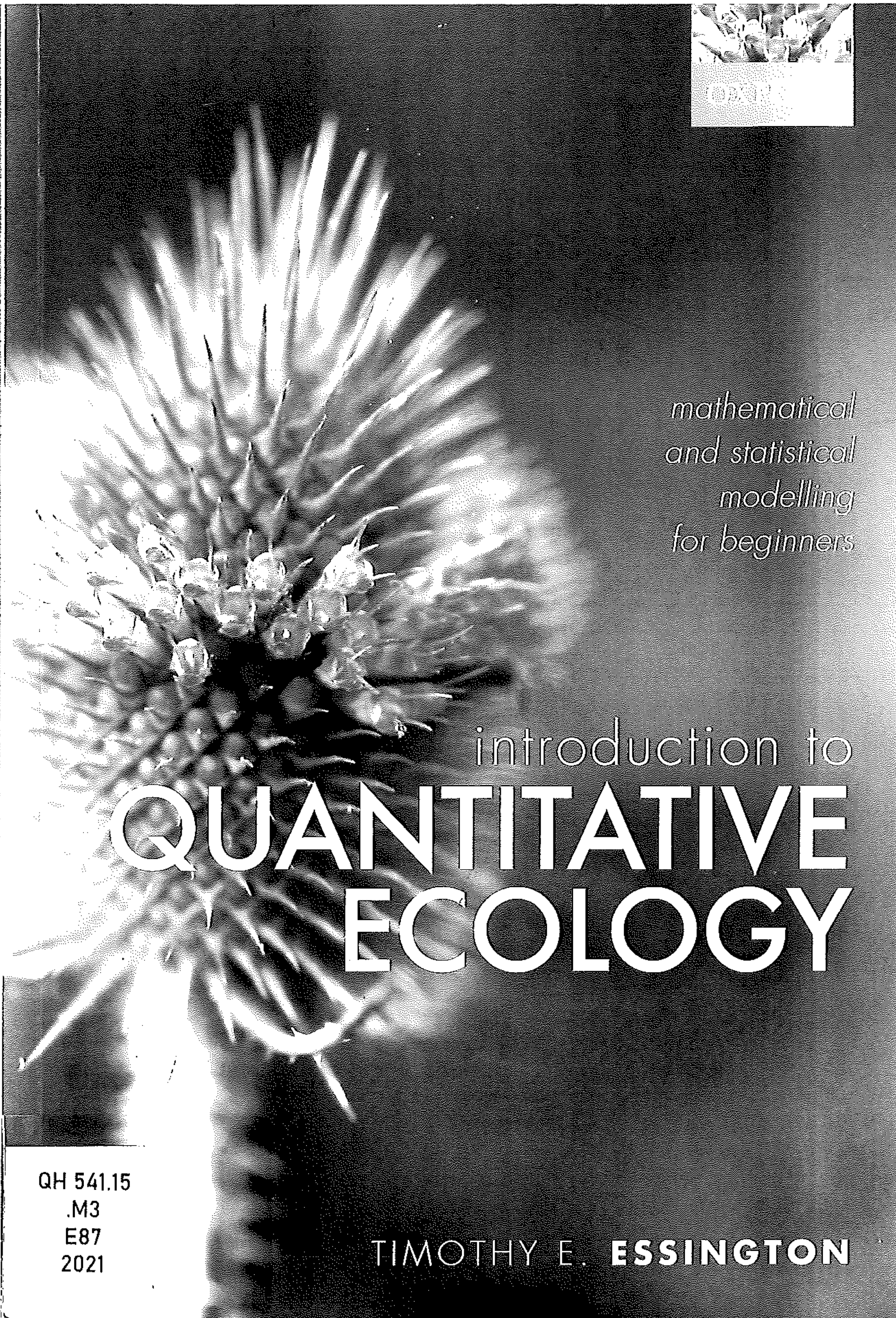

mathematical and statistical modelling for beginners

introduction to QUANTITATIVE ECOLOGY

TIMOTHY E. **ESSINGTON**

7

Random Variables and Probability

7.1 Introduction

Random variables are the core concept in understanding probability, parameter estimation, and model selection. It is a term that you may learned before, but you may not have fully understood its meaning. Here we'll review the basic concept and identify the two main kinds of random variables.

7.1.1 What is a random variable?

Suppose you have a pair of dice. You shake them up, throw them on the table, and add up the two numbers. The outcome is a *random variable*. It takes on different values based on a set of probabilities that are defined for each value. A more formal definition is, any outcome that follows from a given trial, experiment, or sample event. It doesn't matter if it is a lab experiment or a field experiment. In other words, all of that beautiful data you collected? From now on, think of them as a beautiful collection of random variables.

There are two basic categories of random variables. Discrete random variables are those whose outcomes can only take discrete values. The above example of throwing two dice is a good example: there are eleven possible outcomes: all of the integers between 2 and 12. There is no way to throw the dice and get a 8.467. Continuous random variables are those whose outcomes can take any value. For example, suppose while you where throwing the dice, someone was timing how long it took for the dice to stop moving once they left your hand. This could be anything: 2.34524 seconds, 4.02042 seconds, or whatever.

Put another way, anything that you count is a discrete variable. Anything that you measure is a continuous variable.

Below we will briefly review random variables, because they form the basis of all statistics. If you feel that you have a good grasp of these ideas, skip ahead to the next section. In my experience, knowledge of statistics, probability, and random variables is lot like helium in a balloon. While you're taking a course, the balloon is full and buoyant. Come back a year later and that balloon is half-full and lying on the ground. This information just leaks out of our heads for some reason.

Introduction to Quantitative Ecology: Mathematical and Statistical Modelling for Beginners. Timothy E. Essington,
Oxford University Press. © Timothy E. Essington 2021.
DOI: 10.1093/oso/9780192843470.003.0007

Properties of random variables

If a random variable, X, is discrete (e.g., the number of observed predation events), then we define the probability that an observation has some quality A as

$$P(X \text{ is } A) = \sum_{x_i \in A} P(x_i)$$

The right-hand side of this equation is the *Probability Mass Function.* It tells you the probability of each possible outcome. Say we wanted to know the probability of the dice above coming up either 5 or 6; this would simply be the sum of the probability of getting a 5 ($P(x_i = 5)$) plus the probability of getting a 6 ($P(x_i = 6)$).

If X is a continuous random variable (e.g., time until the dice stop moving), then

$$P(X \text{ is } A) = \int_{X \in A} f(x)dx$$

where the left-hand side of the equation denotes the probability that X lies in the region A (defined by lower and upper bounds of x), and $f(x)$ is the *Probability Density Function.* Note that there is an integral (instead of a summation) because x is a continuous variable.

The expected value of the discrete random variable X that can take values $x_1, x_2, x_3, \ldots, x_n$ is:

$$\mathbb{E}[X] = \sum_{i}^{n} x_i P(x_i)$$

where x_i are all of the outcomes that X can take.

The expected value of a continuous random variable X that can take any value x is

$$\mathbb{E}[X] = \int x f(x)dx$$

Random variables have various "moments," which are defined where the jth moment is $\mathbb{E}[X^j]$. The mean of X is the first moment and equals $\mathbb{E}[X]$ (e.g., the expected value of X). The second moment is similar, but, rather than find $\mathbb{E}[X^2]$, we instead look at the "centered" moments (meaning that they are centered around the mean). Thus, the second moment is $\mathbb{E}[(X-\mu)^2]$, where μ is the mean. In other words, the second moment tells us the expected squared deviation from the mean. This is called the variance, and the square root of the variance is called the standard deviation.

7.2 Binomial

The binomial probabilit
of discrete trials, where
example of a binomial pı
means the coin comes u
is a 50% probability of tl
density function to estin
and so on.

7.2.1 Key things

The outcome is discrete
discrete trials (the coin tc
It assumes that trials are
of getting "heads" is the

7.2.2 The probab

We denote the outcome
heads, N is the number c
any given coin toss com
outcomes between X = (

where the notation $N!$ an
side says that X is a rand
tells us the probability th

7.2.3 When woulc

The coin toss is the easiest
where you are counting tl
trials, where there is a coı
instance, the probability o
and successful seed germi
that happen over a series

7.2 Binomial

The binomial probability mass function describes the number of "successes" in a series of discrete trials, where all trials have the same probability of success. The simplest example of a binomial probability mass function is a series of coin tosses, where a success means the coin comes up heads. If you were to flip a coin ten times, and each time there is a 50% probability of the coin coming up heads, you would use the binomial probability density function to estimate the probability of getting zero heads, one heads, two heads, and so on.

7.2.1 Key things about this distribution

The outcome is discrete (you can never have 4.78 heads) and derives from a series of discrete trials (the coin tosses, which are also discrete, as you can't flip a coin 4.78 times). It assumes that trials are independent and have identical properties (i.e., the probability of getting "heads" is the same for each toss).

7.2.2 The probability mass function

We denote the outcome as X. Here, X refers to the number of times the coin turns up heads, N is the number of trials (the number of coin tosses), and p is the probability that any given coin toss comes up heads. From this, we can calculate the probability of all outcomes between $X = 0$ and $X = N$:

$$P(X=x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} \tag{7.1}$$

where the notation $N!$ and $x!$ refer to N factorial x factorial, respectively. The left-hand side says that X is a random variable that can take any value x, while the right-hand side tells us the probability that it will take any value x.

7.2.3 When would I use this?

The coin toss is the easiest way to think about this, but it applies to any analogous situation where you are counting the number of times a given event happens over several discrete trials, where there is a constant probability of the event happening for each trial. So, for instance, the probability of death over survival, a prey encounter over no prey encounter, and successful seed germination over unsuccessful germination are all discrete outcomes that happen over a series of discrete trials.

7.2.4 Properties of the function

The expected value of the outcome is $\mathbb{E}[X] = pN$, which makes sense because it is saying that if the probability of a coin toss heads is p, and you flip the coin N times, on average you will get p times N heads.

The variance of the binomial equals $Np(1-p)$. This means the variance grows as N gets larger. This also makes some sense—there simply aren't that many different outcomes if you only flip the coin once or twice, but there are many different outcomes that can happen if you flip the coin 100 times. The variance is also determined by the product $p(1-p)$. If p is close to 0 or close to 1, this product will be fairly small. Think about why the variance of outcomes is small if p is very nearly 0 or very nearly 1.

7.2.5 Example

Suppose you have a stage-structured population, and you have twenty individuals in a reproductively immature stage. You wish to project the number of individuals that transition into a mature stage over some discrete time interval. The number of individuals that mature is a random variable described by a binomial probability distribution, because the randomness acts on discrete individuals. If the probability is low, $p = 0.1$ (which might happen for small or young individuals), most of the individuals will not become mature, but you may reasonably get as many as five or six that become mature (figure 7.1) over some time increment. If, instead, the maturation probability is $p = 0.5$, then there is a much wider range of outcomes (figure 7.1), consistent with the finding that the variance is proportional to p times $(1-p)$.

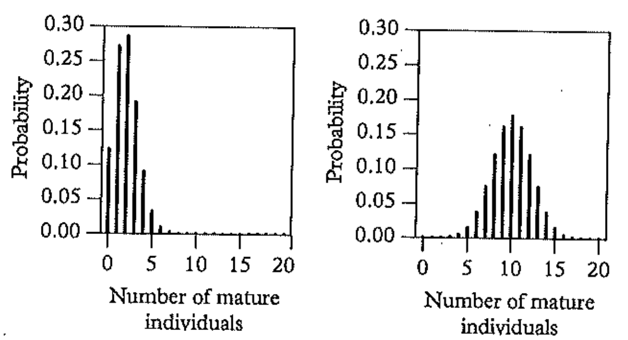

Figure 7.1 *The probability of seeing* x *individuals become mature out of 20 individuals, when the probability of maturation is p = 0.1 (left), and p = 0.5 (right).*

7.3 Poisson

The Poisson probabilit
continuous sampling fr
success. The number o
horribilis) observed usi
mass function, the obse
for example, a certain a
one location for a set
observed. The probabili
of grizzly bears.

7.3.1 Key things

The key things about t
see 4.78 bears perhaps)
either time or space. Als
event density) is consta

7.3.2 The probal

We denote the outcome
the area or amount of ti
this, we can calculate th

Notice that r and t a
model written like this:

Here r times t has bee
numbers of bears prese
Remember that the
calculate the probability
of surveys over space is t
situation where you are c
sampling frame and the

7.3 Poisson

The Poisson probability mass function describes the number of "successes" over a fixed, continuous sampling frame where, at any moment in time or space, there is a some rate of success. The number of "successes" might be the number of grizzly bears (*Ursus arctos horribilis*) observed using an aerial survey. Unlike the case for the binomial probability mass function, the observations come from looking over a continuous sampling frame, for example, a certain amount of area that is surveyed. Alternatively, one could monitor one location for a set period of time and count the number of grizzly bears that are observed. The probability of seeing a grizzly bear at any moment is related to the density of grizzly bears.

7.3.1 Key things about this distribution

The key things about this distribution are that the outcome is discrete (you can never see 4.78 bears perhaps) and that it derives from a continuous observation window, over either time or space. Also, it presumes that the probability of an outcome (or, rather, the event density) is constant over space or time.

7.3.2 The probability function

We denote the outcome as X. Here, X refers to the number bears that are observed, t is the area or amount of time that is surveyed, and r is the density of grizzly bears. From this, we can calculate the probability of all outcomes for $X \geq 0$:

$$P(X=x) = \frac{e^{-rt}(rt)^x}{x!} \tag{7.2}$$

Notice that r and t are always multiplied by each other. Sometimes you will see the model written like this:

$$P(X=x) = \frac{e^{-\lambda}(\lambda)^x}{x!} \tag{7.3}$$

Here r times t has been replaced by λ. You would use this if you cared about the total numbers of bears present in a defined space, rather than the density.

Remember that the left-hand side says that X is a random variable, and we can calculate the probability that X takes any value x, using the right-hand side. The example of surveys over space is the easiest way to think about this, but it applies to any analogous situation where you are counting the number of times an event happens over a continuous sampling frame and there is a constant probability of the event happening.

7.3.3 When would I use this?

You would use this whenever your observations are discrete counts of events over a continuous time or spatial observation window. Most ecological surveys, for instance, have a defined space. You might place a quadrat on a forest floor and count the number of seedlings in the quadrat. The number of seedlings could be described by a Poisson. Needless to say, this is a very common probability density function in ecological research.

7.3.4 Properties of the function

The expected value of the outcome is $\mathbb{E}[X] = rt = \lambda$, which makes sense because it is saying that if the rate of a seeing a "success" is r per given area or time, and you search over area $= t$ or time $= t$, then, on average, you will see r times t "successes." If you think of r as a density, and the true grizzly bear density is 3 bears/km^2, and you search 10 km^2, you expect, on average, to see 30 bears. The variance of the Poisson also equals rt. This is a bit less intuitive, but it means that the ratio of mean to variance always equals 1. This is quite different from many other probability density functions, where the mean and variance are uncoupled.

7.3.5 Example

You have a camera attached to a foraging emperor penguin to examine its foraging behavior. From the video feed, you have continuous-time monitoring of the number of attacks on fish prey. If the Poisson rate parameter is 0.1 min^{-1} and you have fifty minutes of observation, then you might reasonably expect to see from zero to about twelve fish attacks (figure 7.2), although in rare cases you might see many more than twelve. If the Poisson rate parameter is 0.3 min^{-1}, then there is a much wider range of likely outcomes, ranging from five to twenty-five (figure 7.2). This is consistent with our expectation that the variance increase (in this case) is equal to the product r times t.

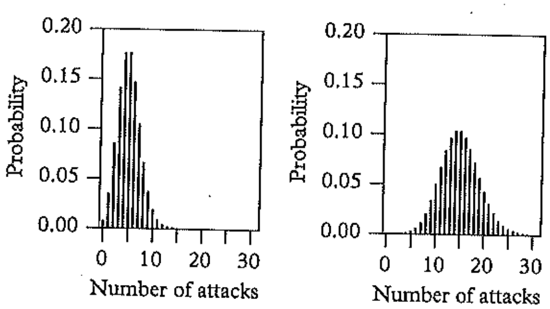

Figure 7.2 *Probability of observed attacks by penguins when r = 0.1 (left) and r = 0.3 (right).*

7.4 Negative bi

The negative binomial
would be used: it desc
that is continuous in ti
some probability of suc
describes patchiness in
Another way to think
example given above va
of bears is very high; ir
flexible than the Poisso
be equal (the variance w

7.4.1 Key things

The outcome is discrete
time or space. Unlike th
bear at any given locatio
varies across the samplin

7.4.2 When woul

You would use this in m
especially when the the
patchy encounters, the n

7.4.3 The probab

The negative binomial is
in different ways and car
and the one that explair
discrete trials, before a su
nod to its roots in the bi
counts of failures (the "n
is different. Instead, it sta
space and time, with son
discrete outcome x. In ou
is the area or amount of t
over some units of time o
but it sets the level of pat
of k produce very patchy
probability of all outcome

7.4 Negative binomial

The negative binomial is used in contexts that are similar to those in which the Poisson would be used: it describes the number of "successes" over an observation window that is continuous in time or space, where, at any moment in time or space, there is a some probability of success. The difference is that there is an additional parameter that describes patchiness in success over time or space.

Another way to think about this is that the grizzly bear density (the parameter r) in the example given above varies over the sampling frame. Thus, in some areas, the density of bears is very high; in others, it is very low. This makes the negative binomial more flexible than the Poisson, because the variance and the mean are no longer required to be equal (the variance will be equal to, or exceed the mean).

7.4.1 Key things about this function

The outcome is discrete and derives from a continuous observation window, over either time or space. Unlike the Poisson, it does not presume that the rate of seeing a grizzly bear at any given location is constant. We tend to use it when we think the event rate r varies across the sampling frame.

7.4.2 When would I use this?

You would use this in most of the same contexts for which you would use the Poisson, especially when the the variance is greater than the expected value. That is, for very patchy encounters, the negative binomial is a good alternative

7.4.3 The probability function

The negative binomial is the source of a great deal of confusion, because it can be derived in different ways and can be expressed using different equations. The classic derivation, and the one that explains its name, is the expected number of failures in a series of discrete trials, before a success happens. It is therefore called a "negative binomial" in a nod to its roots in the binomial distribution, and owing to the fact that it describes the counts of failures (the "negative" part of the name). In ecological circles, the derivation is different. Instead, it starts with the Poisson distribution but assumes that r varies over space and time, with some mean value. The random variable is X, which can take any discrete outcome x. In our example, X refers to the number of bears that are observed, t is the area or amount of time that is surveyed, and r is the rate of observing grizzly bears over some units of time or space. The parameter k has no immediate ecological meaning, but it sets the level of patchiness (variance) of the grizzly bear encounters. Small values of k produce very patchy (high-variance) encounters. From this, we can calculate the probability of all outcomes for $x \geq 0$:

$$P(X=x)=\frac{\Gamma(x+k)}{\Gamma(k)x!}\left(\frac{rt}{k+rt}\right)^{x}\left(\frac{k}{k+rt}\right)^{k} \qquad (7.4)$$

The notation $\Gamma()$ refers to the gamma function (you can think of the gamma function in the same way you think of logarithms—it is a specific mathematical operation, but the specific details of this operation are not critical to know at this point). In Excel, you can get the natural log of the gamma function using the `gammaln()` command, while in R you can get the gamma function calculated at any number by using the `gamma()` command.

7.4.4 Properties of the function

Just like the case for the Poisson distribution, the expected value of X is $\mathrm{E}[X]=rt$. However, the variance of the negative binomial now equals $rt + (rt)^2/k$. In other words, the variance becomes inflated relative to the Poisson by an additional amount equal to $(rt)^2/k$. Small values of k produce high variance.

7.4.5 Example

In our Poisson example, we considered the case of counts of prey encounters from continuous monitoring of emperor penguins. That case assumed that a penguin was equally likely to encounter a prey at any moment in time. There is one huge problem with that assumption: penguins live on ice and dive into water to find fish prey. While they are on ice, they have zero chance of finding food. Although not a perfect example of the negative binomial, it illustrates the idea that the event probability varies over the sampling frame.

When k is large, say, $k = 25$, the expected number of prey attacks looks fairly similar to that in the Poisson model. Compare the left panel of figure 7.3 to the right panel

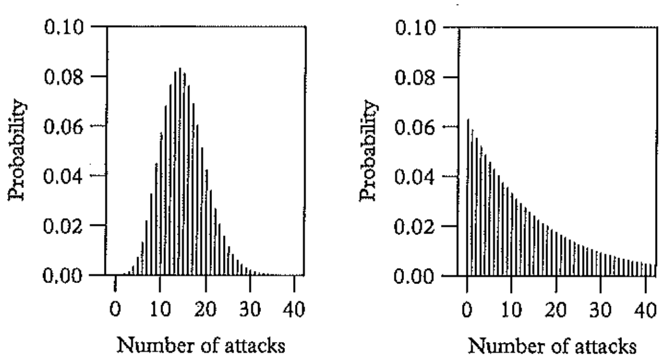

Figure 7.3 *Two examples of the negative binomial probability function, each with the same expected value but with different variances. (Left) k = 25; (right) k = 1. Note that, on the right panel,the right-hand tail of the distribution extends beyond the plot range.*

of figure 7.2. When k is small, say, $k = 1$, the probability density is markedly different (figure 7.3). Notice now that the most common outcome is zero prey encounters. At the same time, it would not be too strange if there were forty or more prey encounters.

7.5 Normal

The normal distribution describes the outcomes of events that are continuous variables. For example, temperature can take any value, not just integer values. The distribution of temperatures is described by any one of several continuous probability density functions. The normal probability density function is a common choice, because it has many desirable features. First, the mean and the variance are defined directly based on its two parameters. Two, it has many helpful statistical properties, such as that given by the central limit theorem, which states that if you have a big enough sample size, the distribution of the mean of the samples is approximately normally distributed.

7.5.1 Key things about this distribution

The outcome is continuous, which means that the probability that a random variable takes any particular value is infinitesimally small. We often bin our observations (e.g., what is the probability that the temperature will be between 12 and 13 degrees) and then integrate to get the probability.

7.5.2 When would I use this?

Whenever you can get away with it. The normal distribution has so many useful features that you might even choose to use it when your observations are discrete (typically, you would do this when the discrete observations are large numbers, say, greater than 100).

7.5.3 The probability density function

Because the normal distribution is continuous, we use a probability density function, which itself has to be integrated to get the probability that a random variable is between two numbers. The normal distribution has two parameters, μ and σ. The probability density is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mu - x)^2}{2\sigma^2}\right) \tag{7.5}$$

where here I adopt the notation $\exp(y)$ to refer to e raised to the power y, which makes the equation easier to read. To get the probability of some set of possible outcomes, k, we integrate over the range of values of k that we are interested in. Assume our range of k is defined by a lower bound, k_{lower} and upper bound, k_{upper}; then

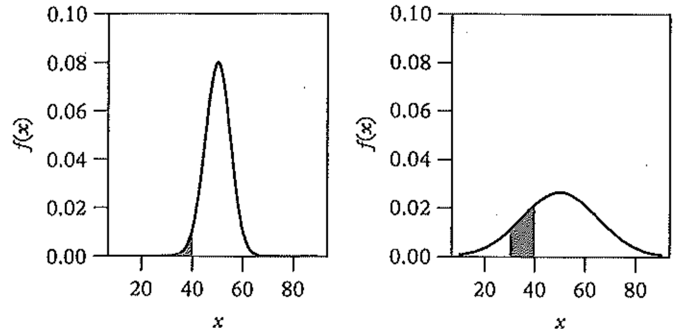

Figure 7.4 *The probability of events are calculated by integrating to get the area under the curve. Here, the probability that the tree diameter is between 30 and 40 cm, given a mean of 50 cm and a standard deviation of 5 (left) and 15 (right).*

$$P(k_{\text{lower}} < x < k_{\text{upper}}) = \int_{k_{\text{lower}}}^{k_{\text{upper}}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mu - x)^2}{2\sigma^2}\right) dx \qquad (7.6)$$

7.5.4 Properties of the function

One reason why the normal distribution is so useful is that the mean and the variance are simply the two parameters of the distribution, that is, $\mathbb{E}[X] = \mu$, $\mathbb{E}[(X - \mu)^2] = \sigma^2$.

7.5.5 Example

Suppose you are measuring tree diameter in a forest. The average tree diameter is 50 cm and has a standard deviation of 5 cm. You want to know the probability that any given tree will have a tree diameter between 30 and 40 cm. If the standard deviation is small, the probability will be relatively low, less than 3% (figure 7.4). If the standard deviation is larger (15), the area under the curve between 30 and 40 is much larger, 16%.

7.6 Log-normal

The log-normal distribution describes the outcomes of events that are continuous variables, where the logarithm of the observations are themselves normally distributed. The log-normal distribution structure is fairly common in ecological data, and there are theoretical reasons for why it arises in nature (Hilborn and Walters 1992). Practically speaking, it is often convenient because the outcome is always a positive number, which means we can use it when we want to ensure that our random variable exceeds 0.

7.6.1 Key things about this distribution

Because the log-normal is the distribution of a variable whose logarithm is normally distributed, the variable cannot be 0 or lower. In other words, the log-normal distribution is only defined for positive values.

7.6.2 When would I use this?

The log-normal is frequently used for continuous random variables that have to be greater than 0.

7.6.3 The probability density function

It is easiest to break the function into two components. The first is the definition of our random variable X:

$$X = e^Y \tag{7.7}$$

where Y is normally distributed with mean μ and variance σ. Given these assumptions, the probability density function of X equals

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(\frac{(\mu - \log(x))^2}{2\sigma^2}\right) \tag{7.8}$$

7.6.4 Properties of the function

You might expect that the mean of the log-normal is simply e^{μ}, but it turns out that the mean of the log-normal is $\mathbb{E}[X] = e^{\mu+0.5\sigma^2}$. This may look vaguely familiar to you (section 5.3). If you want to generate log-normal errors with a specified mean, it's important to correct for the variance. The coefficient of variation (standard deviation divided by the mean) is approximately equal to σ.

7.6.5 Example

We have already seen examples of the log-normal distribution when exploring stochastic density-independent models. Recall that we defined r_t as the log of the λ_t and that r_t were random draws from a normal distribution with mean μ and standard deviation σ. In other words, λ_t equals $\exp(r_t)$. Thus, λ_t, the annual population growth rate, is a log-normal random variable with mean equal to $e^{\mu+\sigma^2/2}$. In addition, the population size at any time t steps into the future is a log-normal random variable with mean equal to $N_0 e^{(\mu+\sigma^2/2t)}$.

7.7 Advanced: Other distributions

7.7.1 The gamma distribution

Like the log-normal distribution, the gamma is only defined for values of X that exceed 0. However, it can take different shapes, which sometimes is useful in fitting models to data (figure 7.5). The probability density function is

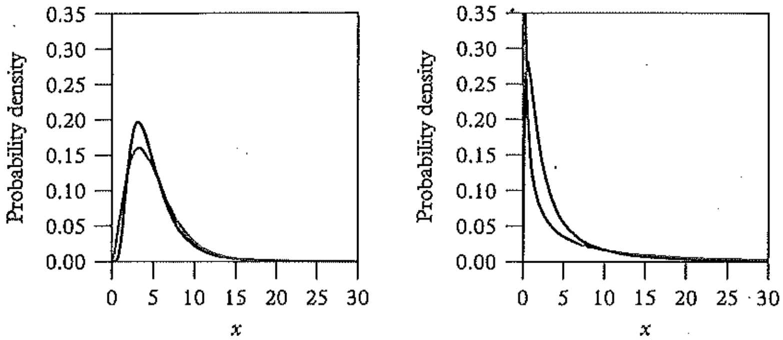

Figure 7.5 *Comparison of the log-normal (black) and the gamma distribution (blue). All distributions have a mean of 5; distributions on the left have a standard deviation of 3, and distributions on the right have a standard deviation of 9.*

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{7.9}$$

The mean is α/β and the variance is α/β^2.

7.7.2 The beta distribution

The beta distribution is a common distribution to use when modeling random variables that must take values between 0 and 1. For instance, in a structured stochastic population model, where you want the survival or transition probabilities to be random draws, the beta distribution is ideal. The probability density function is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (1-x)^{\beta-1} x^{\alpha-1} \tag{7.10}$$

The mean of the beta distribution is $\frac{\alpha}{\alpha+\beta}$. The function is not defined for $x = 0$ or $x = 1$.

7.7.3 Student's t-distribution

Student's t-distribution is an increasingly valued substitute for the normal distribution (Anderson et al. 2017). The normal distribution has one undesirable feature: as you move out on the tails, the probability of events decline exponentially (Taleb 2010). This means that extreme events are quite unlikely. The Student's t-distribution allows for greater probability of events that are far away for the mean, even with the same overall variance as the normal distribution. The probability density function is

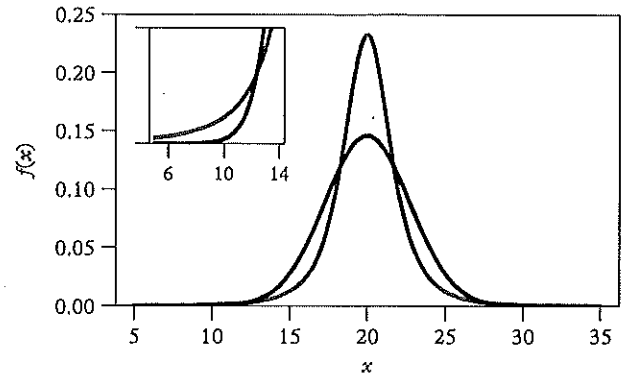

Figure 7.6 *Student's t-distribution (gray) and normal distribution (black). Both distributions have the same mean and variance. The key difference lies in the tails of the distributions. Inset shows the difference in probability density for small values of x.*

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v \sigma^2}} \left(1 + \frac{1}{v}\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{v+1}{2}} \tag{7.11}$$

where v is the degrees of freedom ($v \geq 1$), the mean of the distribution is μ, and the variance is $\frac{v}{v-2}\sigma^2$ (for $v \geq 2$). Note here that the term σ acts as a scaling parameter so that you can make the Student's t-distribution have whatever overall variance you wish. For comparison, consider a normal distribution with mean 20 and variance 7.5. The probability of any event less than 10 is vanishing small (figure 7.6). Yet, a Student's t-distribution, with same mean and variance (here with $v = 3$), can allow events very far from the mean to occur more frequently (figure 7.6).

7.7.4 The beta-binomial distribution

The beta-binomial pairs with the binomial distribution in the same way that the negative binomial pairs with the Poisson distribution. Recall that we use the negative binomial with discrete count data when the data are more dispersed (there is more variability) than the Poisson model can account for. The beta-binomial does the same thing for the binomial distribution. It presumes that the probability of some outcome over some series of discrete trials is not the same over all trials, but instead varies according to a beta distribution. Given this assumption, the probability mass function of the beta-binomial distribution is

$$P(X = x) = \frac{N!}{x!(N-x)!} \cdot \frac{\Gamma(\theta)}{\Gamma(p\theta)\Gamma((1-p)\theta)} \cdot \frac{a\Gamma(x+p\theta)\Gamma(N-x+(1-p)\theta)}{\Gamma(N+\theta)} \tag{7.12}$$

where p is the average probability of the outcome happening in a trial, and θ is the overdispersion parameter that governs the variance in p. The expected value is the same as the binomial (pN), but the variance is given by $Np(1-p)\left(1+\frac{N-1}{\theta+1}\right)$.

7.7.5 Zero-inflated models

Sometimes ecological count data have a lot of zeros, and these cannot be accounted for by our standard probability mass function such as the Poisson or even the negative binomial. In these cases, we might apply a so-called zero-inflated or mixture model. It's called a mixture model because it combines both the binomial model and another model. Usually, we say that the true density is 0 with some probability p and has a nonzero density at probability $1-p$. For instance, a zero-inflated Poisson model of count data is

$$P(X=x)=\begin{cases} p+(1-p)e^{-\lambda} & \text{if} x=0 \\ (1-p)\frac{e^{-\lambda}\lambda^x}{x!} & \text{otherwise} \end{cases} \tag{7.13}$$

where p is the binomial probability per trial, and λ is the mean of the Poisson. Notice in the first line that there are two ways that we could get a 0. One is that the thing we are counting is simply not there with probability p. The second is that the thing we are counting could be present (with probability $1-p$), but we just didn't see them as given by the Poisson (i.e., $e^{-\lambda}$).

Summary

- All data are random variables; that means data can be explained as coming from probability mass functions or probability density functions.
- There are many different probability functions; you choose the function based on some knowledge of the process that generated the random variable.
- You can calculate probabilities of future events easily once you've chosen your probability density function and specified its parameters.
- For continuous probability density functions, you need to integrate the function to get the probability that the random variable takes a value within any specified range.

Exercises

1. You are going to run juvenile salmon mort of salmon that died i twenty salmon. What density function repr
2. If the true mortality r that die in the experi
3. Repeat the calculatic 0.75. Describe and e calculation for the th
4. Geoduck clams are a nearshore habitats an the true density is 0.1 specified number of Why is the Poisso parameters r and t re
5. Given the assumpti probability mass fun area needed so that t of geoducks per squa

Exercises

1. You are going to run an experiment to test the effects of a toxic contaminant on juvenile salmon mortality rates. You will estimate mortality rate as the fraction of salmon that died in the experiment. You choose to run your experiment with twenty salmon. What do the parameters N, p, and k in the binomial probability density function represent in this experiment?
2. If the true mortality rate is 0.5, what is the probability that the fraction of salmon that die in the experiment is within 0.1 unit of the true mortality rate?
3. Repeat the calculation in exercise 2 but with true mortality rates of 0.25 and 0.75. Describe and explain the differences (and similarities) in your probability calculation for the three different mortality rates.
4. Geoduck clams are among the most valuable shellfish in the world. They live in nearshore habitats and can be surveyed by divers monitoring the seabed. Suppose the true density is 0.1 geoducks/m^2. You wish to know the probability of seeing a specified number of geoducks, given some sampling area.

   Why is the Poisson an appropriate probability density function? What do the parameters r and t represent in this situation?
5. Given the assumption that geoduck densities are described with a Poisson probability mass function, with r equal to 0.1/m^2, what is the minimum sample area needed so that there is an 80% or greater chance that the observed number of geoducks per square meter is within 0.05 of the true density?