Dealing with uncertainty requires the proper tools, primary among them, the rules of probability and an understanding of probability distributions. Equipped with these, ecologists can see their way to revealing analysis, regardless of the idiosyncrasies of the particular research problem at hand. These analysis extend logically from first principles rather than from a particular statistical recipe. In the sections that follow we describe these vital tools. Our approach is to start with the definition of probability and develop a logical progression of concepts extending from this definition to a fully specified and implemented Bayesian analyses appropriate for a broad range of research problems in ecology.

3.2 Rules of probability

Ecological research requires learning about quantities that are unobserved from quantities that are observed. Any quantity that we fail to observe, including quantities that are observed imperfectly, involves uncertainty. The Bayesian approach treats all unobserved quantities as random variables to capture that uncertainty. A random variable is is a quantity that can take on values due to chance. A random variable does not have a single value, but instead can take on a range of values, with its "chance" governed by a probability distribution. [2]

It follows that all random variables have probability distributions even though these distributions may be unknown to us. The rules of probability determine how we gain insight about random variables from the distributions that govern their behavior. Understanding these rules lays a foundation for the remainder of the book. This material is not exactly gripping, but we urge you not to skip this section or rush through it unless you already well grounded in formal principles of probability. Understanding these principles will serve you well. We summarize the rules of probability in Table ___ and describe them below.

We start with the idea of a *sample space, S,* consisting of a set of all possible outcomes of an experiment or a sample, shown graphically as a polygon with a specific area (Figure 3.2.1). One

[2] There is some argument among statisticians about whether states of ecological systems and parameters governing their behavior are truly random. Ecologists with traditional statistical training may object to viewing states and parameters as random variables. These objections might proceed like this. Consider the state, "the average biomass of trees in a hectare of Amazon rainforest." It could be argued that there is nothing random about it, that at any instant in time there *is* an average biomass that is fixed and knowable at that instant–it is determined, not random. This is true, perhaps, but the practical fact is that if we were to attempt to know that biomass, which is changing by the minute, we would obtain different values depending on when and how we measured it. These values would follow a probability distribution. So, thinking of unknowns as a random variable is a scientifically useful abstraction with enormous practical benefits, benefits we will demonstrate in later chapters. We will leave arguments about whether states and parameters are "truly random" to metaphysics.

of the possible outcomes of the experiment or sample is the random variable, "event A," a set of outcomes, which we also depict as a polygon (Figure 3.2.1). The area of A is less than the area of S because there are possible outcomes that it does not include. The area of A is proportional to the size of the set of outcomes it *does* include. It follows that the probability of A is simply the area of A divided by the area of S.

We now introduce a second event, B, to illustrate the concept of conditional, independent, and disjoint probabilities. Consider the case when we know that the polygon defining event B intersects with the A polygon (Figure 3.2.1 upper panel) and, moreover, we know that the event A has occurred. We ask, what is the probability of the new event B given our knowledge of the occurrence of A? The knowledge that A has occurred does two things. It shrinks the sample space from all of S to the area of A – if we know A has occurred, we know that everything outside of A has *not* occurred, so in essence we have a new, smaller space for defining the probability of A. Knowing that A has happened also affects what we know about B – we know that everything within B outside of A has not occurred (Figure 3.2.1). This means that

$$\Pr(B|A) = \frac{\text{area shared by } A \text{ and } B}{\text{area of } A} = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{\Pr(A, B)}{\Pr(A)}. \tag{3.2.1}$$

Using the same logic,

$$\Pr(A|B) = \frac{\text{area shared by } A \text{ and } B}{\text{area of } B} = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A, B)}{\Pr(B)}. \tag{3.2.2}$$

The expression $\Pr(A|B)$ reads "the probability of A conditional on knowing B has occurred." The bar symbol (i.e., $|$) reads "conditional on" or "given", expressing the dependence of event A on event B – if we know B our knowledge changes what we know about A. It is important to note that $\Pr(A|B) \neq \Pr(B|A)$. The expression $\Pr(A, B)$ reads the *joint* probability of A and B and is interpreted as the probability that both occur. We will make important use of the algebraic rearrangement of equations 3.2.1, and 3.2.2 to expand their joint probability,

$$\begin{aligned} \Pr(A, B) &= \Pr(B|A)\Pr(A) \qquad (3.2.3) \\ &= \Pr(A|B)\Pr(B). \end{aligned}$$

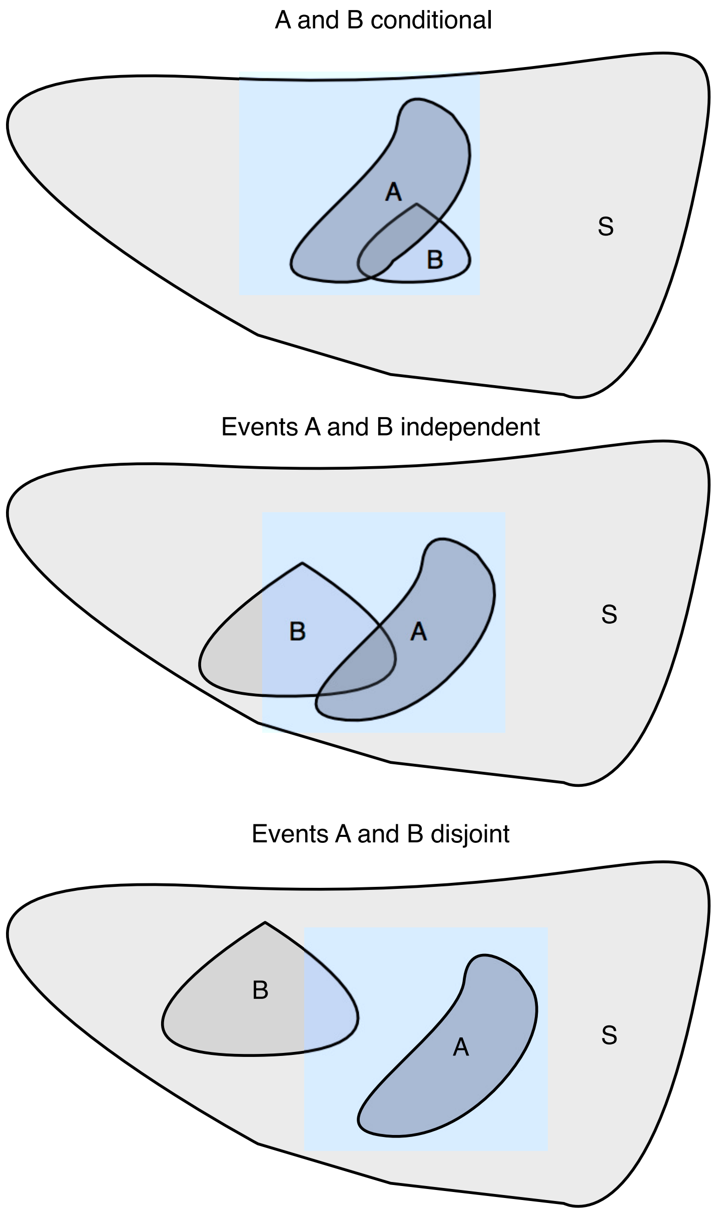

Figure 3.2.1: Illustration of conditional, independent, and disjoint probabilities. The area S defines a sample space including all of the possible outcomes of a sample or an experiment. There are two sets of realized outcomes, A and B. The area of each event is proportional to the size of the set. The probability of $A = \frac{A}{S}$ and the probability of $B = \frac{B}{S}$. Knowledge that the event A has occurred influences our estimate of the probability of B when the intersection of the two events gives us new information about the probability of B. In this case, the probability of B is conditional on A and vice versa (upper panel). There are cases where the events intersect, but there is no new information. In this case the probability of B given A is the same as the probability of B because the $\frac{B \text{ intersect with } A}{A} = \frac{B}{S}$. In this case, we say that A and B are independent (middle panel, areas drawn approximately). If there is no intersection, then the events are disjoint. Knowledge that A has occurred means that we know that B has not occurred.

There will be cases when the area defining the two events overlaps but there is no new information given by the knowledge that either event has occurred (Figure 3.2.1 middle panel). In this case the events are *independent*. Events A and B are independent if and only if

$$\Pr(A|B) = \frac{A \text{ shared by } A \text{ and } B}{\text{area of } B} = \frac{\text{area of } A}{\text{area of } S} = \Pr(A) \tag{3.2.4}$$

or equivalently

$$\Pr(B|A) = \Pr(B). \tag{3.2.5}$$

Using equation 3.2.1 and 3.2.2 we can substitute for the conditional expressions in 3.2.4 and 3.2.5. A little rearrangement gives us the joint probability of independent events:

$$\Pr(A, B) = \Pr(A)\Pr(B). \tag{3.2.6}$$

It is important to throughly understand the difference between the definition of the joint probability of events that are independent (e.g., equation 3.2.6) and those that are not (i.e., equations 3.2.3).

When events are disjoint, there is no intersection between them (Figure 3.2.1 lower panel). In this case, the knowledge that one event has occurred means that the other event has *not* occurred.

We may also be interested in the probability that one event or the other occurs (Figure 3.2.1), which is the total area of A and B without the area they share, i.e.,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A, B). \tag{3.2.7}$$

When A is independent of B,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A)\Pr(B), \tag{3.2.8}$$

but if they are conditional,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A|B)\Pr(B) \tag{3.2.9}$$

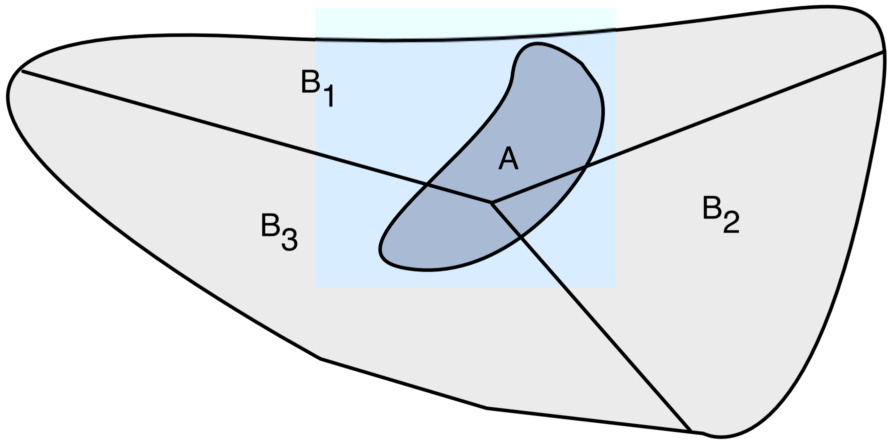

Figure 3.2.2: Illustration of the law of total probability.

or equivalently,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(B|A)\Pr(A). \tag{3.2.10}$$

If A and B are disjoint, then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \tag{3.2.11}$$

which is simply a special case of equation 3.2.7 where $\Pr(A, B) = 0$.

The final probability rule we consider applies to the case when can partition the sample space into several, non-overlapping events (Figure 3.2.2). We define a set of events $\{B_n : n = 1, 2, 3, \ldots\}$, which taken together, cover the entire sample space, $\sum_n B_n = S$. We are interested in the event A that overlaps one or more of the B_n. The probability of A is

$$\Pr(A) = \sum_n \Pr(A \mid B_n)\Pr(B_n). \tag{3.2.12}$$

Equation 3.2.12 is called the law of total probability.

3.3 Factoring joint probabilities

It is hard to avoid a modicum of tedium in describing the rules of probability, but there is a very practical reason for understanding them. They allow us to deal with complexity. These rules permit us to take complicated joint distributions of random variables and break them down in manageable chunks that can be analyzed one at a time as if all of the other random variables were known and constant. The importance of this idea and its implementation will be developed throughout the book. Here, we establish its graphical and mathematical foundation.

Consider the networks shown in Figure 3.3.1. A Bayesian network (also called a directed acyclic graph) depicts dependencies among random variables. The random variables in the network are called nodes. The nodes at head of the arrows are charmingly called children and the tails, parents. Bayesian networks show how we factor the joint probability distribution of random variables into a series of conditional distributions, thereby representing an application of equation 3.2.3 to multiple variables (Figure 3.3.1). This factoring is how we simplify problems that would otherwise be intractably complex.

Bayesian networks are great tools for thinking about relationships in ecology and for communicating them **A box with an example here?** (e.g. Figure 1.2.1). They are useful for thinking because they allow us to visualize a complex set of relationships, encouraging careful consideration of how knowledge of one random variable informs us about the behavior of others. They lay plain our assumptions about dependence and independence. A properly constructed Bayesian network provides a detailed blueprint for writing out a joint distribution as a series of conditional distributions Nodes at the heads of arrows are on the left hand side of conditioning symbols, those at the tails of arrows are on on the right hand sides of conditioning symbols, and any node at the tail of an arrow without an arrow leading into it must be expressed unconditionally, e.g., $P(A)$. The network provides a graphical description of relationships that is easier to see than the corresponding mathematical description, facilitation communication of ecological ideas underlying the network[3].

The mathematics allowing factoring of joint distributions extend directly from the rules of probability we developed above. Given the vector of jointly distributed random variables $\boldsymbol{z} = (z_1, \ldots, z_n)$, their joint probability satisfies:

$$\Pr(z_1, \ldots, z_n) = \prod_{i=1}^{n} \Pr(z_i | \{p_i\}) \tag{3.3.1}$$

where $\{p_i\}$ is the set of parents of node z_i and all of the terms in the product are independent. Independence of the terms in equation 3.3.1 is assured if the equation been properly constructed from a Bayesian network and the network shows relationships that are conditional and independent.

[3] At least Hobbs thinks so. Hooten prefers the equations.

A somewhat more formal way to say the same thing is to generalize the conditioning rule of probability for two random variables (equation 3.2.2) to factor the joint distribution of any number of random variables using

$$\Pr(z_1, z_2, ..., z_n) = \Pr(z_n|z_{n-1}, \ldots, z_1) \ldots.. \Pr(z_3|z_2, z_1) \Pr(z_2|z_1) \Pr(z_1), \tag{3.3.2}$$

where where the components z_i may be scalars or sub-vectors of $\boldsymbol{z}$ and the sequence of the conditioning is arbitrary. It is important to see the pattern of conditioning in equation 3.3.2. [4]. We can use the independence rule of probability (equation 3.2.4) to simplify conditional expressions in equation 3.3.2 for random variables known to be independent. For example, if z_1 is independent of z_2 then $\Pr(z_1|z_2)$ simplifies to $\Pr(z_1)$. If z_1 and z_2 depend on z_3 but not each other, then

$$\Pr(z_1, z_2, z_3) = \Pr(z_1|z_2, z_3) \Pr(z_2|z_3) \Pr(z_3) \tag{3.3.3}$$

simplifies to

$$\Pr(z_1, z_2, z_3) = \Pr(z_1|z_3) \Pr(z_2|z_3) \Pr(z_3). \tag{3.3.4}$$

Another example of this kind of simplification is shown graphically and algebraically in Figure 3.3.1 V and VI. Don't let the formalism in this paragraph put you off. It is simply a compact way to say what have already shown graphically using Bayesian networks, which for many ecologists will be more transparent.

3.4 Probability Distributions

3.4.1 Mathematical foundation

3.4.1.1 Probability mass and density functions

The Bayesian approach to learning from data using models makes a fundamental simplifying assumption: we can divide the world into things that are observed and things that are unobserved. Distinguishing between the observable and unobservable is the starting point for all analyses. We

[4]We say the sequence is arbitrary to communicate the idea that the ordering of the specific z_i is not required for equation 3.3.2 to be true. In other words, z_n doesn't need to come first. However, the word arbitrary should not be takes to mean capricious. As we learn, it our understanding of the *biology* that determines what is conditional on what, ultimately determining the sequence of conditioning.