Williams et al., 2002).

Consequently, we think that it is important in population analysis to include the essential features of the observation process when making inferences from imperfect observations about the underlying ecological process, for example, about the quantitative descriptors of all ecological scales of organization depicted in Table 1.1. Otherwise, we risk describing features of the observation process rather than of the ecological process we are really interested in. We need special data collection designs and methods of interpretation of the resulting data (i.e., models) that take explicitly into account the observation process to tease apart the genuine patterns in the ecological states from those induced in the observations by the observation process (MacKenzie et al., 2006). That is, when the quantities in Table 1.1 need to be studied directly, they must be *estimated* from, and cannot usually be equated with, the observed data.

To explicitly accommodate both the ecological and the observation process, an emerging and very powerful paradigm for population analysis is that of hierarchical models (Link and Sauer, 2002; Royle and Dorazio, 2006, 2008), sometimes also called state-space models (Buckland et al., 2004). One reason why these models are so useful for population analysis is that they simply replicate the hierarchical genesis of ecological data on animals and plants depicted in Fig. 1.2: one level in the model is the un- or only partially observed true latent state (e.g., being alive, abundance, or occurrence) and another level is the observation process, typically represented by detection probability p. Among the several advantages of hierarchical models is that they achieve a clear segregation of the observations into their two (or more) components. Thus, these models greatly foster intellectual clarity.

In this book, we follow Royle and Dorazio (2008) in emphasizing the distinction between the true state of an individual, a population, or

a community, and their observed state, and that the two are linked by an observation process, which imperfectly maps the former onto the latter. An explicit modeling of the observation process is thus essential to our approach of population analysis. Because we are convinced that most ecologists learn best by seeing examples, we next provide a brief numerical illustration for the observation process that shows why it is so important to consider it when making an inference in population ecology.

1.3 THE BINOMIAL DISTRIBUTION AS A CANONICAL DESCRIPTION OF THE OBSERVATION PROCESS

To better understand the key features of the observation process behind most ecological field observations, let us assume 16 sparrows live in our yard and that their population size was constant over a few weeks during which we make some observations (i.e., count them) to find out how many there are. Let us assume that there are no false-positive errors, only false-negative errors. This means that one sparrow cannot be counted for another and that another species cannot erroneously be identified as a sparrow. Let us further assume that each sparrow is independently observed or heard with a constant detection probability of 0.4. This means that if we step out into our yard 10 times, we will expect to see or hear (i.e., detect) that particular sparrow about 4 times. Of course, these are all abstractions of the real-world observation process, but they are very often plausible and adequate assumptions.

When we are interested in the total count of sparrows in our yard, then we have just defined a binomial random variable with sample or trial size $N = 16$ and so-called success probability $p = 0.4$. The binomial distribution is the mathematical abstraction of situations akin to coin flipping, where an event can either happen with a certain probability (p) or not (with $1 - p$), and we watch a number of times (N), all assumed independent, and count how many times (C) that event happens. The event here is the detection of an individual sparrow, and N is the latent state of the local population size of sparrows in the yard. Since detection is a chance process, we will typically not wind up with the same count all the time when we repeat the exercise. We can use a physical model, for example, the flipping of one or several coins, or a computer model to study the features of the observation process.

As we will see throughout the book, program R (R Development Core Team, 2004) is great for conducting quick and simple simulations to better understand a system. We first define the constants in the system:

```
N <- 16      # Population size of sparrows in the yard
p <- 0.4     # Individual detection probability
```

To simulate a single count, we simply draw a binomial random number with sample size N and success probability p (of course, most of you will get different simulated counts from those shown here).

```
rbinom(n = 1, size = N, prob = p)
 [1] 11
```

So the first time, we count 11 sparrows. The next time we go out into the yard, we count again:

```
rbinom(n = 1, size = N, prob = p)
 [1] 4
```

Now only four! This is a large difference to the previous count, so we count again a little later.

```
rbinom(n = 1, size = N, prob = p)
 [1] 9
```

This count is more similar to the first one. But perhaps we are a little worried about the variability in these counts. Were the lower counts made under somewhat inferior conditions? Or did we not pay so much attention to the counting then?

We can very simply use R to study the long-term behavior of the assumed observation process in our example: we simply draw a large number n of samples from the binomial random variable defined by N and p and summarize that sample. Let us draw one million then, since this does not cost us anything and R is free.

```
C <- rbinom(n = 10^6, size = N, prob = p)
```

Next, we describe that big sample in a graph (Fig. 1.3) and numerically to better understand some of the essential features of the counting process, that is, of the observation process behind our counts of a sparrow population of size 16.

```
mean(C)
[1] 6.404259
var(C)
[1] 3.842064
sd(C)
[1] 1.960118
hist(C, breaks = 50, col = "gray", main = "", xlab = "Sparrow count",
  las = 1, freq = FALSE)
```

This simple example illustrates several features about an observation process that is dominated by nondetection error (i.e., where misclassification and double counts are absent):

1. The typical count C is smaller than the actual population size N. Indeed, the mean of a binomial random variable and hence the expected count

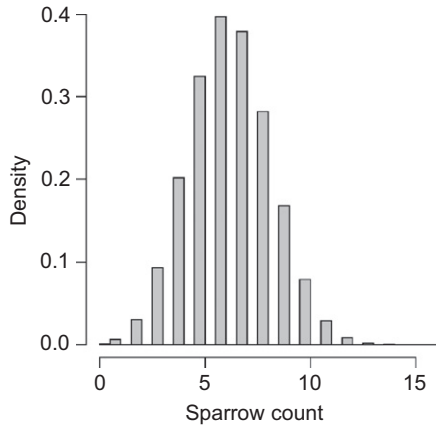

FIGURE 1.3 Frequency distribution of the observations, when 16 sparrows are counted repeatedly and each has independently a detectability of 0.4. The binomial distribution is our canonical model for the observation process involving detections of plant or animal "individuals", where "individual" can mean many different things (see Table 1.1).

of sparrows, equals the product of N and p. This should be 6.4 in the sparrow example and in our sample we get pretty close to that.

2. However, the counts vary quite a lot, *even under totally identical conditions*. Indeed, some of the counts in our sample were 0 and one was 16, meaning that sometimes not a single sparrow was detected but another time, all 16 of them were. Hence, there is nothing intrinsically wrong or inferior with smaller counts. Smaller counts may simply result from the random nature of the counting process in the presence of imperfect detection. Thus, any count with $p < 1$ will automatically tend to vary from trial to trial. Unless $p = 0$ or $p = 1$, it is impossible to eliminate that variation by the sampling design or standardization (though other components of variation may be eliminated).
3. Actually, not only the mean count but also the magnitude of the variation of counts is known from statistical theory. The variance is equal to the product of N, p, and $1 - p$ and thus should be around 3.84. Up to sampling variation, the observed variance of the 1 million counts of sparrows is identical to that.

False negatives will not only affect population counts and thus estimates of population size but also parameter estimates derived from the counts, such as survival or state-transition probabilities. In addition, the observation process will often be affected by explanatory variables and perhaps even by the exact same ones in which we are interested for the ecological process (see Figures 12.3. and 13.2.). Unless detection probability p is estimated then, any patterns in p will be perceived in the apparent state of the

ecological process. For instance, one can often read that state-space models (Chapter 5) correct for observation error. In truth, they only do so in a rather vague way. They only account for the binomial sampling variation around a mean count (i.e., Np), but cannot correct for the general bias in the counts relative to true population size, nor any patterns (e.g., over time) induced in counts by patterns in p (see Section 5.3). These latter two kinds of observation error (detection bias and patterns in detection) cannot be corrected for by the methods in Chapter 5 unless one has extra information about the detection process and uses the methods in Chapters 6, 10, 12, and 13.

We have claimed that the binomial distribution is the canonical description of the observation process. This is true in the sense that it underlies the vast majority of statistical methods that correct for imperfect detection in population analysis (Buckland et al., 2001; Borchers et al., 2002; Williams et al., 2002; Royle and Dorazio, 2008). However, other statistical distributions may be adopted as a description of the observation process in some cases, for instance, a beta-binomial distribution to account for nonindependent detections (Martin et al., 2011). The Poisson distribution is an appropriate description of the observation process if encounter frequencies rather than simply detection events are observed, for example, when we know that y_1 animals were observed once, y_2 twice, and so forth. In fact, there is a natural relationship between a Poisson and a binomial model of detection because the binomial detection probability may be expressed as the probability of observing Poisson counts greater or equal to one (Royle and Gardner, 2011). Finally, the negative binomial distribution may be adopted to model overdispersed encounter frequencies, such as arise from nonindependent detections of individuals, for instance, animals in groups (Boyce et al., 2001).

The binomial distribution is a useful starting point for modeling an observation process that is dominated by false-negative errors, that is, where false-positive errors are rare or absent. False-positive errors arise when one "individual" in Table 1.1 is mistakenly identified as another one. Variants of this kind of error include unoccupied sites being identified as occupied when modeling species distributions or mistakenly identifying one species for another when modeling communities.

In the fields of statistics covered in this book, almost all methodological development over the last decades has dealt with false-negative errors; false-positive errors were essentially assumed away. There are two reasons for this. First, false-negative errors are ubiquitous in ecology and probably much more widespread than false positives. Moreover, false positives can more easily be eliminated by the design of a study, by good training of a field crew or adequate calibration of a measuring device. For instance, any doubtful records, for example, detections of a species when estimating the size of a community (see Section 6.3) may simply be eliminated (J. D. Nichols, personal communication). This will weed out false positives at the risk of losing some valid detections of species. The latter will simply

lower detection probability but obviously not bias estimators from models that account for false-negative errors.

However, a second reason for why false positives have received much less attention than false negatives is that they are harder to deal with mathematically. Although on the whole, false positives may have less biasing effects in population analyses than false negatives, they are nevertheless important to account for in some situations. It has been shown that even relatively rare false-positive events may induce strong biases in site-occupancy models (Royle and Link, 2006). Interestingly, the effects of false positives often become stronger with larger sample sizes, for instance, when more surveys are conducted at each site in the context of site-occupancy sampling (see Chapter 13). Thus, it is important to watch out for this kind of error and if necessary account for it in population analyses. Therefore, it is encouraging that during the last decades a steadily increasing number of papers has dealt with false-positive errors, including Kendall et al. (2003), Nichols et al. (2004), Lukacs and Burnham (2005), Pradel (2005), Royle and Link (2006), Nichols et al. (2007), Conn and Cooch (2009), MacKenzie et al. (2009), Wright et al. (2009), Yoshizaki et al. (2009), Link et al. (2010), McClintock et al. (2010), and Miller et al. (2011). It is likely that we will see much more developments in this important topic in the near future.

In summary, we believe that it is important to account for the observation process when making an inference about any quantity in Table 1.1 for populations of animals or plants in population ecology, management, and conservation.

1.4 STRUCTURE AND OVERVIEW OF THE CONTENTS OF THIS BOOK

In this book, we deal with populations of the first three kinds in Table 1.1 and most intensively with classical population size and the demographic processes underlying its change (survival, fecundity, and dispersal or movement) and with the "population size" of occupied patches, that is, species distributions or occupancy, and its dynamic rates. We also see a little bit of species richness, that is, the size of the third kind of "population" (Section 6.3). The structure of this book is summarized in Table 1.2. We present this overview by distinguishing population studies at single sites from those at multiple sites, which we call studies with a metapopulation design (Royle, 2004c; Kéry and Royle, 2010). The scheme in Table 1.2 covers a fairly large part of the questions and models that ecologists might want to consider when analyzing populations of plants and animals.

We start with simple models for counts in Chapter 3 and then introduce more complexity via random effects in Chapters 4 and 5; it is the presence of

TABLE 1.2 Structure of This Book and Loose Chapter Overview

Quantity Modeled	Single Site	Multiple Sites
Distribution		
Apparent Distribution	–	Logistic Regression (Chapter 3)
True Distribution	–	Site-Occupancy Model (Chapter 13)
Abundance		
Apparent Abundance	Poisson GLM (Chapter 3)	
	Poisson GLMM (Chapter 4)	
	State-space Model (Chapter 5)	
	Integrated Population Model (Chapter 11)	
True Abundance	Closed Population Capture–recapture Model (Chapter 6)	Binomial mix Model (Chapter 12)
	Jolly–Seber Model (Chapter 10)	
	Integrated Population Model (Chapter 11)	
Vital Rates		
Survival Probability	Cormack–Jolly–Seber model (Chapter 7)	
	Ring-recovery Model (Chapter 8)	
	Multistate Model (Chapter 9)	
	Jolly–Seber Model (Chapter 10)	
	Integrated Population Model (Chapter 11)	
Fecundity/Recruitment	Poisson GLM (Chapter 3)	
	Jolly–Seber Model (Chapter 10)	
	Multistate Model (Chapter 9)	
	Integrated Population Model (Chapter 11)	
Movement Probability	–	Multistate Model (Chapter 9)
Leslie-Matrix Modeling	Integrated Population Model (Chapter 11)	

Notes: A distinction is made between models that can be applied to data from a single study site and those that require data from multiple sites. All former can also be applied to data from multiple sites.

these random effects, which converts the models in Chapter 3 to hierarchical models. The so-called state-space models of Chapter 5 attempt a partitioning of the total variation in the observations into one portion due to the ecological system and another portion due to observation error. We have claimed previously that the only accounting possible in this kind of model is for the

random sampling variation (due to imperfect detection) in the counts, but not for detection probability proper. Hence, all models in Chapters 3–5 deal with apparent, not true abundance, where "true" to us means "corrected for imperfect detection". In Chapter 6, we encounter for the first-time models that achieve a comprehensive accounting for detection error: these are the classical capture–recapture models for closed populations.

In Chapters 7–11, we move on to the quantities driving the changes in population size: demographic or vital rates. We note in passing that vital rates that represent probabilities such as survival are still often called rates, but this is not completely correct: a rate, unlike a probability, is not bounded by 0 and 1. We model survival probabilities based on two kinds of data, resighting and dead recoveries, respectively, as well as movement probabilities between states. These models again fully account for detection error. As an aside, in Chapter 3, we show a simple example of the modeling of another important dynamic rate: fecundity. Chapter 10 introduces the Jolly–Seber model, where inferences about abundance, survival, and recruitment from capture–recapture data can be made. In Chapter 11, we introduce a sort of synthesis of several models in preceding chapters and fit simple population models of the Leslie-matrix type. One important feature of these integrated population models is that by the joint modeling of several data sets and data types, parameters can often be estimated that are not identifiable with each data set alone.

Most methods for population analysis at single sites (in the left column in the body of Table 1.2) can also be used for populations at multiple sites. In the last two main chapters of the book, we present two classes of models that explicitly focus on distribution and abundance in a metapopulation setting: the binomial (or N-)mixture model for abundance (Chapter 12) and the site-occupancy model for species distributions (Chapter 13).

There are several important, recurring themes throughout our book. The three most essential ones are actually recurrent themes in much of applied statistical modeling:

1. Linear models
2. Generalized linear models
3. Random effects, or more generally, hierarchical modeling

First, in linear models, the mean of a response is thought to be made up of additive effects of covariates. This is one of the most widespread manner in which relationships between a response and explanatory variables are modeled. If you understand how to build a linear model using a design matrix (Chapter 3), you will achieve an extraordinary flexibility in your modeling. Second, generalized linear models (GLMs) use the same concept of linear models and the design matrix and carry that over to nonnormal responses, such as Poisson or binomial random variables. A Poisson or binomial response may also be the components in a

larger model, for instance, to model survival events over a number of years. We may then encounter a GLM as part of a larger model. We will meet GLMs all over in this book and give a concise introduction to them in Chapter 3. Finally, random effects are another key concept in applied statistical modeling and are introduced in Chapter 4. They give considerable flexibility to our modeling of variation and of correlations. Like GLMs, random effects, or hierarchical models, permeate this entire book. A GLM containing random effects is typically called a generalized linear mixed model (GLMM). We believe that if you understand linear models, GLMs, and random effects, then you understand a large part of applied statistical modeling. You then achieve an organic understanding of many of your models and will be able to build your custom models in a modular, creative, and efficient way (Kéry, 2010). In particular, you may then start to see hierarchical models as a natural way of describing complex stochastic systems by a nested sequence of component GLMs.

Other important concepts or techniques that you may be particularly interested in include the following:

- The distinction between an implicit and an explicit hierarchical model (Chapters 4–6, also see Royle and Dorazio, 2008)
- Data augmentation (Chapters 6 and 10)
- Posterior predictive checking of model adequacy (Chapters 7 and 12)
- The assessment of parameter estimability (Chapter 7)

We emphasize that by combining these basic concepts in a creative way, you will be able to create a large variety of statistical models that will help to obtain better inference from your data.

You will see plenty of WinBUGS and R code in this book. We use Courier font to highlight code. We comment our code quite extensively to make it easier to understand. In both R and WinBUGS, comments are flagged with a hash (#) sign, which means that the line following the hash is ignored.

1.5 BENEFITS OF ANALYZING SIMULATED DATA SETS: AN EXAMPLE OF BIAS AND PRECISION

One key feature of this book is that we work a lot with simulated data sets. There are tremendous benefits in doing so. As argued elsewhere (Kéry, 2010), the advantages of simulating (=assembling) data sets and then analyzing (=disassembling or breaking apart) them again are manifold:

1. Truth is known, so we know what to expect from the analysis.
2. We get a check for whether we have coded things correctly.
3. We can experience sampling error, that is, the variation in results. For instance, we may study long-run characteristics of estimators such as bias or precision (see example below).

4. In particular, repeatedly simulating data under some conditions and analyzing them provides an extremely flexible way of conducting power analyses.
5. We can check the effects of assumption violations: simulate data under a different model than that which is used to analyze a data set.
6. Finally, we can prove to ourselves that we understand an analysis: when you can assemble, that is, generate a data set under a model, you can also disassemble it, that is, break it down again in the analysis of that model.

As we will see many times in this book, we simulate data "from the inside out", that is, we first decide on values for any covariates and on coefficients that relate these covariates to the mean response, and then we add residual variation by drawing random variables with a given expectation and possibly a specified variance. When we disassemble the data set, we go the opposite way and break apart the full response into its ingredients, that is, covariate effects, random effects, etc. Arguably, you will only be able to simulate data under a model if you have really understood that model!

As an illustration, let us look at the concepts of bias and of the precision of an estimator. Both are often misunderstood by ecologists. One can often read an *unbiased estimate*. Strictly, this is wrong because estimates cannot be unbiased, only *estimators* can (Link and Barker, 2010). Estimators are the procedures that produce a particular estimate and if the estimates they produce are right on target, on average, an estimator is said to be unbiased. Right on target means that the mean of all estimates an estimator produces is the same as the population quantity (the parameter) estimated by that estimator. This may sound like counting green peas, and perhaps in ecological writings one could accept *unbiased estimate* as synonymous with the more accurate *unbiased estimator*. However, we must be able to make a clear distinction between a population quantity (a parameter), an estimator (a procedure), and a particular estimate of the population quantity produced by that procedure. Each individual estimate can be far off the target (the parameter), meaning the precision of the estimator is low, and yet the average of all estimates produced by the same procedure may be right on the mark, meaning the estimator is unbiased. Thus, the precision of an estimator expresses the similarity of repeated estimates produced by an estimator. Bias and precision are different aspects of the quality of an estimator: an estimator can be biased but precise, unbiased but imprecise (MacKenzie et al., 2006), or any other combination of these terms.

We next illustrate bias and precision by simulation: we simulate data from a large population, for which we want to estimate the population mean. We assume that we estimate body size of adult asp vipers (Fig. 1.4) by taking a sample of 10 snakes, measuring them, and calculating the mean

FIGURE 1.4 Asp viper (*Vipera aspis*), Germany, 2009 (Photograph by T. Ott).

length. We will see that the mean from one individual sample of 10 will hardly ever be right on target; rather, due to sampling variation (individuals in the population differ and we only sample 10 of them), the mean of 10 will be lower or higher than the true population parameter. However, *repeatedly* sampling that population, by measuring the length of 10 snakes, and producing a histogram of these means of 10 will show that the sample mean is an unbiased estimator of the population mean. Moreover, from statistical theory we know by how much, on average, means of 10 vary around the true population mean: this is given by the standard error of the mean.

So here we go: we assume that full-grown asp vipers in the Jura mountains average 65 cm with a standard deviation of 5. We sample 10 snakes, measure their length and use that sample to say something about the population mean.

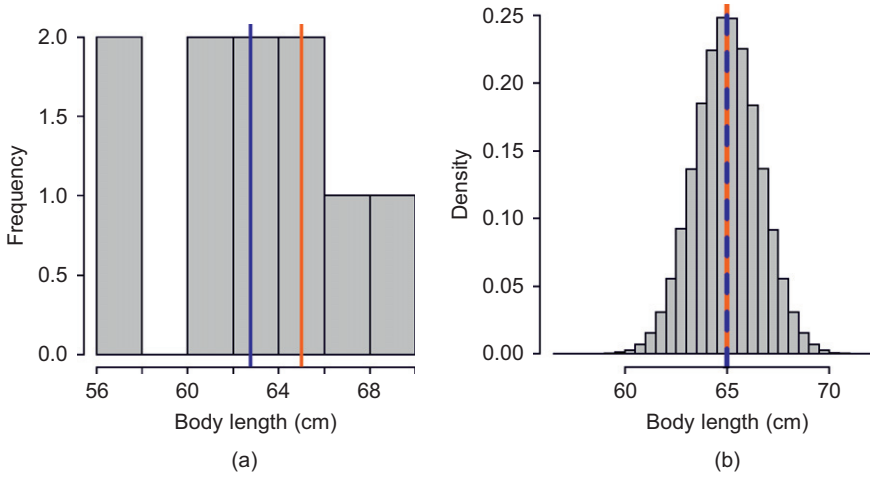

FIGURE 1.5 Histogram of body length of 10 asp vipers (a) and histogram of 1 million sample means of the body length of 10 asp vipers (b). Red line: population mean, blue line: mean of sample (means).

```
# Population values for mean and standard deviation of individual lengths
mu <- 65        # Population mean
sigma <- 5      # Population SD

# Draw a single sample of 10 and summarize it
x <- rnorm(n = 10, mean = mu, sd = sigma)
```

Figure 1.5a shows one data set; do not forget that yours will be different. The red line indicates the population mean (i.e., the value of the parameter that we want to estimate when calculating the sample mean), and the blue line is the mean for our particular sample of 10 snakes. We see that here the sample mean is too small relative to the population mean.

To see whether the difference between the blue line and the red line is simply sampling variation, we can repeatedly draw such samples and plot the means. Hence, let us use R and draw 1 million samples of 10 snakes and plot their means (Fig. 1.5b).

```
reps <- 10^6
sample.means <- rep(NA, reps)
for (i in 1:reps){
   sample.means[i] <- mean(rnorm(n = 10, mean = mu, sd = sigma))
   }

# Produce figure
par(mfrow = c(1, 2), las = 1)
hist(x, col = "gray", main = "", xlab = "Body length (cm)", las = 1)
abline(v = mu, lwd = 3, col = "red")
abline(v = mean(x), lwd = 3, col = "blue")
```

```
hist(sample.means, col = "gray", main = "", xlab = "Body length (cm)",
nclass = 50, freq = FALSE, las = 1)
abline(v = mu, lwd = 5, col = "red")
abline(v = mean(sample.means), lwd = 5, col = "blue", lty = 2)
```

We see that the distribution of the sample means in Fig. 1.5b is much more narrowly centered on the true population value. Moreover, from theory we know the spread of the distribution of sample means, that is, we know the precision of the mean as an estimator of the population mean: this is given by the standard error of the mean, that is, by $\mathrm{sd}(x)/\sqrt{n} = 5/\sqrt{10} = 1.58$. Let us see how closely we get to that in our simulation. Remember, the standard error of our parameter estimator (here, the sample mean as an estimator of the population mean) is the standard deviation of the distribution of the estimates.

```
sd(sample.means)
[1] 1.581443
```

This is pretty close. You can make your estimate of the standard error of the mean arbitrarily exact by increasing your simulation sample size (n).

We always need to clearly distinguish between sample and population quantities. A parameter is a feature of the population and we want to estimate it based on some sample statistic (something that can be measured in the sample), such as the sample mean. Sample statistics, and therefore our parameter estimates, are affected by sampling variation: not all individuals are identical and we only measure a sample, not the entire population. Sampling variation may be quantified by the standard error of our estimator; this quantifies the precision of an estimator. The standard error of the sample mean as an estimator of the population mean can be estimated by the population standard deviation divided by square root of the sample size. Here, the standard error is the standard deviation of a hypothetical distribution of parameter estimates (means) from a sample size of 10. It may be confusing that the standard error is in fact a standard deviation, but one characterizing the variability in a collection of *parameter estimates*, rather than the variability among individual *measurements*.

1.6 SUMMARY AND OUTLOOK

Distribution and abundance, along with their dynamic rates of change such as survival or extinction probability, and the factors that affect them, lie at the heart of the science of ecology. Alas, almost universally, neither state nor rate parameters in natural populations of animals and plants can ever be observed without error. In particular, detection error (manifest in the presence of false-negative observations) is a hallmark of ecological observations of populations. We have seen that the binomial distribution is the typical description of an observation process that is dominated

by false-negative errors, but that other statistical distributions may sometimes be appropriate, for example, the beta-binomial, Poisson, and negative binomial. We have seen that the opposite type of error, false positives, has received far less attention in the literature so far, but that it is the focus of much recent work. We think that the observation process should be accounted for as much and as often as possible in the analysis of wild populations. Therefore, much of this book covers methods that attempt a clean partitioning of the ecological and the observation processes that underlie ecological observations. This partitioning is often achieved using hierarchical models, where separate model components describe the latent ecological process and the observation process. We are big fans of analyzing simulated data sets; perhaps the single biggest advantage of using simulated data sets is that the mere act of simulating data enforces an almost complete understanding of a model. Next, after having motivated the population analysis part in the book title, we give a very brief review of the other half of our book's title, namely "Bayesian" and "WinBUGS".

1.7 EXERCISES

1. Detection probability: convince yourself that very few quantities in nature are ever perfectly detectable. Stand at the window for 1 min and make a list of all the bird species that you see. Repeat this once or twice, then compare the lists among times and observers. If detection were really perfect (for all species, at all times, for all observers, etc.), then all the lists would be the same for all observers and it would not matter for how long you watched. Essentially, you would detect all species instantaneously.

   You may conduct that exercise with a quantity of your choice: for example, the number (or identity) of people in your office hall, and the number of people in your bus. Alternatively, you could also count the brands of cars that pass in front of you.
2. Intrinsic variability of counts: it is our experience that ecologists often do not realize that counts vary intrinsically as soon as detection is imperfect. Moreover, counts that vary more are sometimes viewed as being of inferior quality than counts that vary less, or, that the observers producing these counts are better or worse. It is true that, everything else equal, sloppy counts will usually be more variable than those made by a more dedicated observer. However, owing to the mean–variance relationship in binomial counts, two of the most important factors affecting the variability of counts is (1) the number of things available for counting (i.e., N) and (2) detection probability, p. Produce a plot or a table that makes you better understand these relationships.