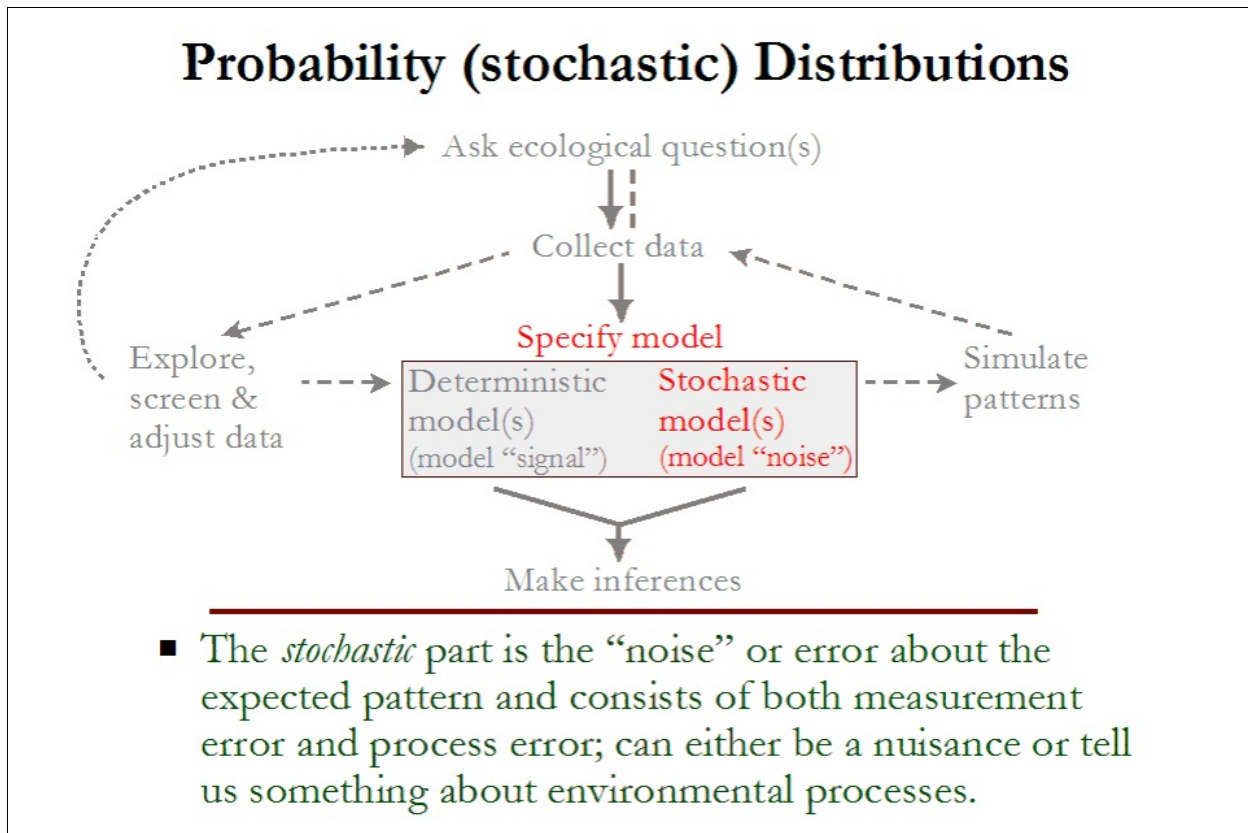


# Analysis of Environmental Data

## Chapter 5: Conceptual Foundations:

### *Probability Distributions*

|   |                    |
|---|--------------------|
| 1. What is a probability (stochastic) distribution? . . . . . | <a href="#">2</a>  |
| 1.1 Measurement error versus process error. . . . .           | <a href="#">3</a>  |
| 1.2 Conventional versus hierarchical models. . . . .          | <a href="#">5</a>  |
| 1.3 Probability distributions. . . . .                        | <a href="#">6</a>  |
| 2. Discrete distributions. . . . .                            | <a href="#">7</a>  |
| 2.1 Probability mass function. . . . .                        | <a href="#">11</a> |
| 2.2 Cumulative probability distribution. . . . .              | <a href="#">12</a> |
| 2.3 Quantile distribution. . . . .                            | <a href="#">13</a> |
| 2.4 Random numbers. . . . .                                   | <a href="#">14</a> |
| 2.5 Binomial examples. . . . .                                | <a href="#">15</a> |
| 3. Continuous probability distributions. . . . .              | <a href="#">16</a> |
| 3.1 Probability density function. . . . .                     | <a href="#">19</a> |
| 3.2 Cumulative probability distribution. . . . .              | <a href="#">20</a> |
| 3.3 Quantile distribution. . . . .                            | <a href="#">21</a> |
| 3.4 Random numbers. . . . .                                   | <a href="#">22</a> |
| 3.5 Normal (Gaussian) examples. . . . .                       | <a href="#">23</a> |
| 4. Bestiary of probability distributions. . . . .             | <a href="#">24</a> |
| 4.1 Discrete distributions. . . . .                           | <a href="#">25</a> |
| 4.2 Continuous distributions. . . . .                         | <a href="#">29</a> |
| 5. Choosing the right probability distribution? . . . . .     | <a href="#">37</a> |
| 6. Statistical Models – putting it all together. . . . .      | <a href="#">38</a> |



### 1. What is a probability (stochastic) distribution?

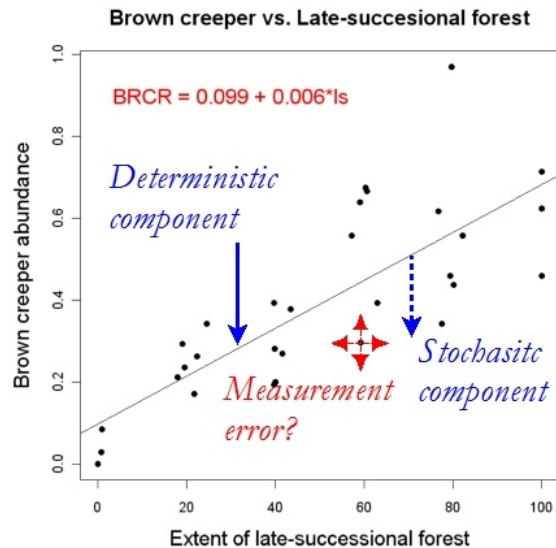
Recall that the deterministic part of the statistical model describes the expected pattern in the absence of any kind of randomness or measurement error. The noise or error about the expected pattern is the *stochastic* component of the model. To formally estimate the parameters of the model, you need to know not just the expected pattern (the deterministic component) but also something about the variation about the expected pattern (the stochastic component). Typically, you describe the stochastic model by specifying a reasonable probability distribution for the variation about the expected pattern (see below).

## Probability Distributions... sources of error

### Measurement, Process, and Model error

Measurement error:

- Variability or “noise” in our measurements having nothing to do with the environmental system
- Makes it hard to make inferences about environmental system
  - ▶ Large confidence intervals
  - ▶ Low statistical power



#### 1.1 Sources of error

The variation about the expected pattern, i.e., the stochastic component of the model, is often termed “noise”. Noise appears in environmental data in two different ways – as measurement error (sometimes called observation error) and as process noise – and it can be exacerbated by model error.

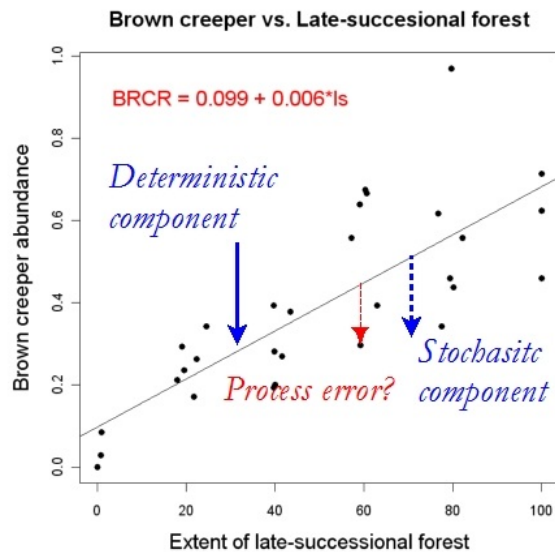
Briefly, *measurement error* is the variability or “noise” in our measurements having nothing to do with the environmental system per se, only in our ability to accurately measure the system. Measurement error makes it hard to estimate parameters and make inferences about environmental systems; more specifically, it leads to large confidence intervals and low statistical power.

## Probability Distributions... sources of error

### Measurement, Process, and Model error

#### Process error:

- Unmeasured but real sources of variability affecting the environmental system
- Makes it hard to make inferences about environmental system
  - ▶ Large confidence intervals
  - ▶ Low statistical power



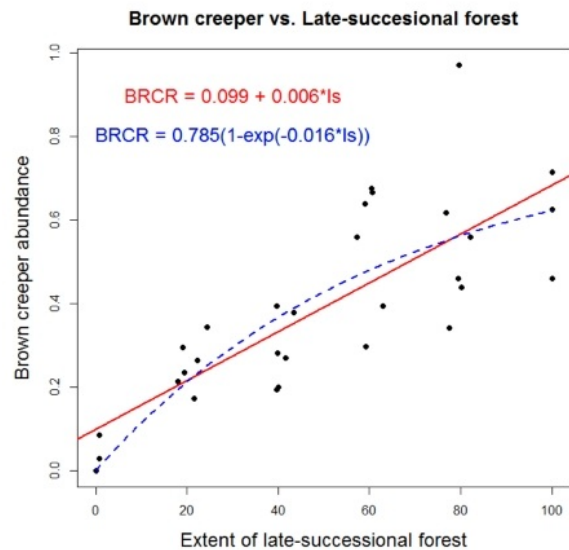
Even if we can eliminate measurement error (or make it trivially small), *process noise* or process error (often so-called even though it isn't technically an error but a real part of the system) still exists because variability affects all environmental systems. For example, as Bolker (2008) describes, we can observe thousands of individuals to determine the average mortality rate with great accuracy. The fate of a group of individuals, however, depends both on the variability in mortality rates of individuals and on the demographic stochasticity that determines whether a particular individual lives or dies. So even though we know the average mortality rate perfectly, our predictions are still uncertain. Additionally, environmental stochasticity – spatial and temporal variability in e.g. mortality rate caused by variation in the environment rather than by the inherent randomness of individual fates – also affects the dynamics, introducing additional variability.

## Probability Distributions... sources of error

### Measurement, Process, and Model error

#### Model error:

- Incomplete, inaccurate, or inappropriate statistical model specified; e.g., wrong error model
- Worsens the fit of the model, decreasing the signal to noise ratio and thus weakening the inferences



Lastly, measurement and process error cooccur with what I will refer to as *model error*, which is the error in the modeled relationship introduced by the choice of a poorly specified or incorrect statistical model. Note, this may be due to the investigator choosing the wrong deterministic function, which exacerbates the apparent error and decreases the signal-to-noise ratio, or choosing the wrong stochastic model (i.e., probability distribution) such that the error is larger than it should or need be. In either case, if the model is improperly specified to reflect the process under investigation, then the signal-to-noise ratio will likely decrease. Even though model error is fundamentally different than error in the process under investigation or its measurement, it nonetheless represents an importance source of uncertainty in the model results. If the model is poorly or incorrectly specified, then there is likely to be much more uncertainty in the model results.

## Probability Distributions... sources of error

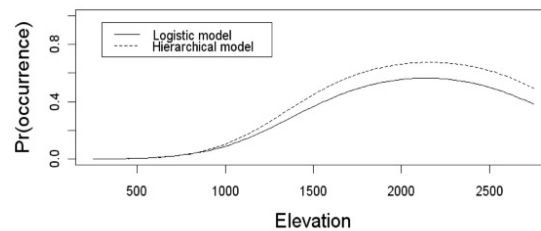
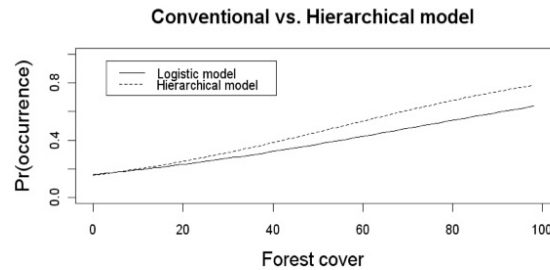
### Conventional versus Hierarchical models

Conventional  
(single-level) model:

- Measurement and process error combined; confounded by design

Hierarchical (multi-level) model:

- Separate observation and process model; requires special study design



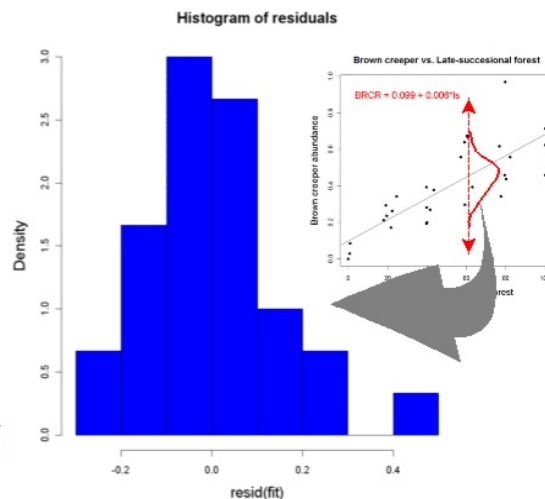
Willow tit example from Royle and Dorazio (2008)

#### 1.2 Conventional versus hierarchical models

The distinction between measurement (or observation) and process error can be very important, and separating these two sources of error is the basis for one whole class of *multi-level* or *hierarchical* models (which we will discuss further in a later section). The example shown here was taken from Royle and Dorazio (2008) and was derived from the Swiss Survey of Common Breeding Birds. The data represent the presence/absence of the Willow Tit on 237 1 km<sup>2</sup> quadrats in relation to two covariates that affect the species' distribution: forest cover and elevation. A conventional (i.e., nonhierarchical) logistic regression model was fit to the data in which the errors associated with the species occurrence on a plot (i.e., process error) and the error associated with detecting the species given it is present (i.e., measurement error) are confounded. A hierarchical model was also fit to the data in which the two sources of error are modeled separately. The figures shown here depict the fitted response in occurrence probability for the willow tit to forest cover and elevation. The solid line and dotted lines show the fitted responses for the conventional and hierarchical models, respectively. Even though the general relationships are the same between models, the response to both covariates is significantly stronger in the hierarchical model. In extreme cases, failure to separately account for both observation and process error can lead to erroneous conclusions.

## Probability Distributions...

- Describes the set of probabilities for all possible *outcomes* of a random variable; based on the *dependent variable* type
- Represents the *stochastic* component of the model; distribution of the error
- Required for all *parametric* statistical models
- Think of it as a *data-generating mechanism* (useful for stochastic simulation)



### 1.3 Probability distributions

Formally, a *probability distribution* is the set of probabilities on a sample space or set of outcomes; it describes the set of probabilities for all possible outcomes of a random variable. Here, we will always be working with sample spaces that are numbers – the number or amount observed in some measurement of an environmental system. Importantly, the probability distribution deals with the type (scale) of the dependent variable, not the independent variable(s). In the context of a statistical model, it is perhaps easiest to think about a probability distribution as a description of the stochastic component of the model, representing the random variability about the expected value (typically the mean). In the simple linear regression example shown here, where the expected value of brown creeper abundance is a linear function of the extent of late-successional forest, the observed values of brown creeper abundance deviate from the expected (fitted) values in every case. These differences are the “errors” in the model (these errors are also called “residuals” as they represent the residual variability not accounted for by the deterministic model). The probability distribution describes the distribution of these errors, and is shown here as a histogram. Note, to fit this model using parametric methods, we had to assume that these errors were drawn from a specific distribution (the normal distribution in this case) having one or more parameters that control its shape, location and/or scale. The histogram reveals that this was a reasonable assumption in this case. Note that the assumption of a particular probability distribution is the basis for all so-called *parametric* statistical methods. Lastly, it can be useful to think of the probability distribution as a sort of *data-generating mechanism*; an engine that randomly generates the data points that we ultimately observe in our sample. This random data generating mechanism is necessary to simulate data, which we will discuss in a later section.

## Probability Distributions... discrete

Definition:

- Outcomes are a set of *discrete* values (integers)
- Most of the discrete distributions we deal with describe counting or sampling processes and have ranges that include some or all of the nonnegative integers



What are some examples?

### 2. Discrete distributions

The simplest distributions to understand are *discrete* distributions whose outcomes are a set of integers; most of the discrete distributions we deal with describe counting or sampling processes and have ranges that include some or all of the nonnegative integers. Note, discrete data generally arise from counting indivisible units. For example, a count of the number of individuals on a plot is discrete since individuals are indivisible – at least if they want to remain an individual. However, counting the basal area of trees on a plot (e.g.,  $\text{m}^2/\text{ha}$ ) but recording it to the nearest integer value doesn't mean that basal area is discrete, only that your measurement precision is to the nearest integer. Recording a continuous variable on a discrete scale doesn't make the variable truly discrete, even if it allows you to treat the data as if it were discrete – and thus use a discrete probability distribution. However, there are occasional cases when it may make sense or even be necessary to make your continuous data discrete for the analysis or, conversely, treat count data as if it were continuous, but in general it is probably best to remain faithful to the intrinsic scale of the variable.



## Probability Distributions... discrete

### Example: *Binomial distribution*

- It applies when you have samples with a fixed number of subsamples or “trials” in each one, and each trial can have one or two values (e.g., present/absent, alive/dead), and the probability of “success” (present, alive) is the same in every trial

#### Example:

Size(#trials) = 10  
 prob(present) = 0.3

|         |   |   |   |   |   |   |   |   |   |    |
|---------|---|---|---|---|---|---|---|---|---|----|
| Trial   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Outcome | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0  |

Total # successes ( $k$ ) = 4

---

What is the probability of observing  $k = 4$  successes?

As an example, one of the most common discrete distributions in environmental science and the easiest to understand is the *binomial*. It applies when you have samples (observational units) with a fixed number of subsamples or “trials” in each one, and each trial can have one of two values (e.g., present/absent, alive/dead), and the probability of “success” (present, alive) is the same in every trial. Consider the presence/absence of a focal species across 10 plots (subsamples) within a sample (observational) unit, say a watershed. At each plot, the species is either “present” or “absent”, which we usually represent as a 1 and a 0, respectively. If we survey all 10 plots (trial  $size=10$ ) within each sample and the probability of the species being present is  $prob=0.3$ , then the probability of observing the species present at say 4 plots ( $k=4$ ) within a sample will have a binomial distribution with parameters  $size=10$  and  $prob=0.3$ . What type of data does this example represent?

Hopefully you answered *proportional*, since each sample (observational) unit has a fixed number of trials, each of which can take on a binary response, such that the number of successes can be expressed as a proportional of the total number of trials (e.g., 3 out 10). What if each sample unit had only a single trial (i.e.,  $size=1$ ). What kind of data would this be? Hopefully you answered *binary*, since each sample unit can take on only one of two values (e.g., present or absent, success or failure). The key here is to note that binary data is just the special case of proportional data when the trial size equals 1. As a result, the binomial distribution is appropriate for both data types.

## Probability Distributions... discrete

Example: *Binomial distribution*

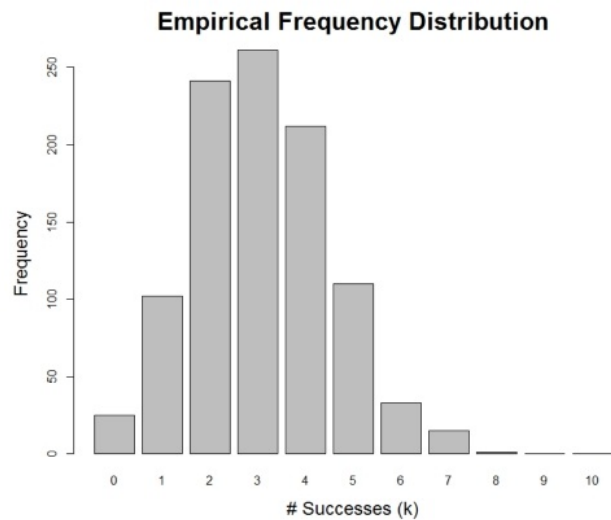
Example:

Size(#trials) = 10

prob(present) = 0.3

| Sample   | Trial outcome       | $k$ |
|----------|---------------------|-----|
| Sample 1 | 0 1 0 0 0 1 0 1 1 0 | 4   |
| Sample 2 | 0 0 0 0 0 0 1 0 0 0 | 1   |
| Sample 3 | 0 1 1 0 0 0 0 0 1 0 | 3   |
| etc...   |                     |     |

Aside: What is the most likely outcome? And does it depend on sample size?



The easiest way to think about this is as follows. For a single sample consisting of 10 plots, we will observe the species to be present a certain number of times, right? The range of possibilities is 0-10, since there are 10 plots – this constitutes the full set of possible outcomes. If the probability of the species being present on any single plot is  $prob=0.3$ , then on average we should expect to observe the species on 3/10 plots. However, given random variability in the environmental system (and the error in our ability to accurately detect the species), it is quite possible that for any given random sample of 10 plots, we might easily observe the species present on say only 1 or 2 plots or perhaps 4 or 5 plots, but it seems highly unlikely that we would observe the species present on all 10 plots. Numerically, we can figure out the likelihood of the species being present 0, 1, ... 10 times by drawing repeated samples, each time recording the number of presences across the 10 plots ( $k$ ). After we draw a large number of samples, we can plot the distribution of  $k$ .

## Probability Distributions... discrete

Example: *Binomial distribution*

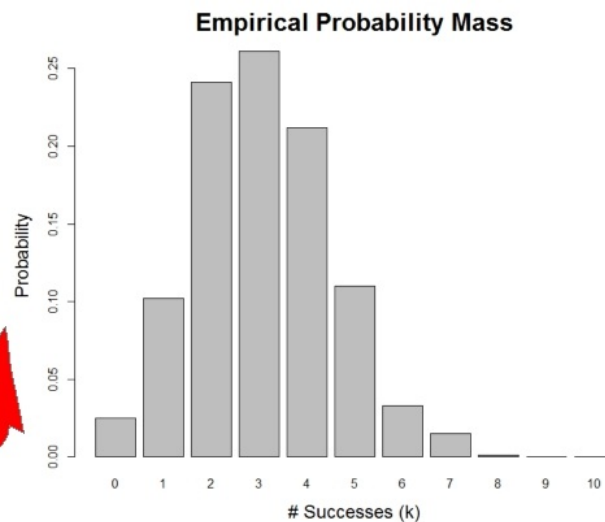
Example:

Size(#trials) = 10

prob(present) = 0.3

| Sample   | Trial outcome       | $k$ |
|----------|---------------------|-----|
| Sample 1 | 0 1 0 0 0 1 0 1 1 0 | 4   |
| Sample 2 | 0 0 0 0 0 0 1 0 0 0 | 1   |
| Sample 3 | 0 1 1 0 0 0 0 0 1 0 | 3   |
| etc...   |                     |     |

Note, divide frequencies  
by total frequency to  
convert to a probability



The frequency of samples with any particular value of  $k$  can be converted into a probability simply by dividing it by the total number of samples in the distribution (or the cumulative area of the frequency distribution). If we plot the result, we have the empirical probability distribution (given a special name below) which depicts the probability of observing any particular value of  $k$ . Of course, the empirical probability distribution shown here is specific to this data set; it gives the actual probability that we observed each value of  $k$  (# successes) in this dataset. However, we typically want to know the theoretical probability distribution; the probability of observing each value of  $k$  from an infinitely large sample, under the assumption that our sample was drawn from such a distribution. Of course we don't have to actually collect hundreds or thousands of samples in order to calculate the approximate theoretical probability distribution, since the mathematical properties of the binomial distribution (and many others, see below) are known.

## Probability Distributions... discrete

Example: *Binomial distribution*

Probability mass function (pmf):

$$f(x) = \text{Prob}(X=x)$$

Binomial pmf:

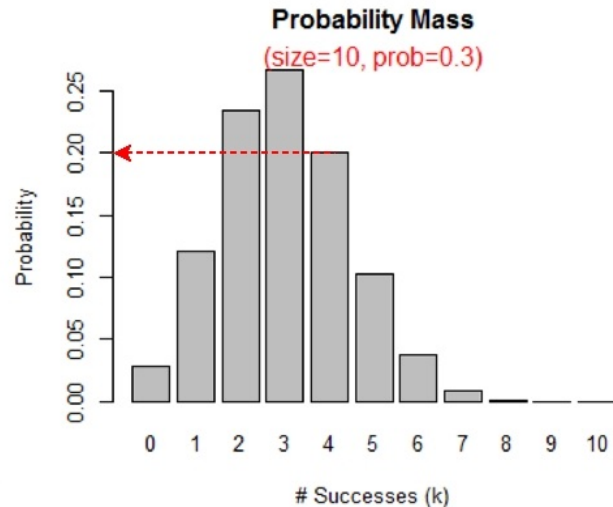
$$\binom{N}{x} p^x (1-p)^{N-x}$$

$N$  = trial size

$p$  = per trial prob(success)

$x$  = #successes ( $k$ )

$$\text{dbinom}(x=4, \text{size}=10, \text{prob}=0.3) \\ = 0.2$$



### 2.1 Probability mass function

For a discrete distribution, the probability distribution (often referred to as the *probability mass function* or *pmf* for short) represents the probability that the outcome of an experiment or observation (called a random variable)  $X$  is equal to a particular value  $x$  ( $f(x) = \text{Prob}(X=x)$ ). In the case of the *binomial distribution*, for example, we can calculate the probability of any specified number of successes ( $x=k$ ) given the number of trials (*size* or  $N$ ) and the probability of success (*prob* or just  $p$ ) for each trial. In the example shown here, the probability of the species being present at 4 of the 10 plots given a per trial probability of success  $\text{prob}=0.3$  is 0.2.

## Probability Distributions... discrete

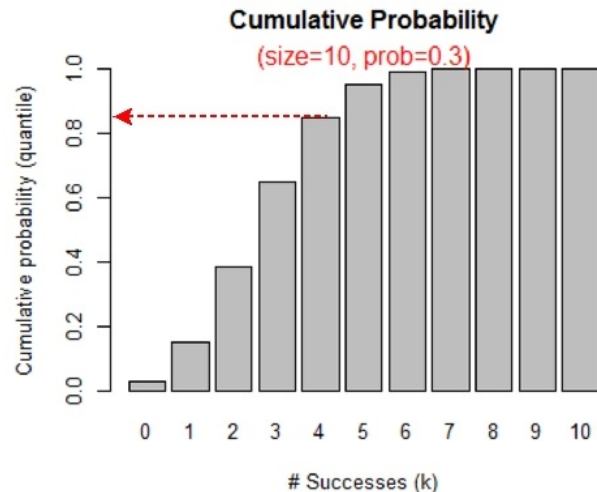
Example: *Binomial distribution*

Cumulative probability distribution:

$$f(x) = \text{Prob}(X \leq x)$$

- Denotes probability of  $x$  being less than or equal to any particular value (basis for  $p$ -values)

$$\text{pbinom}(x=4, \text{size}=10, \text{prob}=0.3) = 0.85$$



### 2.2 Cumulative probability distribution

The cumulative probability distribution represents the probability that a random variable  $X$  is less than or equal to a particular value of  $x$  ( $f(x) = \text{Prob}(X \leq x)$ ). Cumulative distribution functions are most useful for frequentist calculations of so-called tail probabilities, or  $p$ -values. In the case of the binomial distribution, for example, we can calculate the probability of the species being present  $k$  or fewer times given the number of trials ( $size$ ) and the probability of success ( $prob$ ) for each trial. In the example shown here, the probability of the species being present at 4 or fewer plots out of 10 given a per trial probability of success  $prob=0.3$  is 0.85. Thus, if we were to repeatedly draw samples of 10 (trials) given a per trial probability of success  $prob=0.3$ , then we would expect to observe the species present on 4 or fewer plots 85% of the time. Conversely, then, we would expect to observe the species present on 5 or more plots 15% of the time. Thus, the probability of observing the species present on 5 or more plots if the sample was drawn from a binomial distribution with  $size=10$  and  $prob=0.3$  would be 0.15, or  $p=0.15$ . We refer to this as the “tail” probability because it represents the probability of being in the tail of the probability distribution; the probability of being 5 or greater in this case. We also call this the  $p$ -value, which is the standard basis for hypothesis testing, which we will return to later.

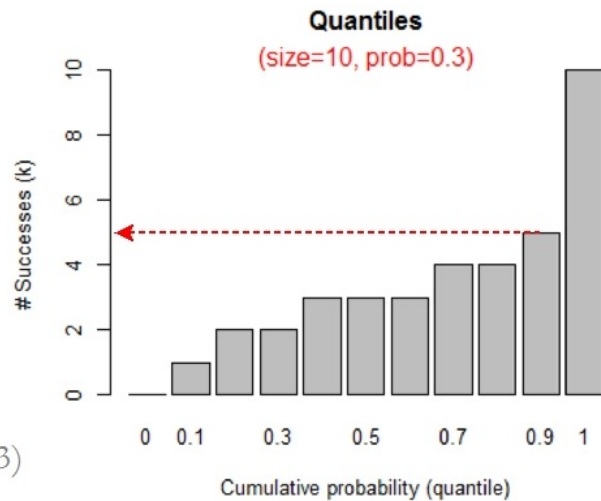
## Probability Distributions... discrete

Example: *Binomial distribution*

Quantile distribution:

- Denotes value of  $x$  for any given quantile of the cumulative probability distribution; i.e., it is the opposite of the cumulative probability distribution

```
qbinom(p=.9,size=10,prob=0.3)
= 5
```



### 2.3 Quantile distribution

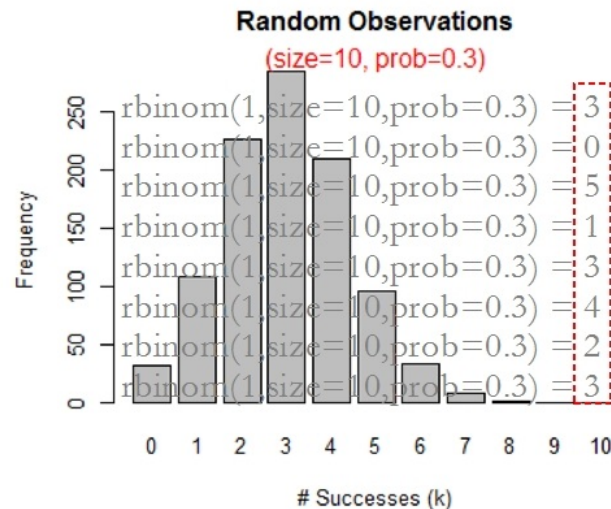
The quantile distribution represents the value of  $x$  for any given quantile (proportion) of the cumulative probability distribution; i.e., it is the opposite of the cumulative probability distribution. In the case of the binomial distribution, for example, we can calculate the number of times the species is expected to be present  $k$  or fewer times for any specified probability level given the number of trials ( $size$ ) and the per trial probability of success ( $prob$ ). In the example shown here, the species is expected to present on 5 or fewer plots out of 10 in 90 out of every 100 samples given a per trial probability of success  $prob=0.3$ .

## Probability Distributions... discrete

Example: *Binomial distribution*

Random numbers:

- Random draws from the specified distribution
- Each draw will produce a different result, which gives rise to the probability distribution
- Extremely useful for simulations



### 2.4 Random numbers

Random numbers can be drawn from a specified probability distribution. For example, in the binomial example shown here, we can draw a random sample from the specified binomial distribution (i.e., given trial size and per trial probability of success). The result is the number of successes (or species' presences, in this case) in a single sample of  $size=10$  given a per trial probability of success  $prob=0.3$ . Note, if we repeat this process, we would expect to observe a different result each time owing to chance. Indeed, it is this variability that gives rise to the probability distribution.

## Probability Distributions... discrete

Example: *Binomial distribution*

Real-world examples:

- # adult female salamanders that successfully breed in a pond (observational unit)
- # point intercepts intersecting focal plant species out of # point intercepts for each vegetation plot



What are some other examples of proportional or binary data where the binomial distribution makes sense?

### 2.5 Binomial examples

There are lots of examples of data well suited to the binomial distribution:

- # adult female salamanders that successfully breed in a pond (observational unit). In this example, the pond is the sample (observational) unit, the # females attempting to breed is the trial size, and the # successful females is the outcome of the set of trials (breeding attempts). The proportion of breeding females that are successful is distributed binomially. Note, in this case, if multiple ponds are sampled, the trial size is likely to vary among ponds – this is OK with proportional data.
- # point intercepts along a transect or grid that intersect a focal plant species out of a total # of point intercepts taken on a vegetation plot. In this example, the vegetation plot is the sample (observational) unit, the # of point intercepts taken is the trial size, and the # of point intercepts intersecting the focal plant species is the outcome of the set of trials (point intercepts). The proportional cover of the focal plant species is distributed binomially. Note, in this case, in multiple vegetation plots are sampled, the trial size is likely to be fixed by study design.
- Can you think of other examples of either proportional data or binary data where the trial size is 1?



## Probability Distributions... continuous

Definition:

- Outcomes exist over a *continuous* range (real numbers)
- The most common continuous distribution in environmental science and all of statistics is the “normal” distribution



What are some examples?

### 3. Continuous probability distributions

A probability distribution over a continuous range (such as all real numbers, or the nonnegative real numbers) is called a *continuous* distribution.

As an example, the most common continuous distribution in ecology (and all of statistics) and the easiest to understand is the *normal*.

The normal distribution applies when you have samples that can take on any real number and for which there is a central tendency (usually described by the mean). Consider the size of fish in a particular stream. Each fish has a size that can take on any nonnegative real number ( $>0$ ). If the size distribution of fish in the population is symmetric about the mean ( $\mu$ ) with a standard deviation ( $sd$ ) or spread about the mean equal to  $\sigma$ , then the probability of observing a fish of size  $y$  will have a normal distribution with parameters  $\mu$  and  $\sigma$ .

### Probability Distributions... continuous

Example: *Normal distribution*

- It applies when you have a random variable that can take on any real number ( $-\infty$  to  $+\infty$ ) and the distribution is unimodal and symmetric with mean  $\mu$  and standard deviation  $\sigma$

Example:

Mean = 10  
sd = 2

Fish 1 2 3 4 5 ...  
size 7.1 4.5 12.2 13.1 10.0 ...

---

What is the probability of observing size = 8?

### Probability Distributions... continuous

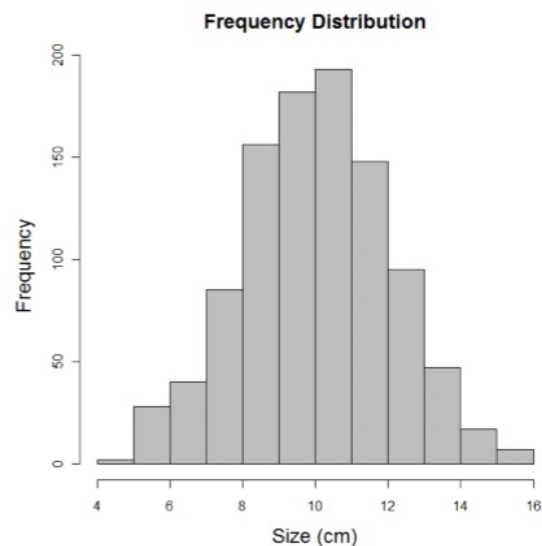
Example: *Normal distribution*

Example:

Mean = 10  
sd = 2

Fish 1 2 3 4 5 ...  
size 7.1 4.5 12.2 13.1 10.0 ...

Aside: What is the most likely outcome? And does it depend on sample size?



## Probability Distributions... continuous

Example: *Normal distribution*

Example:

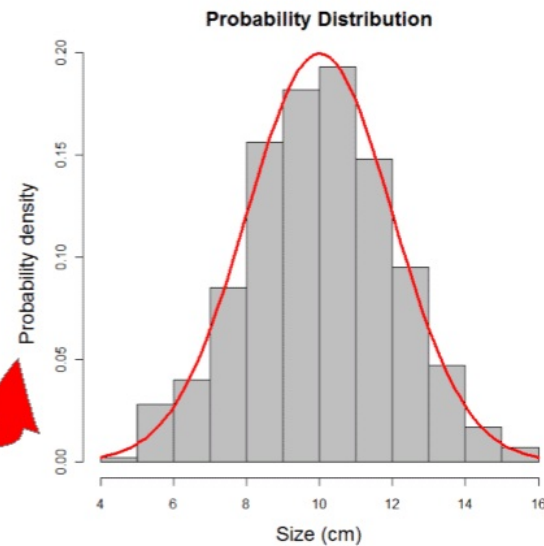
Mean = 10

sd = 2

Fish 1 2 3 4 5 ....

size 7.1 4.5 12.2 13.1 10.0 ....

Note, can't just divide frequencies by cumulative frequencies to convert to a probability with continuous distributions



Continuous probability distributions are a bit more confusing to understand than discrete distributions. Let's consider the fish size example further. The frequency of fish of any particular size (i.e., the frequency distribution, as shown) can't be converted into a probability simply by dividing it by the total number of samples in the distribution (or the cumulative area of the frequency distribution) as we did with the discrete distribution. This is because, unlike the discrete distribution, we can't enumerate all possible outcomes of the continuous random variable because it can take on any real number – at least theoretically. In the fish size example, you may imagine that a measurement of size is *exactly* 7.9 cm, but in fact what you have observed is that the size is between 7.85 and 7.95 cm – if your scale has a precision of  $\pm 0.1$  cm. Thus, continuous probability distributions are expressed as *probability densities* rather than probabilities – the probability that random variable  $X$  is between  $x$  and  $x+\Delta x$ , divided by  $\Delta x$  ( $\text{Prob}(7.85 < X < 7.95)/0.1$ , in this case). Dividing by  $\Delta x$  allows the observed probability density to have a well-defined limit as precision increases and  $\Delta x$  shrinks to zero. Unlike probabilities, probability densities can be larger than 1. In practice, we are concerned mostly with relative probabilities or likelihoods, and so the maximum density values and whether they are greater than or less than 1 don't matter much.

## Probability Distributions... continuous

### Example: *Normal distribution*

Probability density function (pdf):

$$f(x) = \frac{\text{prob}(x \leq X \leq (x + \Delta x))}{\Delta x}$$

Normal pdf:

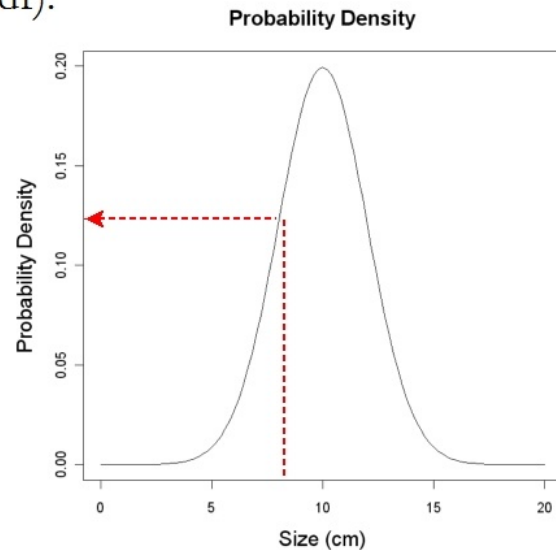
$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

$\mu$  = mean

$\sigma$  = standard deviation

$x$  = value of random variable

$$\text{dnorm}(x=8, \text{mean}=10, \text{sd}=2) \\ = 0.12$$



### 3.1 Probability density function

For a continuous distribution, the probability distribution is referred to as a *probability density function* (or *pdf* for short). In the case of the *normal distribution*, for example, we can calculate the probability density for any specified value of  $X$  given the mean ( $\mu$ ) and standard deviation ( $\sigma$ ). In the example shown here, the probability density for an individual with size = 8 cm given an underlying mean of say 10 cm and a standard deviation of say 2 cm is 0.12.

## Probability Distributions... continuous

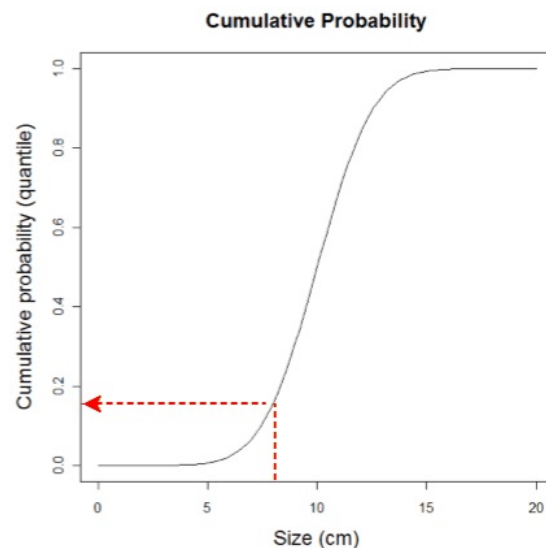
Example: *Normal distribution*

Cumulative Probability  
Distribution:

$$f(x) = \text{Prob}(X \leq x)$$

- Denotes probability of  $x$  being less than or equal to any particular value (basis for  $p$ -values)

$$\text{pnorm}(x=8, \text{mean}=10, \text{sd}=2) = 0.16$$



### 3.2 Cumulative probability distribution

The cumulative probability distribution represents the probability that a random variable  $X$  is less than or equal to a particular value of  $x$  ( $f(x) = \text{Prob}(X \leq x)$ ). Remember, cumulative distribution functions are most useful for frequentist calculations of tail probabilities. In the case of the normal distribution, for example, we can calculate the probability of a fish being smaller than or equal to any given size given the mean fish size ( $\mu$ ) and the standard deviation ( $\sigma$ ). In the example shown here, the probability of a fish less than or equal to 8 cm given an underlying *mean* of 10 cm and *sd* of 2 cm is 0.16.

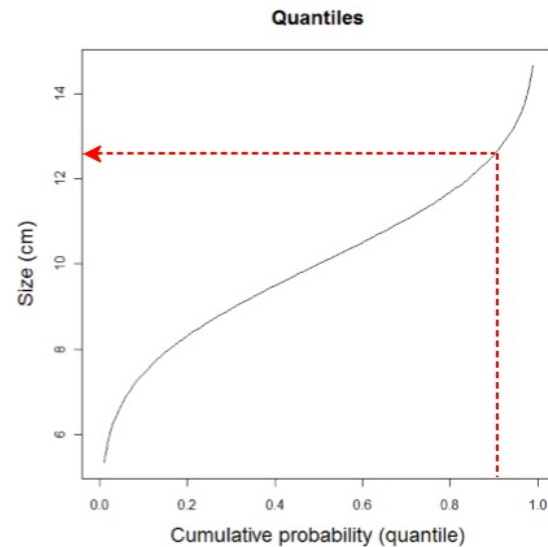
## Probability Distributions... continuous

Example: *Normal distribution*

Quantile Distribution:

- Denotes value of  $x$  for any given quantile of the cumulative probability distribution; i.e., it is the opposite of the cumulative probability distribution

```
qnorm(p=.9,mean=10,sd=2)
= 12.56
```



### 3.3 Quantile distribution

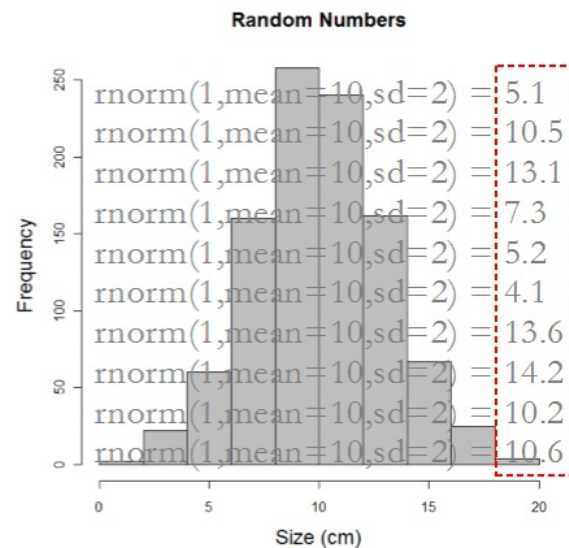
The quantile distribution represents the value of  $x$  for any given quantile (proportion) of the cumulative probability distribution; i.e., it is the opposite of the cumulative probability distribution. In the case of the normal distribution, for example, we can calculate the upper size limit below which any specified percentage of the distribution lies given the mean fish size ( $\mu$ ) and the standard deviation ( $\sigma$ ). In the example shown here, fish are expected to be smaller than 12.56 cm 90 percent of the time given an underlying *mean* of 10 cm and *sd* of 2 cm.

## Probability Distributions... continuous

Example: *Normal distribution*

Random numbers:

- Random draws from the specified distribution
- Each draw will produce a different result, which gives rise to the probability distribution
- Extremely useful for simulations



### 3.4 Random numbers

Random numbers can be drawn from a specified probability distribution. For example, in the normal distribution example shown here, we can draw a random sample from the specified normal distribution (i.e., given the mean and sd). The result is the size (in this case) of a single fish drawn at random from a normal distribution with *mean* of 10 cm and *sd* of 2 cm. Note, if we repeat this process, we would expect to observe a different result each time owing to chance. Indeed, it is this variability that gives rise to the probability distribution.

## Probability Distributions... continuous

Example: *Normal distribution*

Real-world examples:

- Home range size (ha) of individuals in a population
- Basal area ( $\text{m}^2/\text{ha}$ ) of white pine stands
- Temperature (degrees celsius) on December 21<sup>st</sup> over the past 100 years



What are some other examples of continuous data where the normal distribution makes sense?

### 3.5 Normal (Gaussian) examples

There are lots of examples of data well suited to the normal distribution:

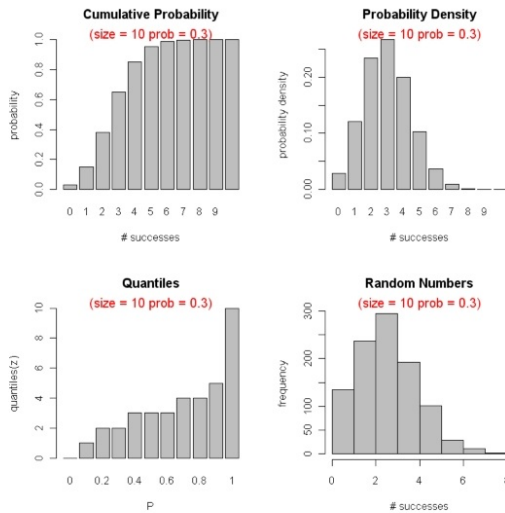
- Home range size (ha) of individuals in a population. The individual is the sample (observational) unit and home range size is the measured continuous response that is likely to roughly symmetrically distributed about a central mean value.
- Basal area ( $\text{m}^2/\text{ha}$ ) of white pine stands. The forest stand is the sample (observational) unit and basal area is the measured continuous response that is likely to be roughly symmetrically distributed about a central mean value.
- Temperature (degrees Celsius) on December 21<sup>st</sup> over the past 100 years. The year or day of December 21 is the sample (observational) unit and temperature is the measured continuous response that is likely to roughly symmetrically distributed about a central mean value.
- Can you think of other examples of continuous data where the normal distribution makes sense?



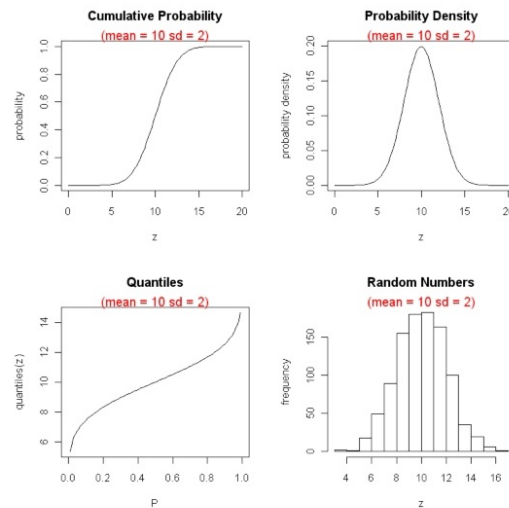
# Probability Distributions... review

## Discrete versus Continuous

### Discrete distributions (binomial)



### Continuous distributions (normal)



## 4. Bestiary of probability distributions

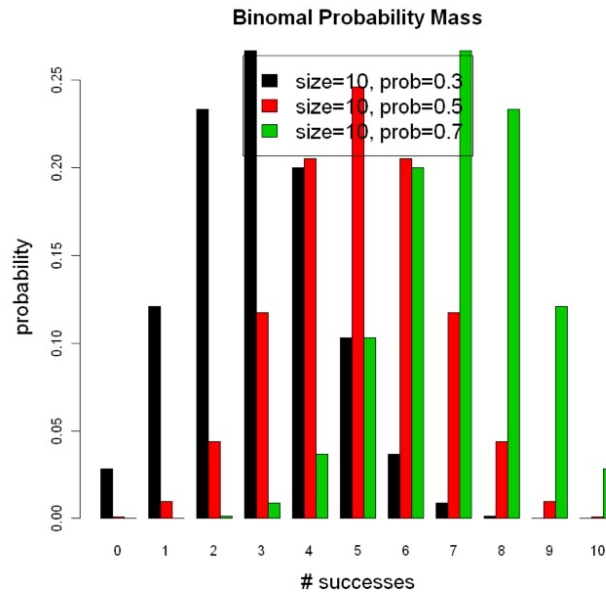
There are a large number of well-established probability distributions. Some of these, like the binomial and normal distributions discussed above, are already quite familiar to ecologists. Indeed, the normal distribution should be familiar to everyone as it is the distribution underlying most of the classical statistical methods that we learned in introductory statistics courses. However, it is very important to understand that there are many other probability distributions for both discrete and continuous data that have widespread application to environmental data. Indeed, some of these distributions address the shortcomings or limitations of the normal distribution for environmental data. Of course, a detailed description of all of the available distributions is way beyond the scope of this course. Instead, here we will briefly review a small bestiary of distributions that are useful in environmental modeling (summarized from Bolker 2008 and Crawley 2007, but do see these references for a more complete description).

# Probability Distributions... bestiary of functions

## Discrete Distributions

### Binomial Distribution:

- Gives number of successes ( $k$ ) in a sample with fixed number of trails ( $size$ ) and equal probability of success ( $prob$ ) in every trial
- Used when  $k$  has an upper limit; when  $size$  is large and  $prob$  is small, approaches Poisson
- Lots of examples (very common)



## 4.1 Discrete distributions

### 4.1.1 Binomial

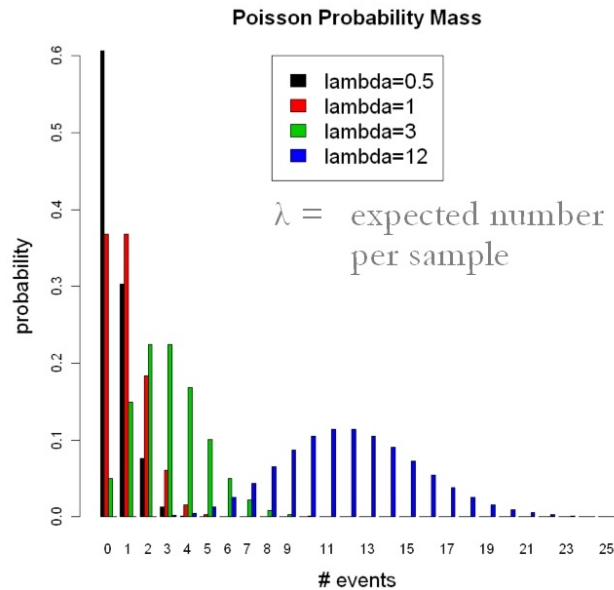
The binomial distribution gives the number of success ( $k$ ) in a sample with a fixed number of trials ( $size$ ) and an equal probability of success ( $prob$ ) in every trial. In each trial, there are only two possible outcomes (success or failure, presence or absence, etc.). Thus, the response in a given trial is binary and can be represented as a 0,1 response. The binomial is used when  $k$  has an upper limit; that is, when the trial size is known and places a limit on the number of possible successes. When the trial size is effectively unlimited, the Poisson distribution is more appropriate. In addition, when the sample size (number of trials) gets large and the probability of success per trial ( $prob$ ) is small, the binomial distribution approaches the Poisson (below). There are numerous uses of the binomial in ecology.

# Probability Distributions... bestiary of functions

## Discrete Distributions

### Poisson Distribution:

- Gives number of events in a given unit of sampling effort if each event is independent
- Used when you expect the number of events to be effectively unlimited
- Used only for count data (very common)



#### 4.1.2 Poisson

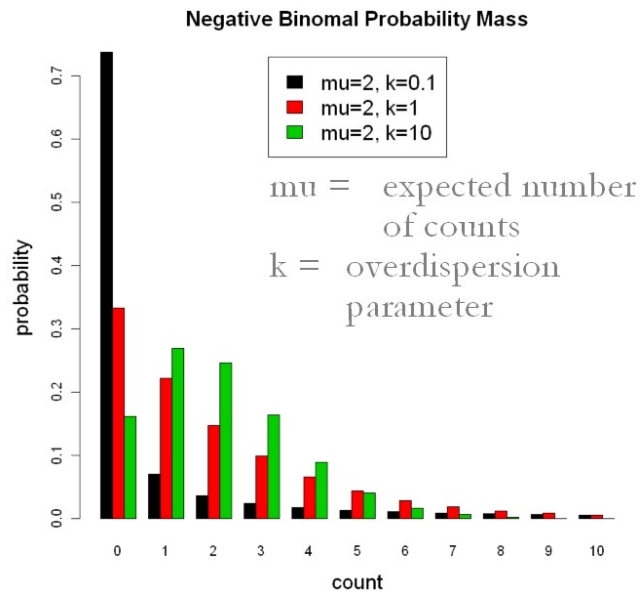
The Poisson distribution gives the number of events in a given unit (space or time) of sampling effort if each event is independent. The Poisson is used when you expect the number of events to be effectively unlimited; if the number of events is limited, use the binomial distribution. The Poisson is used only for count data; that is, events that can't be meaningfully divided into smaller units simply by changing the scale of the measurement (e.g., changing meters to centimeters). The Poisson is very common and familiar to ecologists because count data is so common to population and community ecology.

## Probability Distributions... bestiary of functions

### Discrete Distributions

Negative binomial  
Distribution:

- Gives number of failures before a predetermined number of successes ( $n$ ) occurs in a sample with equal probability of success in every trial ( $prob$ )
- Typically used as an overdispersed Poisson (variance > mean)
- Used as with Poisson, but with heterogeneity in data (overdispersion)



#### 4.1.3 Negative binomial

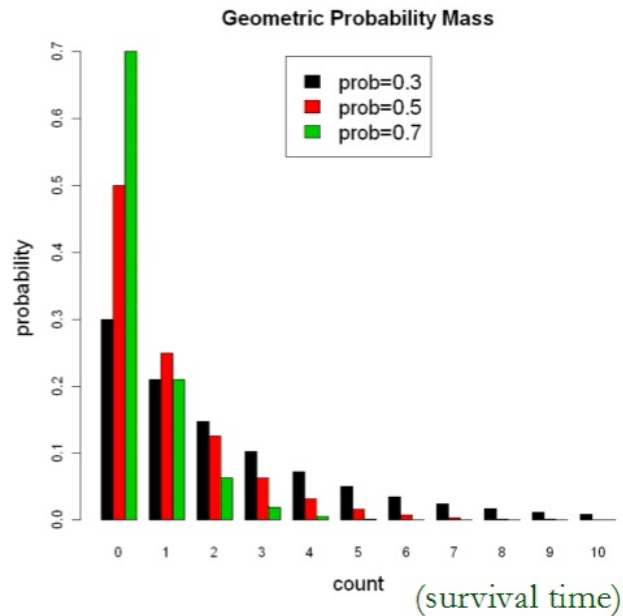
The negative binomial distribution gives the number of failures before a predetermined number of success ( $n$ ) occurs in a sample with an equal probability of success ( $prob$ ) in every trial. While the negative binomial has a mechanistic use in ecology, it is more often used phenomenologically to describe a patchy or clustered distribution with no intrinsic upper limit that has more variance than the Poisson; i.e., an overdispersed Poisson distribution in which the variance is greater than the mean. In this case, the negative binomial can be parameterized with two parameters,  $\mu$  and  $k$ , where  $\mu$  is the expected number of counts and  $k$  is the overdispersion parameter. The overdispersion parameter measures the amount of clustering, or aggregation, or heterogeneity in the data: a smaller  $k$  means more heterogeneity. The negative binomial can be hard to tell from the Poisson when the heterogeneity is low. The negative binomial can arise from a Poisson sampling process where the rate ( $\lambda$ ) itself varies according a Gamma distribution (below). In this case, the negative binomial reflects unmeasured (“random”) variability in the population. Negative binomial distributions can also result from a homogeneous birth-death process, making it difficult to reason backwards from the observed probability distribution to the particular mechanisms that generated it. Nevertheless, the negative binomial is very useful in ecology and is most often used in the same situations as the Poisson, but allowing for heterogeneity in the data.

## Probability Distributions... bestiary of functions

### Discrete Distributions

#### Geometric Distribution:

- Gives number of trials until you get a single failure in a sample with equal probability of failure (*prob*) in every trial; alternative form gives number of trials until first success given equal per trial probability of success (*prob*)
- Special case of *nbinom* when  $k=1$
- Used for survival data



#### 4.1.4 Geometric

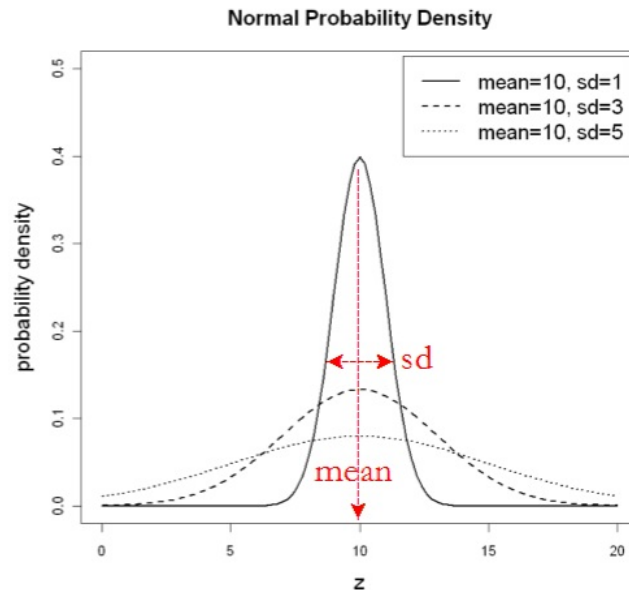
The geometric distribution has two closely related forms. One form gives the number of trials based on the set  $\{1, 2, \dots\}$  until you get a single success in a sample with equal probability of success (*prob*) in every trial. The alternative form gives the number of failures based on the set  $\{0, 1, 2, \dots\}$  until you get a single success in a sample with an equal probability of success (*prob*) in every trial. Note the subtle differences between formulations. The probability mass distribution, along with the mean and variance, are slightly different between forms. Note, the geometric is also a special case of the negative binomial, with  $k=1$ . The geometric distribution is often used to describe survival time data, e.g., the number of successful/survived breeding seasons (i.e., interpreted as failures) for a seasonally reproducing organism until it dies (i.e., interpreted as success), when time is considered as discrete units (e.g., years). Alternatively, and more directly, it can be used to describe for example the number of failed nesting attempts until the first success.

# Probability Distributions... bestiary of functions

## Continuous Distributions

### Normal Distribution:

- Distribution of the *sum* of many independent samples from the same distribution; ubiquity lies in the Central Limit Theorem
- Mean and variance are independent, so used when variance is constant
- Used with continuous, unimodal and symmetric distributions (everywhere), and the basis for most classical methods



## 4.2 Continuous distributions

### 4.2.1 Normal

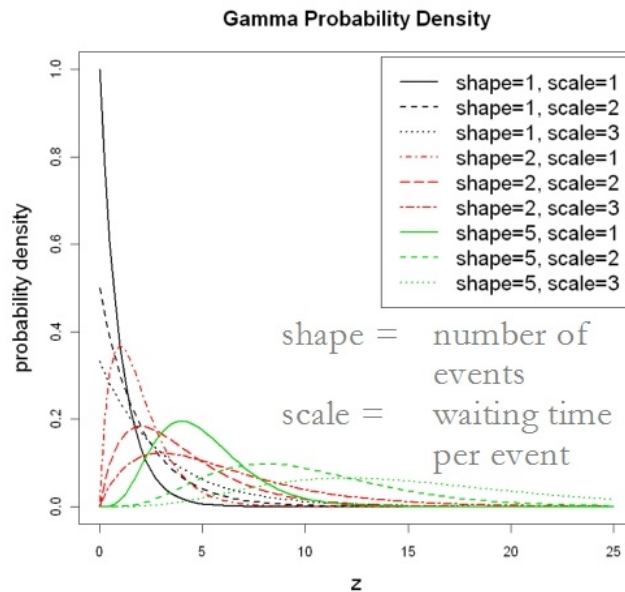
The normal distribution is used everywhere and is the basis for most classical statistical methods. It arises from the sum of many independent samples drawn from the same distribution (of any shape) and its ubiquity stems from the Central Limit Theorem, which says that if you add a large number of independent samples from the same distribution, the distribution of the sum will be approximately normal. The normal distribution has two parameters, the mean and standard deviation, which are independent. Consequently, one often assumes constant variance (as the mean changes), in contrast to the Poisson and binomial distributions where the variance is a fixed function of the mean. The normal distribution is used with continuous, unimodal and symmetric distributions, which are reasonably common in ecology (e.g., temperature, pH, nutrient concentration).

## Probability Distributions... bestiary of functions

### Continuous Distributions

#### Gamma Distribution:

- Distribution of waiting times until a certain number of events (*shape*) takes place given an average waiting time per event (*scale*)
- Continuous counterpart of negative binomial
- Used phenomenologically with continuous, positive data having too much variance (overdispersed normal) and right skew



#### 4.2.2 Gamma

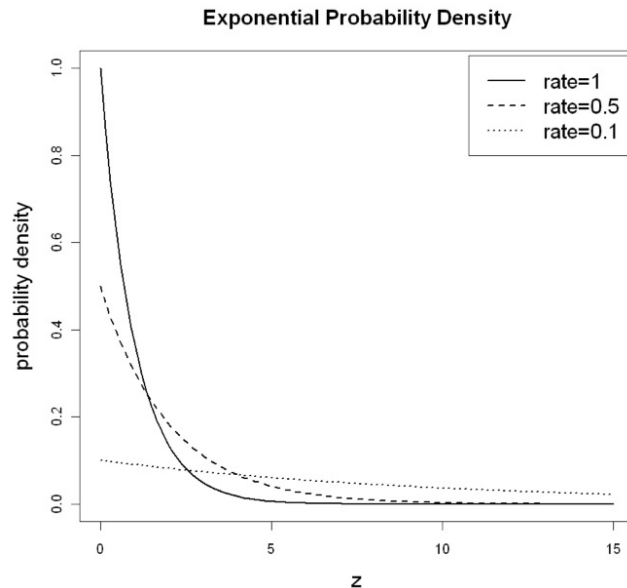
The Gamma distribution gives the distribution of waiting times until a certain number of events (*shape*) takes place, for example, the distribution of the length of time (in days) you would expect to have to wait for 3 deaths in a population, given an average survival time. The Gamma distribution is the continuous counterpart of the negative binomial, which recall is used to describe an overdispersed Poisson with heterogeneous data. The Gamma is similarly used in phenomenological fashion with continuous, positive data having too much variance and a right skew; in other words, an overdispersed normal distribution with a right skew. The Gamma is an excellent choice when only positive real numbers are possible (note the normal is not lower bounded like Gamma).

## Probability Distributions... bestiary of functions

### Continuous Distributions

#### Exponential Distribution:

- Distribution of waiting times for a single event to happen given a constant probability per unit time (*rate*) that it will happen
- Continuous counterpart of geometric and special case of Gamma (*shape*=1)
- Used phenomenologically and mechanistically (very common)



#### 4.2.3 Exponential

The exponential distribution gives the distribution of waiting times for a single event to happen given a constant probability per unit time (*rate*) that it will happen. The exponential is the continuous counterpart of the geometric and is a special case of the Gamma with *shape* = 1. The exponential is used mechanistically when the distribution arises via waiting times, but it is also frequently used phenomenologically for any continuous distribution that has highest probability for zero or small values.

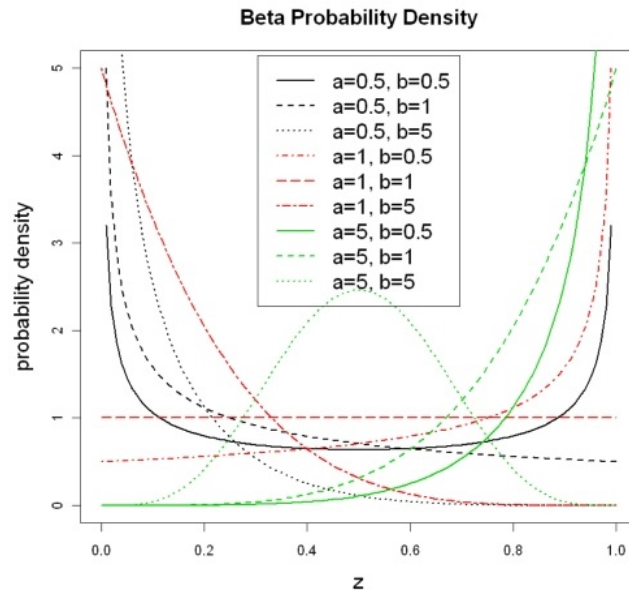


# Probability Distributions... bestiary of functions

## Continuous Distributions

### Beta Distribution:

- Closely related to binomial; distribution of probability of success in a binomial trial with  $a-1$  observed successes and  $b-1$  observed failures
- Used when you have to define any continuous distribution on a finite range (0-1)
- Used with probabilities or proportions (when derived from continuous data)



#### 4.2.4 Beta

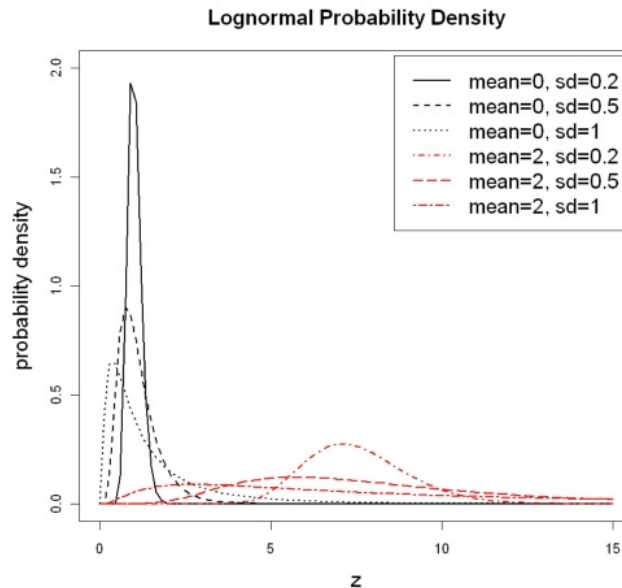
The beta distribution is closely related to the binomial; it gives the distribution of the probability of success in a binomial trial with  $a-1$  observed successes and  $b-1$  observed failures. It is the only continuous probability distribution, besides the uniform, with a finite range, from 0 to 1. This makes it very useful when you have to define any continuous distribution on a finite range. The beta distribution is used with data in the form of probabilities or proportions (e.g., proportional cover).

## Probability Distributions... bestiary of functions

### Continuous Distributions

#### Lognormal Distribution:

- Distribution of the *product* of independent variables, instead of the *sum* in the normal; e.g., sizes or masses of individuals that grow exponentially
- Taking the log of a lognormal variable converts it to a normal
- Often used phenomenologically when Gamma also fits



#### 4.2.5 Lognormal

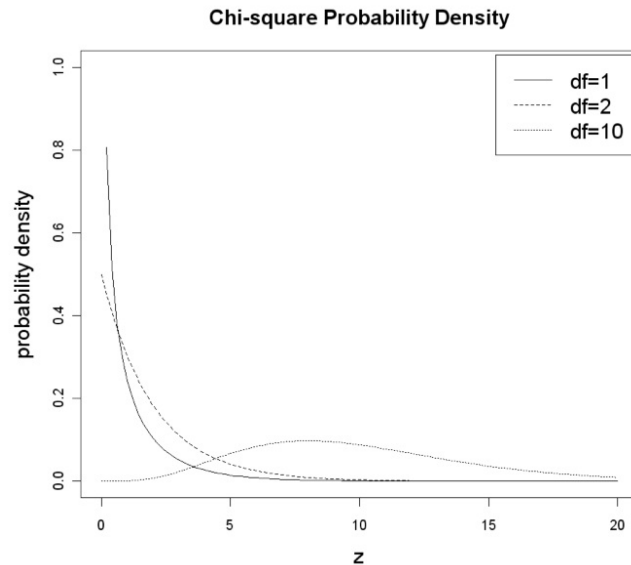
The lognormal gives the distribution of the *product* of independent variables with the same distribution, much like the normal except it stems from the product instead of a sum. The best example of this mechanism is the distribution of the sizes of individuals or populations that grow exponentially, with a per capita growth rate that varies randomly over time. At each time step (e.g., daily, yearly), the current size is *multiplied* by the randomly chosen growth increment, so the final size (when measured) is the product of the initial size and all of the random growth increments. The easiest way of thinking about the lognormal is that taking the log of a lognormal variable converts it to a normal variable. The lognormal is often used phenomenologically in situations when the Gamma also fits.

## Probability Distributions... bestiary of functions

### Continuous Distributions

#### Chi-square Distribution:

- Distribution of the sum of squares of  $n$  (degrees of freedom) normals each with variance one
- Famous for its use in contingency table analysis
- Important because Likelihood ratio statistics are distributed approximately chi-square



#### 4.2.6 Chi-square

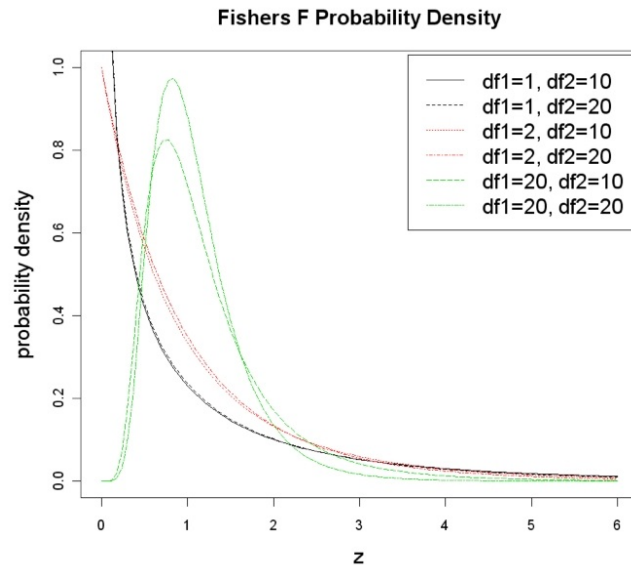
The chi-square distribution gives the sum of squares of  $n$  (degrees of freedom) normals each with variance one. The chi-square distribution is famous for its use in contingency table analysis (cross-classified categorical data) and the analysis of count data. For example, in a table of cross-classified counts (males versus females and present versus absent), the Pearson's chi-squared statistic is equal to the sum of the squared deviations between observed counts and expected counts divided by expected counts, where expected counts are typically under the null hypothesis of independence among factors (sex and presence in this case). The chi-square is also especially important to us because Likelihood ratio statistics (which we will discuss in a later chapter), which are used to test for the difference between nested models, are also approximately distributed chi-squared.

## Probability Distributions... bestiary of functions

### Continuous Distributions

Fisher's  $F$  Distribution:

- Distribution of the *ratio* of two variances – treatment variance to error variance
- Famous for its use in analysis of variance (ANOVA)
- Central to hypothesis testing because of its use in assessing the significance of the differences between two variances



#### 4.2.7 Fisher's $F$

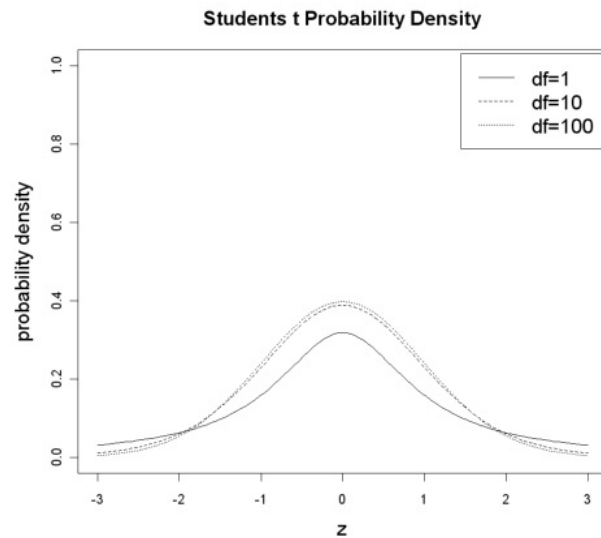
Fisher's  $F$  distribution gives the ratio of the mean squares (variances) of two independent standard normals, and hence of the ratio of two independent chi-squared variates each divided by its degrees of freedom. The  $F$  distribution is famous for its use in analysis of variance (ANOVA) tables, involving the ratio of treatment (or model) and error variance. Note, the  $F$  distribution is used in both classical ANOVA as well as linear regression.

# Probability Distributions... bestiary of functions

## Continuous Distributions

### Student's $t$ Distribution:

- Distribution of a random variable that is the ratio of the difference between a sample statistic and its population value to the standard deviation of the distribution of the sample statistic (standard error)
- Famous for its use in testing for difference between the means of two normally distributed samples and for whether a parameter estimate differs from zero.

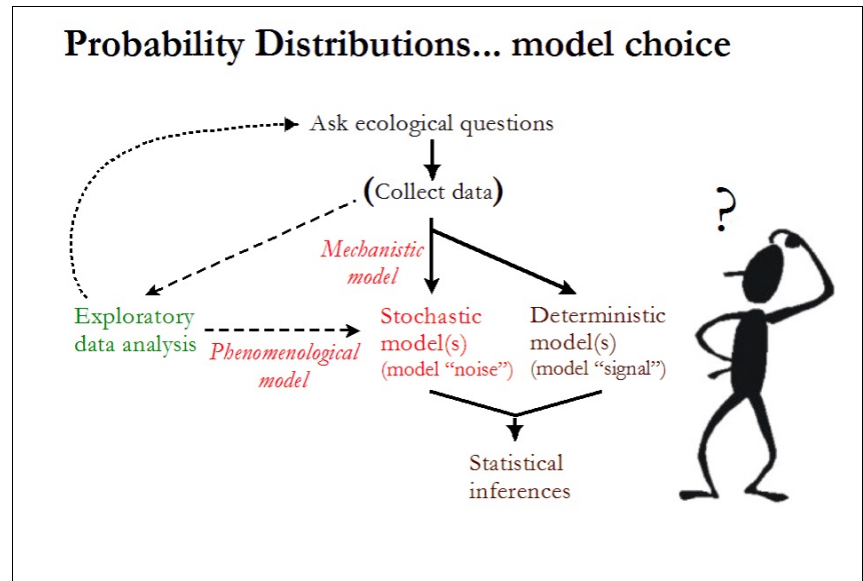
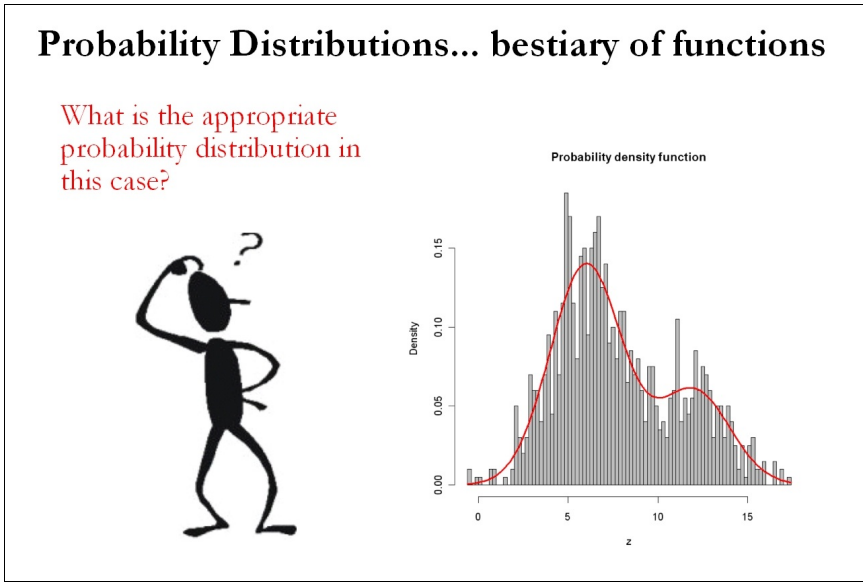


#### 4.2.8 Student's $t$

The Student's  $t$  distribution represents the probability distribution of a random variable that is the ratio of the difference between a sample statistic and its population value to the standard deviation of the distribution of the sample statistic (known as the standard error of the estimate). It is also the overdispersed counterpart for the normal distribution which results from mixing the normal sampling distribution with an inverse gamma distribution for the variance. The  $t$  distribution has fatter tails than the normal and is famous for its use in testing the difference in the means of two normally distributed samples. It is commonly used in testing whether a parameter estimate differs from zero. The  $t$  distribution is also proportional to the  $F$  distribution ( $F=t^2$ ) when there is a single degree freedom in the numerator.

## 5. Choosing the right probability distribution?

Determining the appropriate probability distribution can be just as challenging as choosing the right deterministic function, although in some cases it may require little thought. In the modeling process, the choice of a stochastic model is especially critical because it is the basis for most of the assumptions of parametric statistical methods. In particular, the errors are assumed to follow a specified distribution. If they do not, then the statistical inferences, such as any hypothesis tests, are strictly invalid. Fortunately, in many cases, the methods are surprising robust to minor or even modest infractions of this assumption, but the further your data stray away from the assumptions of the model, the weaker the inferences will be and the greater the chances you have of reaching erroneous conclusions. Thus, choosing a stochastic model carefully is of paramount concern. In many cases, the model choice is made a priori or is intrinsic to the study design. For example, a sample of a fixed number of plots to determine species presence/absence will have a binomial distribution by design. However, there will be many times where an initial examination of the data will provide important insights on the adequacy of a particular model and suggest a different model or perhaps competing models. Time spent carefully considering the right model for the question, given the data, is time well worth spending, as any inferences made are going to be contingent on the model, as are any insights that are gained from the study.



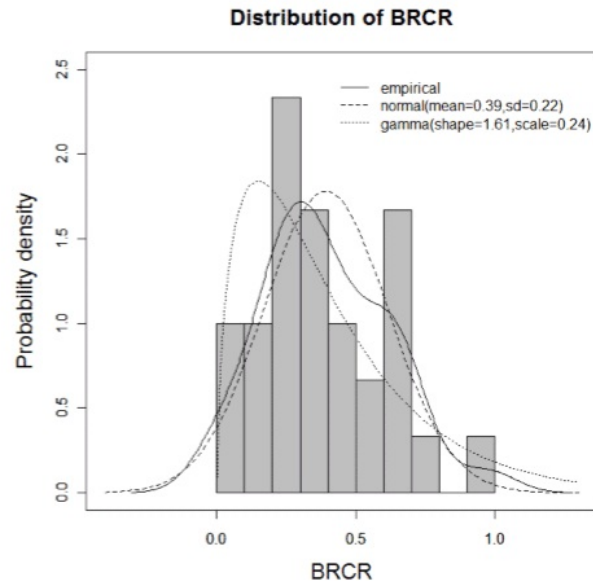
## Statistical Models... putting it all together

Example: *linear model (brown creeper abundance)*

$$Y_i \sim \text{Normal}(\text{mean} = \mu, \text{sd} = \sigma)$$

Alternatively

$$Y_i \sim \text{Gamma}(\text{shape} = \alpha, \text{scale} = \text{mean}/\alpha)$$



### 6. Statistical Models – putting it all together

Let's go back to our three initial real-world examples introduced in the previous section on deterministic functions and now consider the full statistical model, comprised of both the deterministic component and the stochastic component, for each example.

*Example 1: Linear model (brown creeper abundance)*

In this example, recall that the dependent variable  $Y$  (brown creeper abundance) is hypothesized to be linearly dependent on the independent variable  $X$  (extent of late-successional forest). Before exploring this linear relationship, however, let's begin with an even simpler model of the distribution of brown creeper abundance without relationship to any independent variable.

The raw empirical distribution is given by the histogram or, more appropriately (since it is a continuous variable), by the smoothed kernel density line shown in the figure. We can fit a parametric statistical model by assuming a particular probability distribution. Let's begin with the normal distribution. This simple statistical model contains no deterministic component – it is entirely error; actually, it does contain a deterministic component but it is just a constant, the mean. We can specify this model as follows:

$$Y_i \sim \text{Normal}(\text{mean} = \mu, \text{sd} = \sigma)$$

Read this as “ $Y$  is a random variable distributed (represented by tilda) normally with a *mean* equal to  $\mu$  and a standard deviation *sd* equal to  $\sigma$ .” If we estimate the mean from the sample mean (0.39) and the standard deviation from sample standard deviation (0.22), the normal distribution matches the empirical distribution pretty well as shown in the figure, so the normal probability distribution might be a reasonable choice for this dataset. However, since the normal distribution allows for negative values and our random variable (brown creeper abundance) cannot be negative, we might also try the Gamma distribution which allows only positive values. We can specify the gamma model as follows:

$$Y_i \sim \text{Gamma}(\text{shape} = \alpha, \text{scale} = \text{mean}/\alpha)$$

Note, in this model there are two parameters, shape and scale. Unfortunately, neither of these parameters are equal to the mean. The mean of the Gamma is equal to the product of shape and scale. Consequently, we can reparameterize the Gamma as shown, with scale equal to the mean divided by shape. If we were to estimate shape equal to 1.61 and scale equal to 0.24 (mean of  $Y$  divided by 1.61), the Gamma distribution is positively skewed and does not appear to match the empirical distribution as well as the normal.

If we were only interested in describing the distribution of brown creeper abundance, we would be done with our statistical modeling. However, we are usually interested in trying to explain variation in the dependent variable by one or more independent variables. This is where the deterministic model comes into play. We use a deterministic function to model how the expected value of the dependent variable (the mean) varies as some function of the variation in one or more independent variables. In this example, we want to know if the expected value of brown creeper abundance increases linearly with increasing extent of late-successional forest.



## Statistical Models... putting it all together

Example: *linear model (brown creeper abundance)*

$$Y_i = a + b \cdot X_i + \varepsilon_i$$

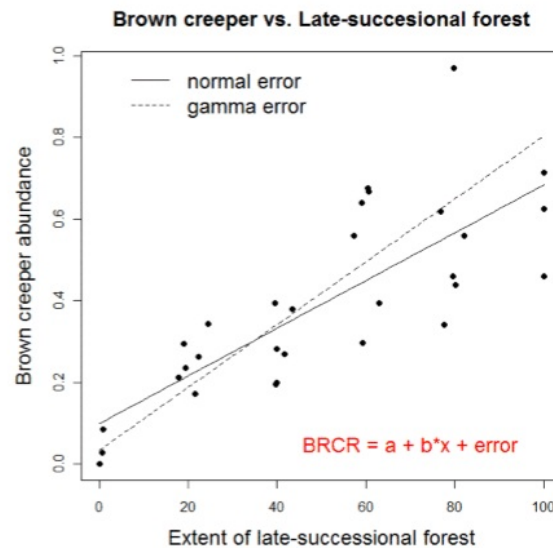
$$\varepsilon_i \sim \text{Normal}(\text{mean} = 0, \text{sd} = \sigma)$$

Or

$$Y_i \sim \text{Normal}(\text{mean} = a + b \cdot X_i, \text{sd} = \sigma)$$

Alternatively

$$Y_i \sim \text{Gamma}(\text{shape} = \alpha, \text{scale} = a + b \cdot X_i / \alpha)$$



In this example, we might hypothesize that the dependent variable  $Y$ , brown creeper abundance, is linearly dependent on the independent variable  $X$ , extent of late-successional forest, with normally distributed (and independent) errors. The conventional specification of this model is as follows:

$$Y_i = a + b \cdot X_i + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(\text{mean} = 0, \text{sd} = \sigma)$$

An appealing feature of this model specification is that it makes explicit the *distinction* between the deterministic and stochastic components of the model. An alternative and more succinct specification of this model is as follows:

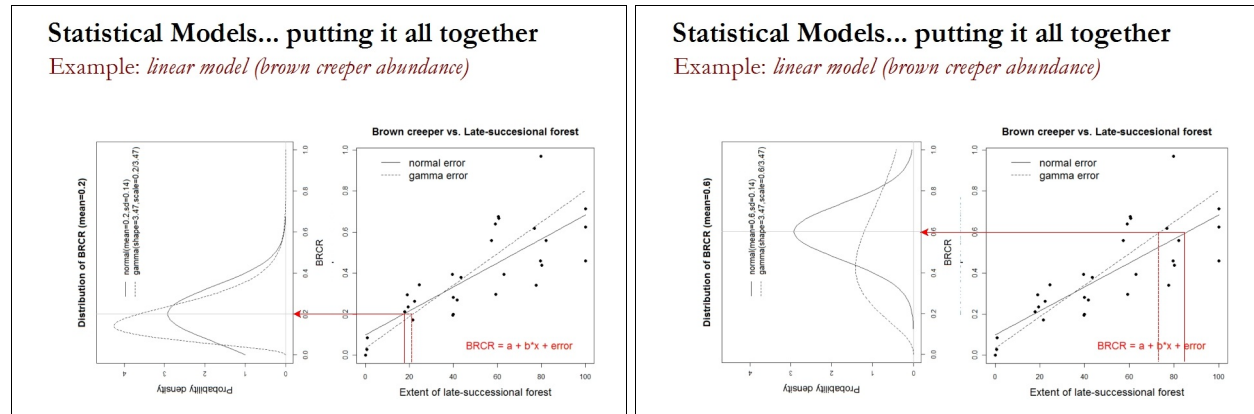
$$Y_i \sim \text{Normal}(\text{mean} = a + b \cdot X_i, \text{sd} = \sigma)$$

An appealing feature of this model specification is that it makes explicit the *link* between the deterministic and stochastic components of the model. This formulation is the one we will use throughout and is interpreted as follows:  $Y$  is distributed (represented by tilda) normally with a mean equal to a linear function of  $X$  and a standard deviation equal to  $\sigma$ . In this model, there are three parameters (to be estimated):  $a$  and  $b$  are the intercept and slope of the linear model and are thus associated with the deterministic component of the model, and  $\sigma$  is the standard deviation of the normal probability distribution and is thus associated with the stochastic component of the model.

Alternatively, we might specify the model with a Gamma error distribution as follows: It is important to recognize the link between the deterministic model and the stochastic model. In most cases (including all that we have discussed thus far), the deterministic model is used to represent how the independent variable(s) relates to the expected value (mean) of the dependent

$$Y_i \sim \text{Gamma}\left(\text{shape} = \alpha, \text{scale} = \frac{a + b \cdot X_i}{\alpha}\right)$$

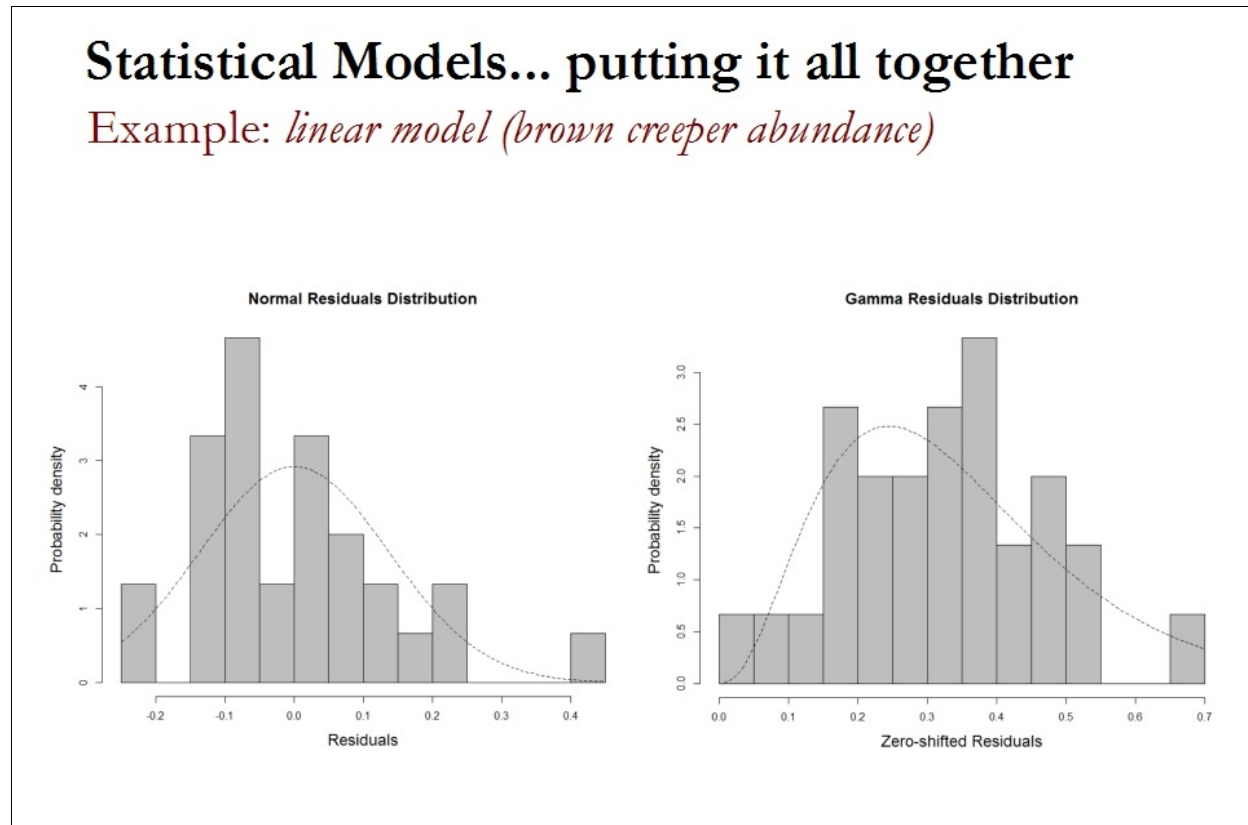
variable. Thus, the link is via the *mean* of the probability distribution. Recall that the normal probability distribution has two parameters: the mean and the standard deviation. If all we are interested in is explaining the distribution of the dependent variable without relation to any independent variable, then we would not have a deterministic model (but in truth, it would simply be a constant, the mean) and the entire model would be comprised of the stochastic component alone. In this case, we would have two parameters, the mean and standard deviation. However, if we are interested in explaining the dependent variable in relation to one or more independent variables, which we usually are, then we also need a deterministic model. Instead of simply modeling the mean as a single parameter, we replace the mean parameter with our deterministic function (the linear model in this case), which we use in order to specify the mean as a function of an independent variable. Because the normal distribution has a parameter that is equal to the mean, the link between the deterministic function and the stochastic distribution is straightforward. Unfortunately, this is not always the case. In the Gamma model, the mean is the product of shape and scale, so we have to work the mean in a little differently. In the case shown here, we specified the scale parameter to be equal to the mean divided by shape. For all practical purposes, this parameterization is no real concern. We are still modeling the mean as a linear function of the independent variable and the errors are distributed Gamma about that mean.



Let's take a graphical look at how the deterministic and stochastic components of the model link together. The deterministic linear model (once the slope and intercept parameters have been estimated) tells us the expected value (mean) of brown creeper abundance. For example, if extent of late-successional forest is approximately 20%, then we expect or predict brown creeper abundance to be 0.2. If our data is being generated from a normal distribution, then we expect the observed values of brown creeper abundance to vary normally about a mean of 0.2 with a standard deviation (spread) estimated from the data to be 0.14. Thus, for any observed value of brown creeper abundance in a landscape with roughly 20% late-successional forest, we can use the probability density function to determine the probability of observing that value. Values close to the mean will have a relatively high probability and values far from the mean will have a relatively low probability. Similarly, if our data is being generated from a Gamma distribution, then we expect the probability density function to be different. In this case, we estimated the shape to be equal to 3.47 and thus the scale to be equal to 0.2/3.47. The resulting probability distribution gives the probability of observing any particular value of brown creeper abundance if the data came from a Gamma distribution. Note, the Gamma distribution is considerably flatter (i.e., more spread) than the normal and is positively skewed, which might be quite reasonable for this data. We can repeat this process for any value of the independent variable. For example, if extent of late-successional forest is roughly 80%, then we expect brown creeper abundance to be 0.6, and the probability of observing any particular value to be given by the corresponding probability distributions.

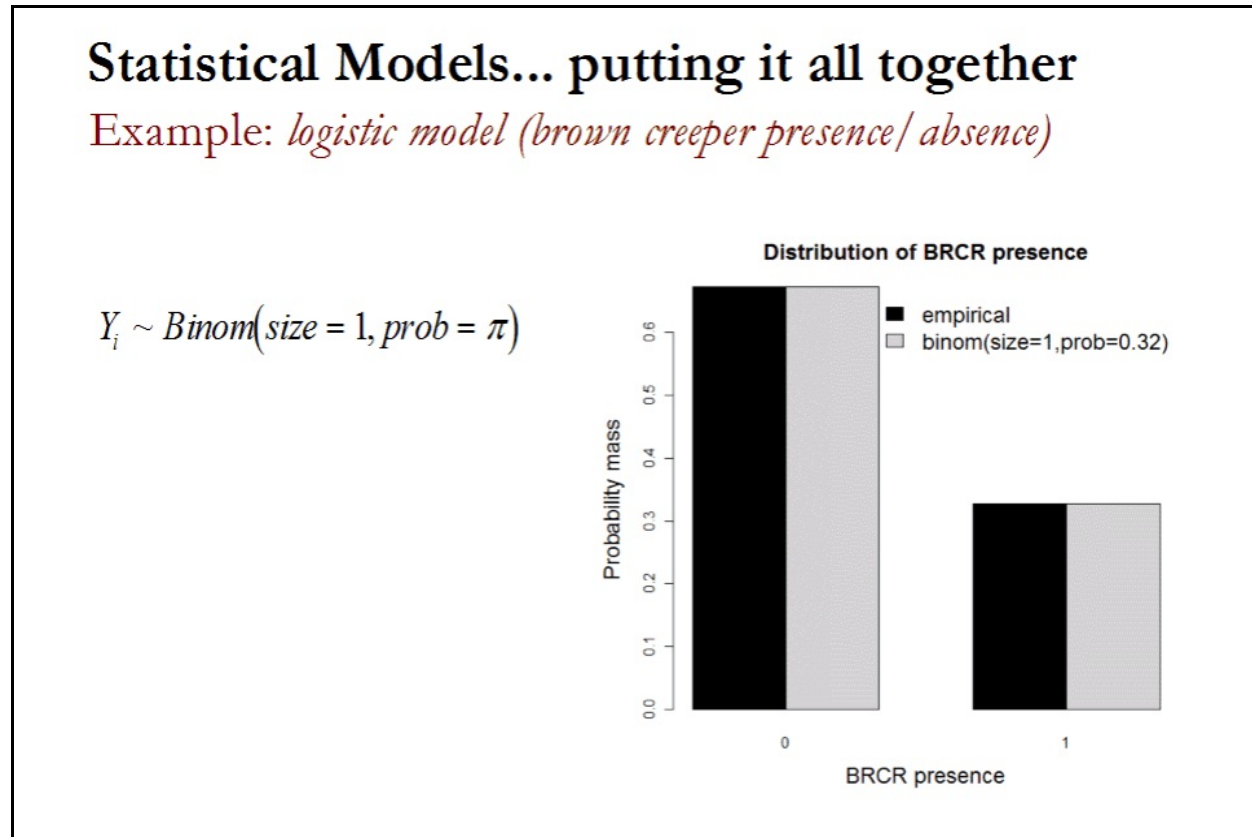
## Statistical Models... putting it all together

Example: *linear model (brown creeper abundance)*



Although the probability distribution is determined by the type (and scale) of the dependent variable, and that should be the paramount consideration in choosing a probability distribution, the formal statistical assumption of the final model is that the residuals (or errors) are (typically) independently and identically distributed according to the specified probability distribution. Thus, the customary way to evaluate the adequacy of a statistical model and whether the statistical assumptions have been met is to analyze the residuals of the model. One way to do this is simply to plot the residuals and verify that the shape of the distribution is consistent with the chosen probability distribution.

In the figures shown here, the left plot shows the distribution of the residuals from the model fit with the Normal errors, and I have overlaid a normal curve with a mean of zero (the expected value for the residuals in this case) and a standard deviation estimated from the model fitting process (to be discussed later). In this case it seems that the residuals plausibly come from a Normal distribution, although the residual distribution seems to be slightly right-skewed when a Normal distribution should be perfectly symmetrical. The right plot shows the distribution of the residuals from the model fit with Gamma errors, and I have overlaid a Gamma curve (fit by the model) after shifting the residuals to have a minimum value of zero. In this case, since the Gamma can take on right skew, the distribution seems to correspond better to the observed residuals and we might thus conclude that the Gamma is the more appropriate distribution to use in this case.



*Example 2: logistic model (brown creeper presence/ absence)*

In this example, recall that the dependent variable  $Y$  (brown creeper presence) is hypothesized to be logistically dependent on the independent variable  $X$  (total basal area). Before exploring this relationship, however, let's begin with a simpler model of the distribution of brown creeper presence without relationship to any independent variable.

The raw empirical distribution is given by the barplot shown here. We can fit a parametric statistical model by assuming a particular probability distribution. Let's assume the binomial distribution, which is perfectly suited for binary response data. This simple statistical model contains no deterministic component – it is entirely error (but in truth, it is simply a constant, the mean). We can

$$Y_i \sim \text{Binom}(\text{size} = 1, \text{prob} = \pi)$$

specify this model as follows:

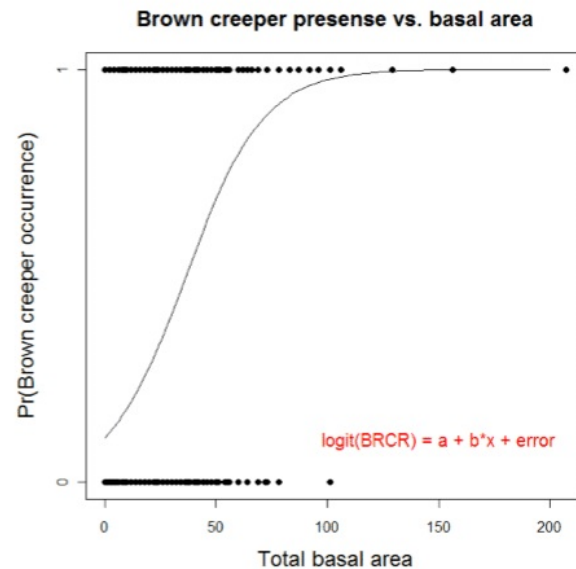
Read this as “ $Y$  is a random variable distributed binomially with a trial *size* equal to 1 and a per trial probability of success *prob* equal to  $\pi$ .” Note, in this case the trial size is fixed by study design, only *prob* is free to vary. If we estimate the probability of presence to be equal to the proportion of presences in our data set (0.32), the binomial distribution matches the empirical distribution exactly – not surprisingly.

## Statistical Models... putting it all together

Example: *logistic model (brown creeper presence/absence)*

$$\pi_i = \frac{e^{a+b \cdot X_i}}{1 + e^{a+b \cdot X_i}}$$

$$Y_i \sim \text{Binom}(\text{size} = 1, \text{prob} = \pi_i)$$

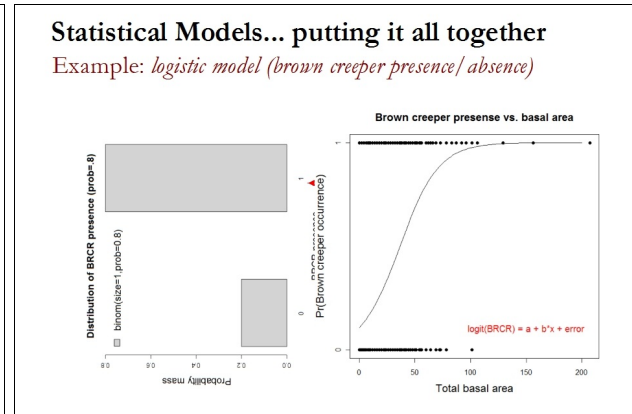
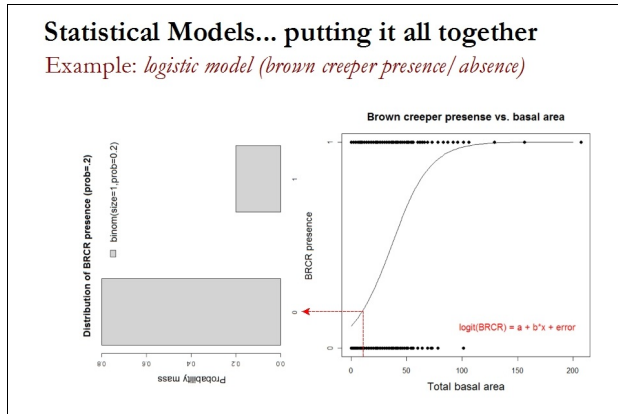


Now let's expand this model to include the logistic relationship between probability of presence and total basal area. In this example, the dependent variable  $Y$  (probability of brown creeper presence) is hypothesized to vary as a logistic function of the independent variable  $X$  (total basal area) with binomially distributed (and independent) errors. The fully specified model is as follows:

$$\pi_i = \frac{e^{a+b \cdot X_i}}{1 + e^{a+b \cdot X_i}}$$

$$Y_i \sim \text{Binom}(\text{size} = 1, \text{prob} = \pi_i)$$

This model is interpreted as follows:  $\pi$  is the per trial probability of success (presence, in this case) and it is a logistic function of  $X$ ;  $Y$  is distributed binomially with a trial size equal to one and a per trial probability of success equal to  $\pi$ . In this model, there are only two parameters (to be estimated):  $a$  and  $b$  are the location and scale parameters of the logistic function and move the inflection point left/right (location) and make the sigmoid curve steeper or shallower (scale). Both parameters are associated with the deterministic component of the model. There are no additional parameters for the stochastic component of the model, because the binomial distribution has only one free parameter when the trial size is fixed at one, and it is the per trial probability of success which is being specified via the deterministic model.



As in the previous example, we use the independent variable to predict the mean of the dependent variable, in this case to determine the expected probability of presence. For example, if total basal area is roughly 10, then we predict there to be a 0.2 probability of presence. The probability of observing either outcome, presence or absence, is then given by the binomial probability mass function for a trial size equal to 1 and a per trial probability of success equal to 0.2. Similarly, for a total basal area equal to roughly 60, we predict there to be a 0.8 probability of presence, and the binomial gives the probability of presence versus absence.

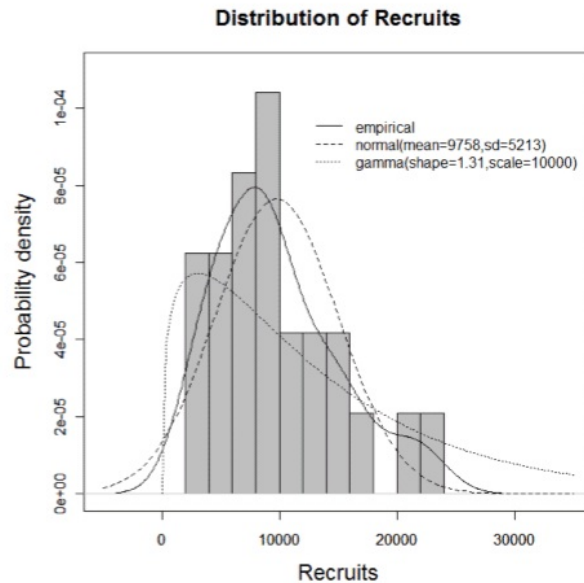
## Statistical Models... putting it all together

Example: *Ricker model (striped bass stock-recruitment)*

$$Y_i \sim \text{Normal}(\text{mean} = \mu, \text{sd} = \sigma)$$

Alternatively

$$Y_i \sim \text{Gamma}(\text{shape} = \alpha, \text{scale} = \text{mean}/\alpha)$$



*Example 3: Ricker model (striped bass stock-recruitment)*

In this example, recall that the dependent variable  $Y$  (recruits) is hypothesized presumed to vary as a Ricker function of the independent variable  $X$  (stock) with either normal or Gamma distributed (and independent) errors. Before exploring this nonlinear relationship, however, let's begin with an even simpler model of the distribution of recruits without relationship to any independent variable.

The raw empirical distribution is given by the histogram or, more appropriately (since it is a continuous variable), by the smoothed kernel density line shown in the figure. We can fit a parametric statistical model by assuming a particular probability distribution. Let's begin with the normal distribution. This simple statistical model contains no deterministic component – it is entirely error (but in truth, it is simply a constant, the mean). We can specify this model as follows:

$$Y_i \sim \text{Normal}(\text{mean} = \mu, \text{sd} = \sigma)$$

Read this as “ $Y$  is a random variable distributed normally with a *mean* equal to  $\mu$  and a standard deviation *sd* equal to  $\sigma$ .” If we estimate the mean from the sample mean (9758) and the standard deviation from sample standard deviation (5213), the normal distribution matches the empirical distribution reasonably well as shown in the figure, so the normal probability distribution might be a reasonable choice for this dataset. However, since the normal distribution allows for negative values and our random variable (recruits) cannot be negative, and the normal distribution is symmetrical



but our data appear to be positively skewed, we might also try the Gamma distribution which allows only positive values and allows for positive skew. We can specify the gamma model as follows: Note, as in the first example, we have parameterize the Gamma as shown, with scale equal to the mean divided by shape. If we were to estimate shape equal to 1.31 and scale equal to 10000, the Gamma distribution is positively skewed and also appears to match the empirical distribution reasonably well.

$$Y_i \sim \text{Gamma}(\text{shape} = \alpha, \text{scale} = \text{mean}/\alpha)$$

If we were only interested in describing the distribution of recruits, we would be done with our statistical modeling. However, in this example, we want to know if the expected value of recruits varies according to a Ricker function of stock.

## Statistical Models... putting it all together

Example: *Ricker model (striped bass stock-recruitment)*

$$Y_i = a \cdot X_i \cdot e^{-b \cdot X_i} + \varepsilon_i$$

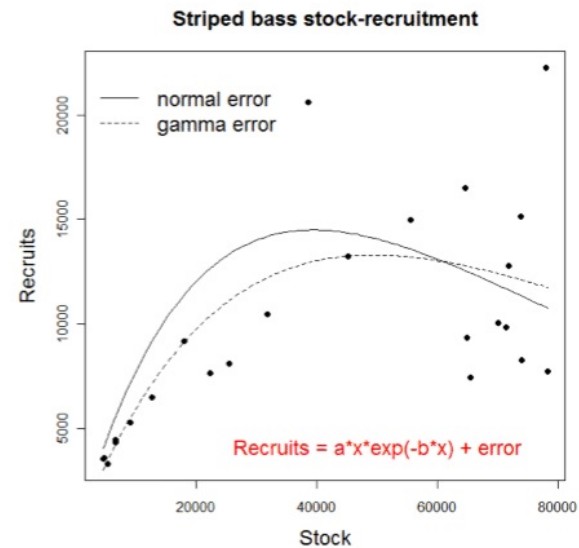
$$\varepsilon_i \sim \text{Normal}(\text{mean} = 0, \text{sd} = \sigma)$$

Or

$$Y_i \sim \text{Normal}(\text{mean} = a \cdot X_i \cdot e^{-b \cdot X_i}, \text{sd} = \sigma)$$

Alternatively

$$Y_i \sim \text{Gamma}\left(\text{shape} = \alpha, \text{scale} = \frac{a \cdot X_i \cdot e^{-b \cdot X_i}}{\alpha}\right)$$



The fully specified model is as follows:

$$Y_i = a \cdot X_i \cdot e^{-b \cdot X_i} + \varepsilon_i$$

$$\varepsilon_i \sim \text{Normal}(\text{mean} = 0, \text{sd} = \sigma)$$

Alternatively, we can specify the model more succinctly as follows:

$$Y_i \sim \text{Normal}(\text{mean} = a \cdot X_i \cdot e^{-b \cdot X_i}, \text{sd} = \sigma)$$

This model is interpreted as follows:  $Y$  (recruits) is a random variable distributed normally with a mean equal to a Ricker function of  $X$  (stock) and standard deviation equal to sigma.

Alternatively, we might specify the model with a Gamma error distribution as follows:

$$Y_i \sim \text{Gamma}\left(\text{shape} = \alpha, \text{scale} = \frac{a \cdot X_i \cdot e^{-b \cdot X_i}}{\alpha}\right)$$

In this case, the model is read as follows:  $Y$  (recruits) is a random variable distributed Gamma with a shape equal to alpha and scale equal to a Ricker function of  $X$  (stock) divided by alpha. In this

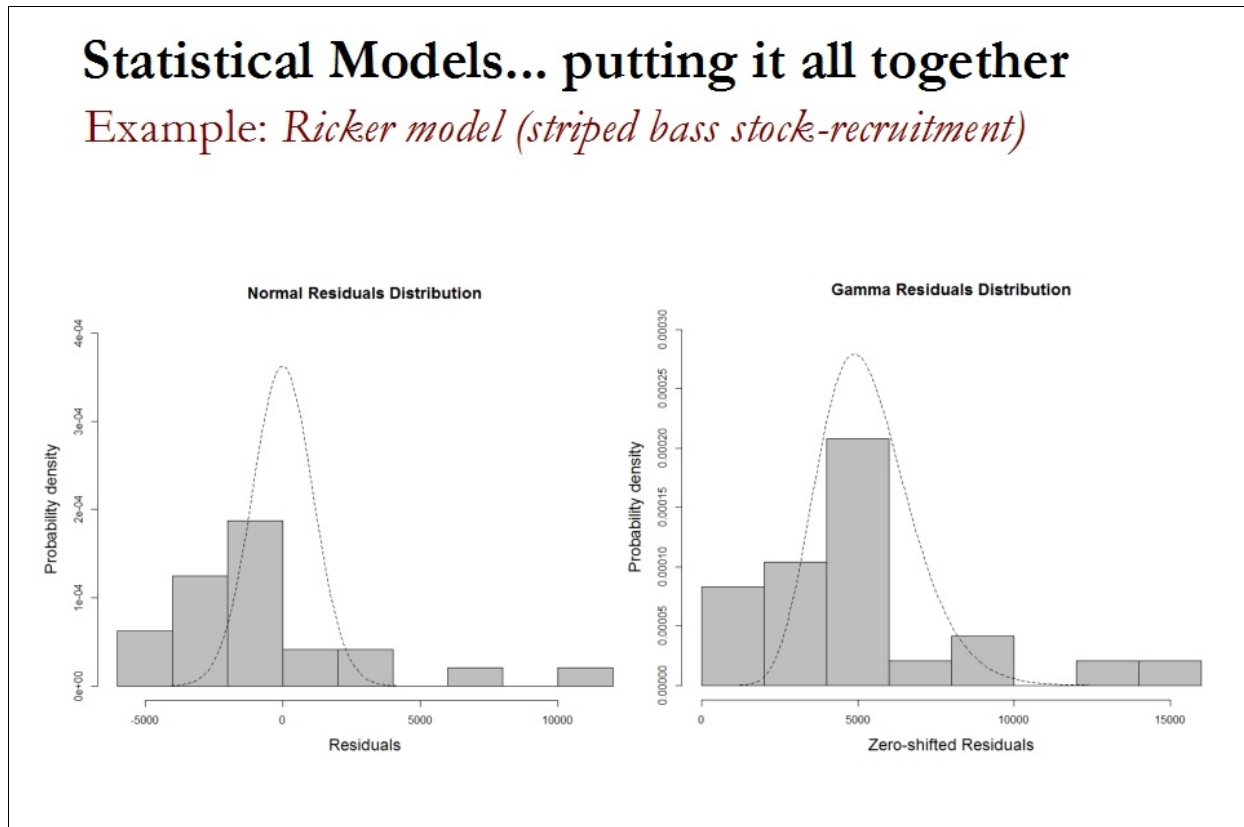
model, there are three parameters (to be estimated):  $a$  and  $b$  are parameters of the Ricker function and influence the location and shape of the curve and are thus associated with the deterministic component of the model, and  $\alpha$  is the shape parameter of the Gamma distribution and is thus associated with the stochastic component of the model. Note, the mean of the Gamma distribution is equal to the product of the scale parameter and the shape parameter (see Bolker). Thus, to link the deterministic model to the stochastic model, we have to redefine the scale parameter to be equal to the mean (expected value of  $Y$ ) divided by shape.



Now let's take a graphical look at how the deterministic and stochastic components of the model link together. Once the parameters of the deterministic model ( $a$  and  $b$ ) have been estimated, it tells us the expected value (mean) of the dependent variable recruits. For example, if stock is approximately 1000, then we expect or predict recruits to be roughly 7500. If our data is being generated from a normal distribution, then we expect the observed values of recruits to vary normally about a mean of 7500 with a standard deviation (spread) estimated from the data to be 1099. Thus, for any observed value of recruits coming from a stock of roughly 1000, we can use the probability density function to determine the probability of observing that value. Values close to the mean will have a relatively high probability and values far from the mean will have a relatively low probability. Similarly, if our data is being generated from a Gamma distribution, then we expect the probability density function to be different. In this case, we estimated the shape to be equal to 12.91 and thus the scale to be equal to 7500/12.91. The resulting probability distribution gives the probability of observing any particular value of recruits if the data came from a Gamma distribution. Note, the Gamma distribution is considerably flatter (i.e., more spread) than the normal and is positively skewed, which might be quite reasonable for this data. We can repeat this process for any value of the independent variable. For example, if stock is roughly 5000, then we expect recruits to be 1350, and the probability of observing any particular value to be given by the corresponding probability distributions.

## Statistical Models... putting it all together

Example: *Ricker model (striped bass stock-recruitment)*



In the figures shown here, the left plot shows the distribution of the residuals from the model fit with the Normal errors, and I have overlaid a normal curve with a mean of zero (the expected value for the residuals in this case) and a standard deviation estimated from the model fitting process (to be discussed later). In this case it seems that the residuals plausibly come from a Normal distribution, although the residual distribution seems to be slightly right-skewed when a Normal distribution should be perfectly symmetrical. The right plot shows the distribution of the residuals from the model fit with Gamma errors, and I have overlaid a Gamma curve (fit by the model) after shifting the residuals to have a minimum value of zero. In this case, since the Gamma can take on right skew, the distribution seems to correspond better to the observed residuals and we might thus conclude that the Gamma is the more appropriate distribution to use in this case.