

# Analysis of Environmental Data

## Distributions: Notation, Functions, and Probability

Michael France Nelson

Eco 602 – University of Massachusetts, Amherst – Fall 2021  
Michael France Nelson

# Announcements Sep 30

- It's almost October!
- I'm delaying the in-class activity until Tuesday.
- No assignments are due this Sunday!
- Week 5 reading questions are due one week later (Oct 10).
  - There are a lot of probability questions – I want to have at least one hands-on-acorns in-class session before you submit.
  - Probability in-class activity on Tuesday will cover the relevant reading questions.
  - Week 6 reading questions will still be due on Oct 10.

# Probability and Distributions: Probability Theory

## **Probability theory concerns the likelihood of events**

Distributions are tools for describing the likelihood of observing specific events from the set of all possible events.

- They map *events* to *likelihoods*

There are many named *parametric* distributions with well-understood, useful, and sometimes surprising properties.

## **Probability theory gets complicated and difficult *very* quickly!**

- I'll attempt to help you develop intuition about the most essential parts.

# Probability Theory Essentials

## Probabilities are non-negative

- A probability can be any value between zero and 1.0, inclusive.
- The probability of a specific event is usually less than 1.0
- **Law of total probability:** The sum of the probabilities of all possible events is 1.0

## Sample space: the set of all possible events

- Events: a possible outcome of a stochastic process
- The definition of event is context specific:
  - “What is the probability of catching a fish that weighs 405 grams?”
- “What is the probability of catching a fish that weighs between 399 and 411 grams?”
- “What is the probability of catching a fish that weighs less than 200 grams?”

# Probability Notation Basics

## Basic probability

- $\Pr(A) = 0.05$ 
  - Read as: “The probability that event A occurs is 5%”

## Joint probability

- $\Pr(A \text{ and } B) = \Pr(A \cap B) = 0.05$ 
  - Read as: “The probability that both events A and B occur is 5%”

## Conditional Probability

- $\Pr(A|B) = 0.05$ 
  - Read as: “The probability that event A occurs, given that B has already occurred is 5%”

# Independent Events

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)$$

- The probability that A and B both occur is equal to the product of the individual probabilities...
- We'll dissect this surprisingly important definition.



# Independent events

Events are independent if knowing the value of one observation gives us no information about the value of another observation:

1. I measure the temperature in Neuquén, Argentina on November 23, 1823.
2. I measure the temperature in Amherst on July 4, 2020.

The Neuquén temperature in 1823 probably doesn't tell me much about Amherst in 2020

- Likewise, the temperature here today probably won't tell me much about what to expect there! [Other than knowing that it is fall here, and spring there!]

# Independent events

## Non-Independent Temperatures

Compare the previous temperature example to:

1. I measure the temperature in Amherst on July 4, 2020 at 4:05PM (it is 20C)
2. I measure the temperature in Amherst on July 4, 2020 at 4:11PM (it is 21C)

The temperature at 4:05 gives me a lot of information about what the temperature will be in the same location six minutes later.





# Independent events

Suppose we are equally likely to observe these temperatures:

Temperatures on July 4<sup>th</sup>:

$$Pr(temp = 19C) = 0.05$$

$$Pr(temp = 20C) = 0.05$$

$$Pr(temp = 21C) = 0.05$$

**Independent events: joint probability is product of individual probabilities**

If successive temperature measurements were independent:

- $Pr(20) * Pr(21) = 0.05 * 0.05 = 0.0025$  or about 0.25%

Do you think observing a temperature of 20, followed by a temperature of 21 in the same location 6 minutes later is only 0.25%

It's probably closer to 5%, i.e. the probability of observing 20C.

# Independent events

**If events are independent, the probability of observing a *specific* set of events (the joint probability) is the same as the product of the events of the individual events.**

- I pick up an acorn in each hand simultaneously, from a *very large* collection of acorns.
- Does knowing that the acorn in my left hand is from a Bur Oak tell me anything about the acorn in my right hand?

## **Independence and Maximum Likelihood**

- This may not seem important now, but it is *crucial* to the likelihood concepts we'll examine later.
- It's also key to understanding Bayes' Rule.

# Functions and Formulae

components and intuition

# Key concepts

## Notation

- Variables and Constants
- Arithmetic operators
- Summation notation
- Product notation (we'll use this later with likelihood)
- Set notation
- Bar notation
- Capital and lower-case letters

## Common Formula Chunks

- Means: summation and bar notation:
  - $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
- Sums of Squares: summation and bar notation
  - $SSE = \sum_{i=1}^n (x_i - \bar{x})^2$

# Interpreting formulae

## Strategies

- Identify constants and variables
- Learn to recognize common 'chunks'
- Try to identify long-term behavior
- Try to identify the class of the function

## This lecture focuses on three common chunks

- Means
- Sums of Squares
- Variance/Covariance

# Common chunks

# Sums of squares

- often used to quantify some sort of 'error'
- use in variance and covariance formulas

## Normalizing constants

- These are often nightmarish, but if you look closely, you can usually ignore them!

## Sample size, sample size correction

- $N-1, n-1$



# Starting simple: the mean

## Arithmetic Mean

$$\frac{\sum_{i=1}^n x}{n}$$

### The mean is a simple concept, right?

- It's just the average value...
- It's what we get if we add up all the numbers and divide by the count.

### What do we need to know?

- Our data:
  - A vector of numbers (in R-speak)
- Our quantities:
  - The number of observations
  - The sum of all the observations

# Starting simple: the mean

## We can practice our notation skills:

- Set notation
- Capital/lowercase notation
- Summation notation
- Bar notation
- Normalizing and Sample size notation

## Our x-values in set notation:

$$X = \{x_1, x_2, x_3, \dots, x_n\}$$

\*note the capital X for the set, and the lowercase x for the elements

## The sum of values in sigma notation:

$$\sum_{i=1}^n x_i$$

## The sample size: n



# Starting simple: the mean

Putting it all together: the overall formula

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

The mean is also known as the Expected Value

$$E(X) = \frac{\sum_{i=1}^n x_i}{n}$$

# Sum of squared errors

## This is a common chunk

- *Error* is the difference between an observation and the expected value.
- In words: “The sum of the squared differences between each value and the mean”
- In words: “The sum of the squared errors”

## SSE

$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2$$

# Sum of squared errors

**We already know the mean:  $\bar{x}$**

**The SSE chunk:**

$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2$$

**The sigma decorations are often dropped:**

$$SSE = \sum (x_i - \bar{x})^2$$

**It can be expanded to:**

$$SSE = \sum (x_i - \bar{x})(x_i - \bar{x})$$

**Not so bad. We can learn to see this as a chunk, rather than individual terms!**

# Sum of squared errors

## Why do we want to know this anyway?

Squaring has some desirable properties

- Converts negative values to positive.
- Penalizes high values, i.e. large *errors* are very influential because the squaring function is nonlinear.
- What happens if you take the sum of the non-squared errors?



# Variance and Covariance

# Variance and Covariance: Key Concepts

- Variance and covariance are measures of *spread* or *dispersion*.
  - What are some other measures of dispersion that we know about?
- Variance and covariance are sample statistics - we can use them to estimate population parameters.
- The formulas look intimidating... You won't need to memorize them
- My goal in this section is to build intuition about what they do and how they are notated.
- Variance is a univariate statistic.
- Covariance measures an association between two variables: this is directly analogous to correlation!

# Variance

## What does variance tell us?

- Recall this form for the SSE:

$$SSE = \sum (x_i - \bar{x})(x_i - \bar{x})$$

- It includes the  $(x_i - \bar{x})$  twice.

## It is a *univariate* statistic!

- It characterizes how much spread there is within a variable, with reference to *itself*.
  - That's why we have the  $(x_i - \bar{x})$  two times.

What if we want to know if two variables change in a *coordinated* way?

Var(x) = 0.71



Var(x) = 1.96



# Variance

Variance is a measure of *dispersion* or *spread*

Why do we want to use squared differences?

$$(x_i - \bar{x})^2$$

- In words: “The variance is the average of the squared differences from the mean.”
- It’s just the SSE normalized by the [adjusted] sample or population size.
  - For a population we use  $N$ , a sample uses  $n - 1$
- What is the sign of this term?
- What would happen to the sum if we used the unsquared differences?
- Why not just use the absolute value?



# Variance

**Variance is a measure of *dispersion* or *spread***

- In words: “The variance is the average of the squared differences from the mean.”
- It’s just the SSE normalized by the [adjusted] sample or population size.
  - For a population we use  $N$ , a sample uses  $n - 1$

**Why do we want to use squared differences?**

$$(x_i - \bar{x})^2$$

- What is the sign of this term?
  - It is always positive!
- What would happen to the sum if we used the unsquared differences?
  - They would sum to zero for both dispersed and clustered data.
- Why not just use the absolute value?
  - The squaring penalizes large deviations, this has desirable theoretical and practical consequences.

# Variance

## Formulae: Populations and Samples

- for populations

$$Var(x) = \frac{1}{N} \sum (x_i - \bar{x})^2 = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{N}$$

- Samples require a sample size correction:

$$Var(x) = \frac{1}{n - 1} \sum (x_i - \bar{x})(x_i - \bar{x}) = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

# Covariance

Covariance measures the *dispersion* of one variable,  $x$ , in the context of the *dispersion* of a second variable,  $y$ .

It turns out that *variance* is a special case of *covariance*.

$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

- Covariance tells us the amount by which the changes in one variable are *coordinated* with changes in another.
- $(x_i - \bar{x})(y_i - \bar{y})$  is like a *crossed* version of the squared errors...
  - But the term *cross product* is already taken.

$$Cov(x, x) = Var(x)$$

# Variance and Covariance

## Case 1: positive covariance.

- If the values of  $x$  and  $y$  were somehow coordinated, we might expect that high values of  $x$  would tend to co-occur with high values of  $y$ .
- We use  $(x_i - \bar{x})$  to symbolize the deviation of a sampling unit's  $x$ -value from the mean of  $x$ .
- Similarly  $(y_i - \bar{y})$  is the deviation of a sampling unit's  $y$ -value from the average value of  $y$ .
- The (non-squared) sum of all the deviations of  $x$  is zero (by the definition of the mean).



# Covariance

**Case 1: Positive covariance: High x-values tend to co-occur with high y-values.**

**Most terms will be positive**

**$(x_i > \bar{x})$  AND  $(y_i > \bar{y}) = \text{positive}$**

**$(x_i < \bar{x})$  AND  $(y_i < \bar{y}) = \text{positive}$**

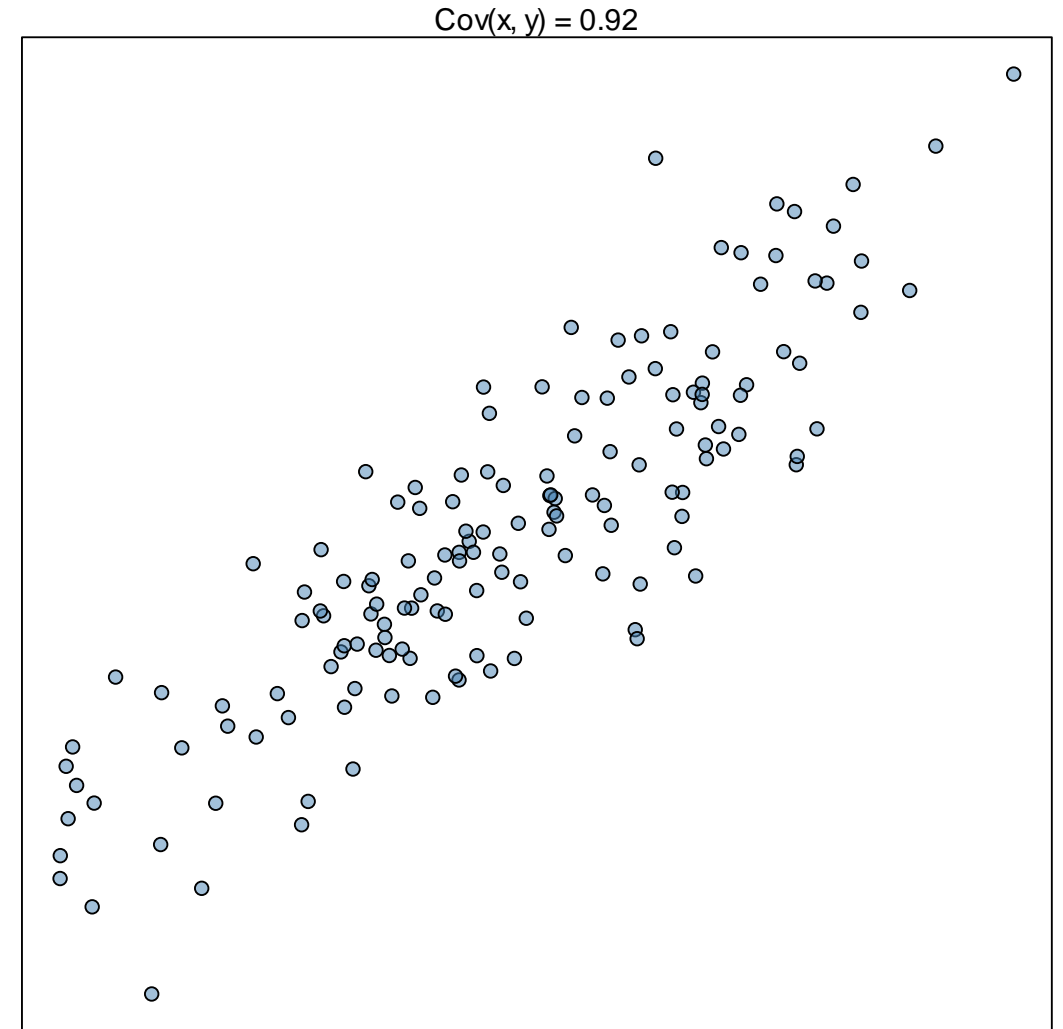
**Few terms will be negative**

$(x_i > \bar{x})$  AND  $(y_i < \bar{y}) = \text{negative}$

$(x_i < \bar{x})$  AND  $(y_i > \bar{y}) = \text{negative}$

# Covariance

Positive Covariance



# Covariance

Case 2: **Negative** covariance: High x-values tend to co-occur with low y-values.

Few terms will be positive

$(x_i > \bar{x})$  AND  $(y_i > \bar{y})$  = positive

$(x_i < \bar{x})$  AND  $(y_i < \bar{y})$  = positive

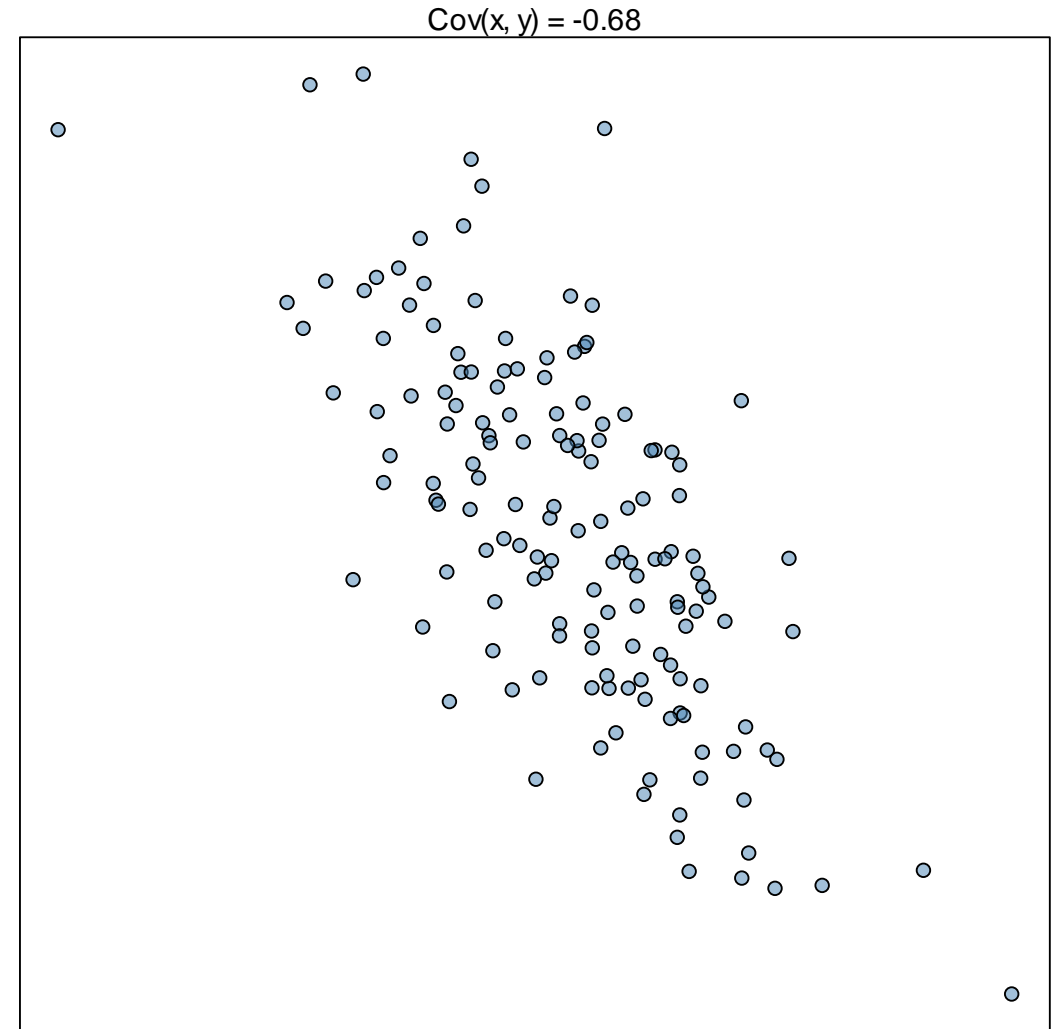
Most terms will be negative

$(x_i > \bar{x})$  AND  $(y_i < \bar{y})$  = negative

$(x_i < \bar{x})$  AND  $(y_i > \bar{y})$  = negative

# Covariance

## Negative Covariance





# Covariance

**Case 3: no covariance: No association between above average x and above average y**  
**Negative and positive values cancel – sum is near zero**

**About half the terms will be positive**

**$(x_i > \bar{x})$  AND  $(y_i > \bar{y})$  = positive**

**$(x_i < \bar{x})$  AND  $(y_i < \bar{y})$  = positive**

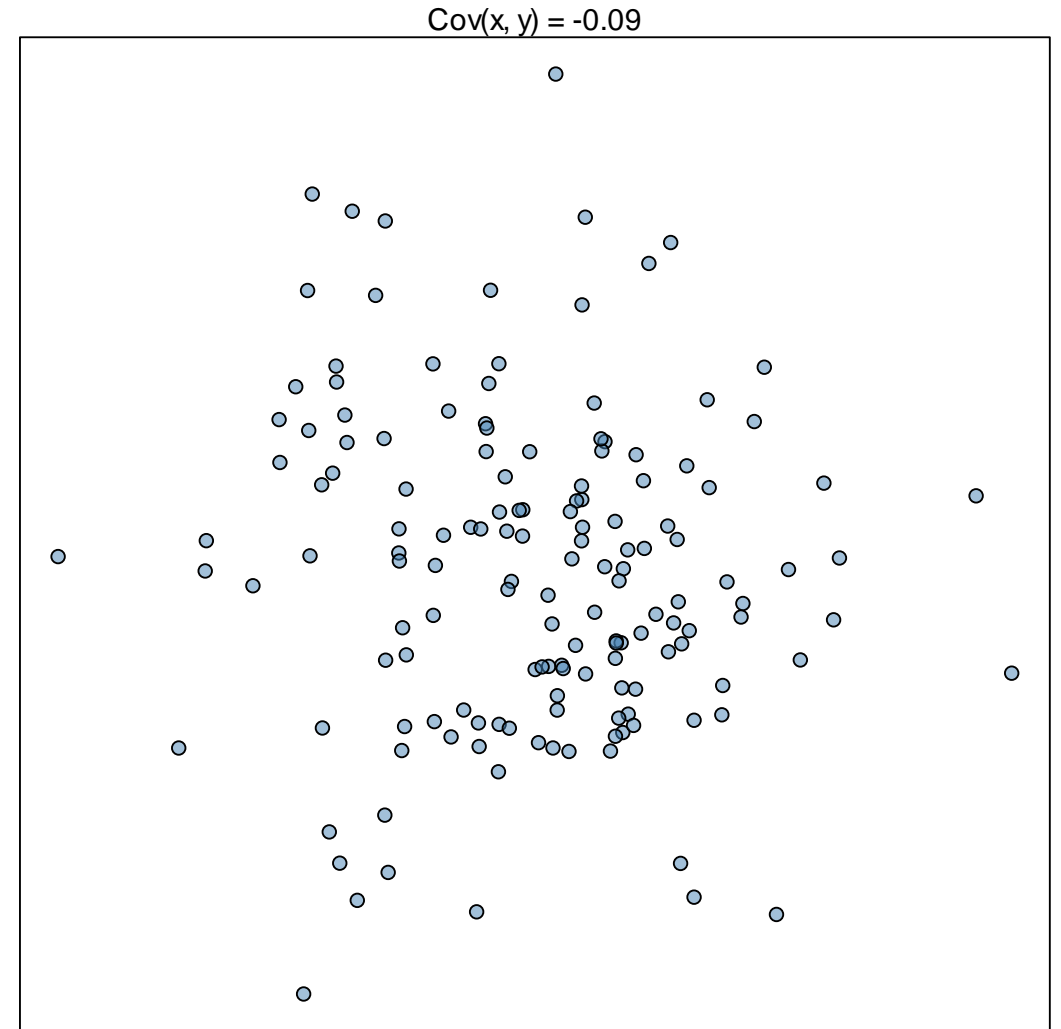
**About half the terms will be negative**

**$(x_i > \bar{x})$  AND  $(y_i < \bar{y})$  = negative**

**$(x_i < \bar{x})$  AND  $(y_i > \bar{y})$  = negative**

# Covariance

Zero Covariance



# Recap

## Common chunks

- Means:  $\bar{x} = \frac{\sum x}{n}$
- Sums of errors:  $\sum (x_i - \bar{x})$ 
  - Remember this sum is zero!
- Sums of squared errors:  $\sum (x_i - \bar{x})^2$
- Sums of squared errors:  $\sum (x_i - \bar{x})(x_i - \bar{x})$
- Sums of crossed errors:  $\sum (x_i - \bar{x})(y_i - \bar{y})$
- Normalizing by population size:
  - $\frac{1}{N}$  and  $\frac{1}{N-1}$

# Next time...

**Pearson correlation walkthrough**

**Normal PDF**

# Discrete Distributions

A Parametric Frequentist Approach

# Key concepts

- What is discrete?
- Discrete sample spaces
- Combinations and permutations
- Bernoulli, Binomial, and Poisson distributions

# Discrete Distributions

**In the world of theoretical distributions, discrete refers to a measurement scale that is:**

- Numeric, i.e. not categorical
  - Not ordinal because the interval between adjacent points is constant:  $4 - 3 = 5 - 4$  but  $medium - low \neq high - medium$
- Cannot take on fractional values
  - They are integers

**Counts, or censuses, are [usually] considered discrete data type**

# Sample Spaces

Events in a discrete distribution can only be integer values.

- That means a discrete distribution has a *finite sample space*, right?



# Sample Spaces

Events in a discrete distribution can only be integer values.

- That means a discrete distribution has a *finite sample space*, right?

No! Non! Nej! Não! ¡No!

It may seem unintuitive, but many discrete [theoretical] distributions have infinite sample spaces.

# The Simplest Distribution?

**One of the easiest distributions to understand is the *Bernoulli Distribution*.**

- It's sample space has only two elements.
- Realizations of a *Bernoulli process* produces *binary* outcome.
- It has one parameter: the probability of *success*.

**It's a special case of the *binomial distribution***

- A realization of the *Bernoulli process* is called a **trial**

# The Binomial distributions

**A binomial process is a collection of  $n$  independent Bernoulli trials.**

- Each Bernoulli trial must have the same probability of success
- Binomial has two *parameters*:  $n$  and  $p$
- $n$  is the number of trials
- $p$  is the probability of *success* in an individual trial (just like the Bernoulli dist.)
- The *sample space* of a binomial distribution has  $n + 1$  elements:
- It's the possible counts of successes, i.e. the set  $\{0, 1, 2, \dots, n\}$

# Classic Example: Coin Tosses

## A series of independent coin tosses is a lot like a binomial process...

Wait a minute, I think I remember something about *independent events* and *joint probabilities*....

- A single coin flip is not very interesting, but consider the sample space for two flips ( $n = 2$ ):

$$\{(T, T), (T, H), (H, T), (H, H)\}$$

- The *sample space* of a binomial distribution has  $n + 1$  elements, which should be 3, but there are 4 elements in the set!
  - Something seems wrong.


# Independent Coin Flips

- Think of each flip as a junction in a tree.
- The first flip has two branches:



# Independent Coin Flips

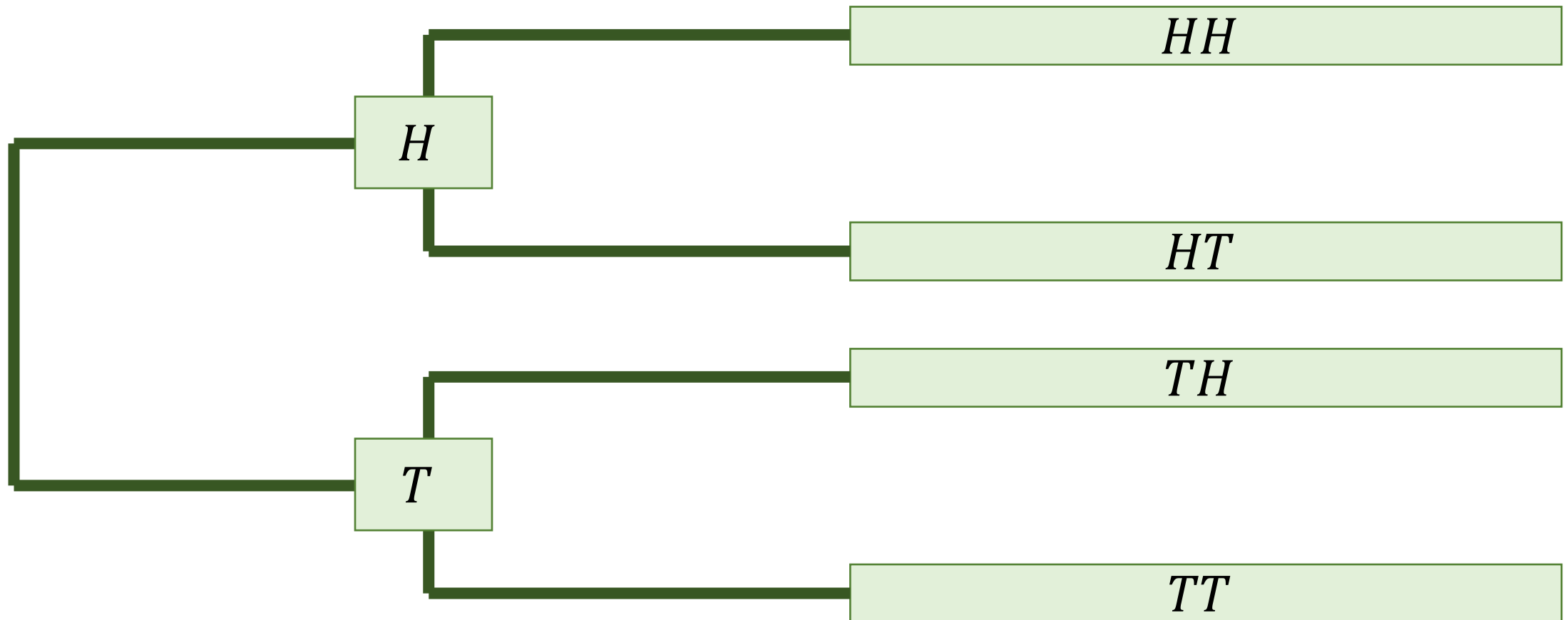
- If the probabilities are equal:


$$\Pr(H) = 0.5$$

$$\Pr(T) = 0.5$$

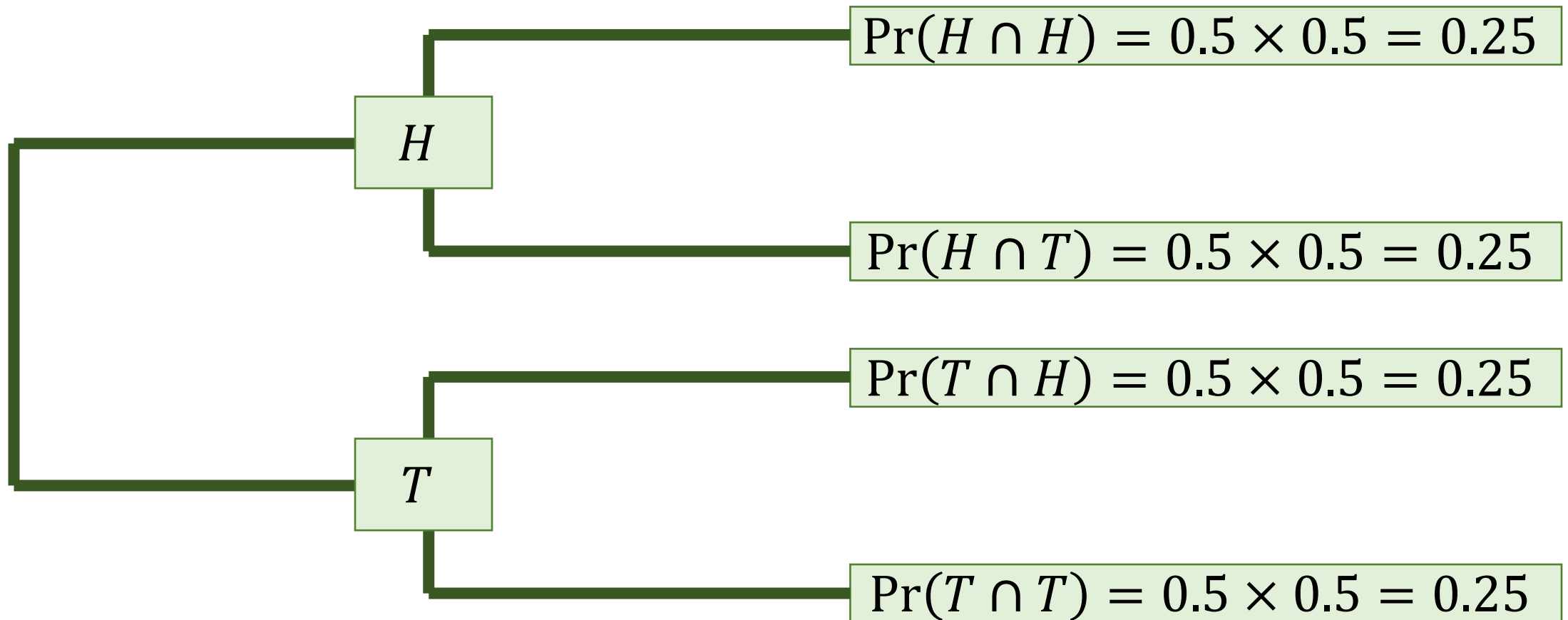
# Independent Coin Flips

- Each of those branches has two branches:



# Independent Coin Flips

- Probabilities of independent events are multiplied:





# Binomial processes and order

**Consider the sample space for two flips ( $n = 2$ ):**

$$\{(T, T), (T, H), (H, T), (H, H)\}$$

**There are *four* elements... But what if we think of them as the number of heads?**

$$\{0, 1, 1, 2\}$$

- Now the size of the sample space makes more sense.
- The *sample space* is really only *three* elements because  $(H, T)$  and  $(T, H)$  are equivalent in the binomial world.

**The *sample space* of a binomial is the range of possible count of successes, i.e. the set  $\{0, 1, 2, \dots, n\}$**

# Combinations and Permutations

**When we consider a binomial process, we care about *combinations*...**

- but we also need to know about *permutations* to characterize the sample space.
- Recall the possible outcomes for 2 coin flips:  $\{(T, T), (T, H), (H, T), (H, H)\}$ 
  - Let's assume that  $Pr(H) = Pr(T) = 0.5$ .

**If the flips are independent, what is  $Pr(H, H)$ ?**

- The joint probability of independent events is the produ.....
- That seems relevant, but the wording doesn't feel intuitive.
- More on combinations and permutations later

# The Poisson Distribution

**Another important discrete distribution is the Poisson.**

It has a single parameter:  $\lambda$

- The Poisson distribution describes counts:
  - A Poisson event is a count, or census.
  - The sample space is  $\{0, 1, 2, \dots, \infty\}$
  - It has an *infinite sample space*!

**How can a discrete distribution have an infinite sample space?**

- Recall that it's a *theoretical* distribution
- Compare the sample to a binomial sample space.

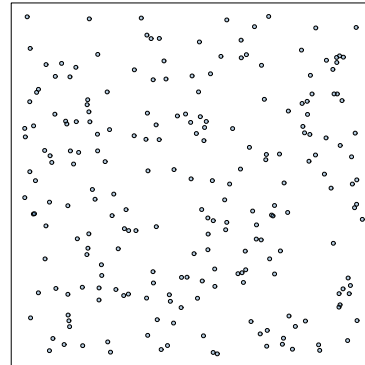
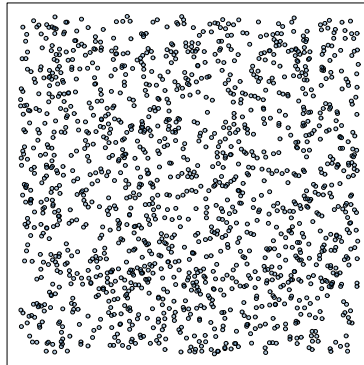
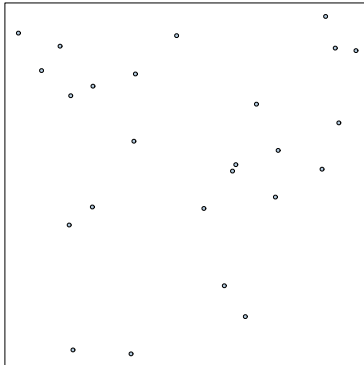
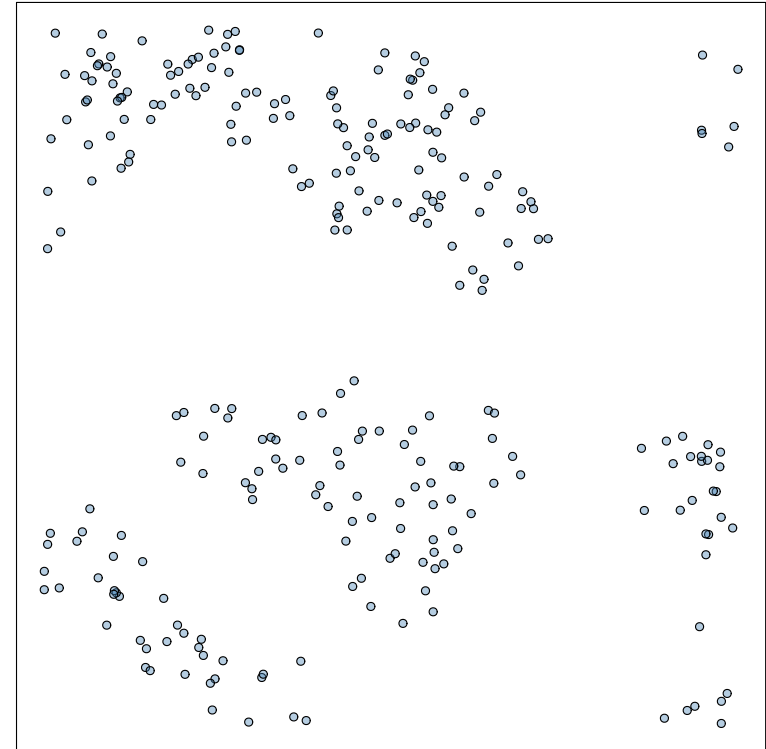
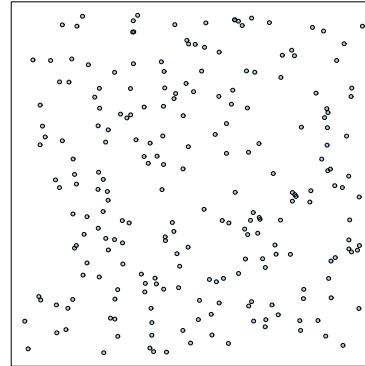
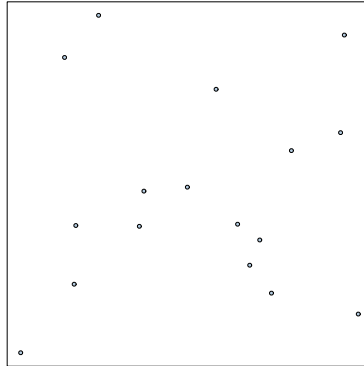
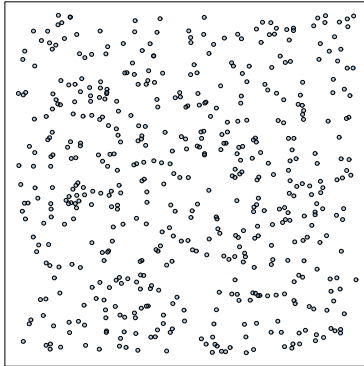
# Poisson Processes

**Poisson distribution is often appropriate for things that occur *randomly* but at a certain *constant rate*.**

- If you could repeat a census many times (either in the same location, or simultaneously in many similar locations) your data can be modeled with a Poisson distribution.
- We'll talk more about modeling counts with the Poisson, binomial, Bernoulli, and other discrete distributions when we talk about extending the simple linear model later in the course.

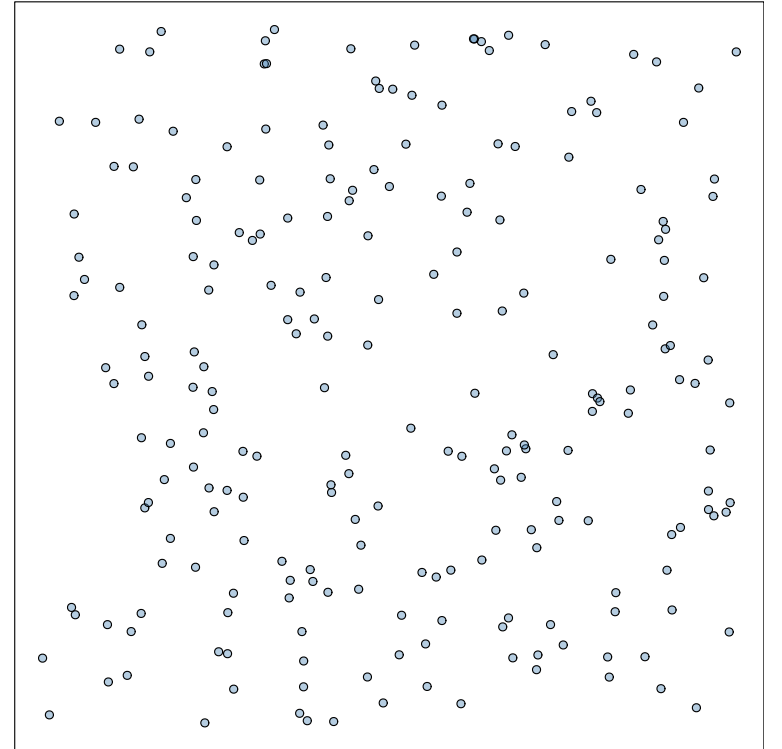
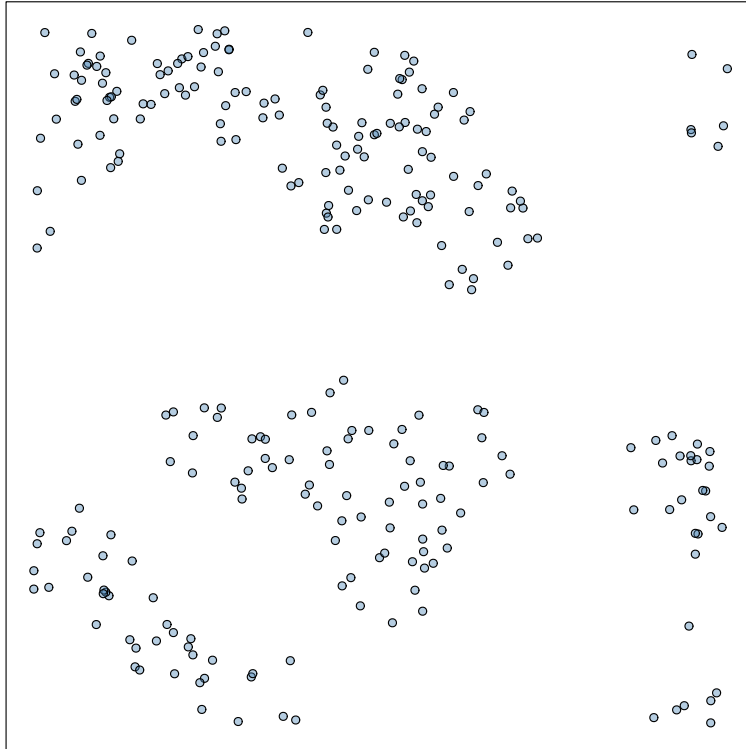
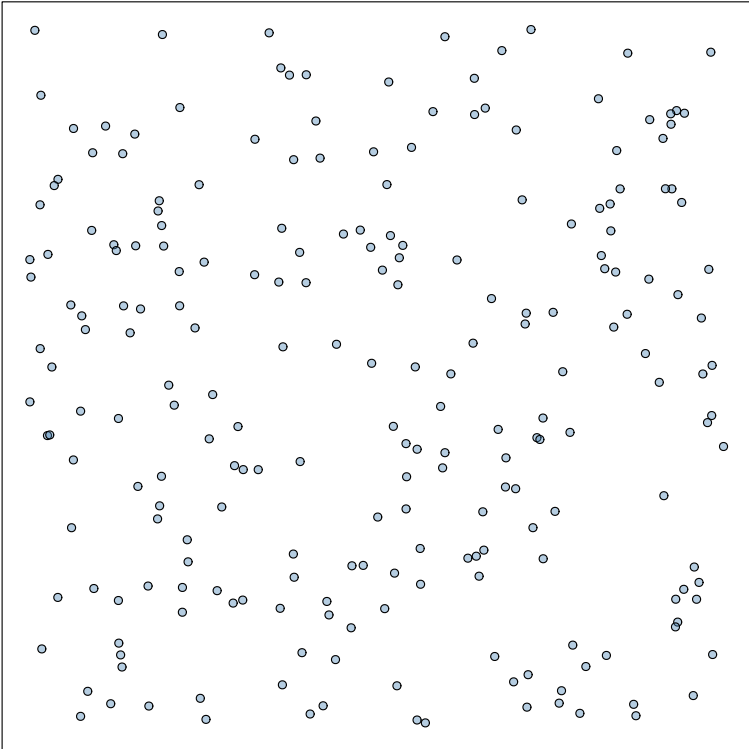
# The Poisson Distribution

The Poisson distribution is very important *null model* in spatial statistics



# The Poisson Distribution

Completely Spatially Random (CSR) point patterns follow a Poisson Distribution – It is a great model for **point processes**.



# Sampling with Replacement

- The Hypergeometric Distribution models sampling with replacement:
- Binary outcomes, traditionally described using balls of 2 colors in an urn.
  - But acorns in a bag work too!
- Fixed number of trials: the total number of acorns removed from the bag.
- The number of red and brown acorns in the bag is fixed, but you might not be able to know ahead of time.
- Sampling without replacement: removal of the first red acorn changes the proportion of red and brown acorns remaining in the bag.

# Sampling With and Without Replacement

## With Replacement: Binomial

- Used to infer  $\text{Pr}(\text{Success})$
- Fixed number of trials
- Trials are independent
  - Because the acorns are replaced after each trial.
- $\text{Pr}(\text{Success})$  is constant for each trial
- Can be used instead of hypergeometric if  $N$  is large and  $\text{Pr}(\text{success})$  is small.

## Without Replacement: Hypergeometric

- Used to infer numbers of successes/failures
- Fixed number of trials
- Trials are non-independent
  - You keep the acorn, changing the proportions in the remaining acorns.
- $\text{Pr}(\text{Success})$  changes as you remove individuals from the sample.
- Useful when  $N$  is small, and trials are not independent



# Recap

## Important concepts

- What is a discrete distribution?
- Discrete sample spaces
- Combinations and permutations
- Bernoulli, Binomial, and Poisson distributions



# Combinations and Permutations

**Combinatorics studies the possible ways we can *arrange* sets of objects.**

- Is order important?
- do we consider  $(T, H)$  and  $(H, T)$  to be the same or different events.
- How many *categories* of objects are there?
- Sampling - to replace or not to replace?

**Combinatorics is the key to understanding probability in discrete distributions.**

# Combinations and Permutations

## Does order matter?

- If we care about order we consider  $(T, H)$  and  $(H, T)$  to be *distinct* events.
- If we do not care about order  $(T, H)$  and  $(H, T)$  are *equivalent* events.

**When order is important, we work with *permutations*.**

**When order is unimportant, we have *combinations*.**

- There are usually more permutations than combinations.

# Si ektywqzuxim qldebxow\*

**\*An interesting question. I looked for 'an interesting question' in the library, but the best I could find was three successive words with the right number of letters!**

**Does the spelling of words concern *combinations* or *permutations*?**

- What can we learn from the Library of Babel?

# Combinations, Permutations, and Cards

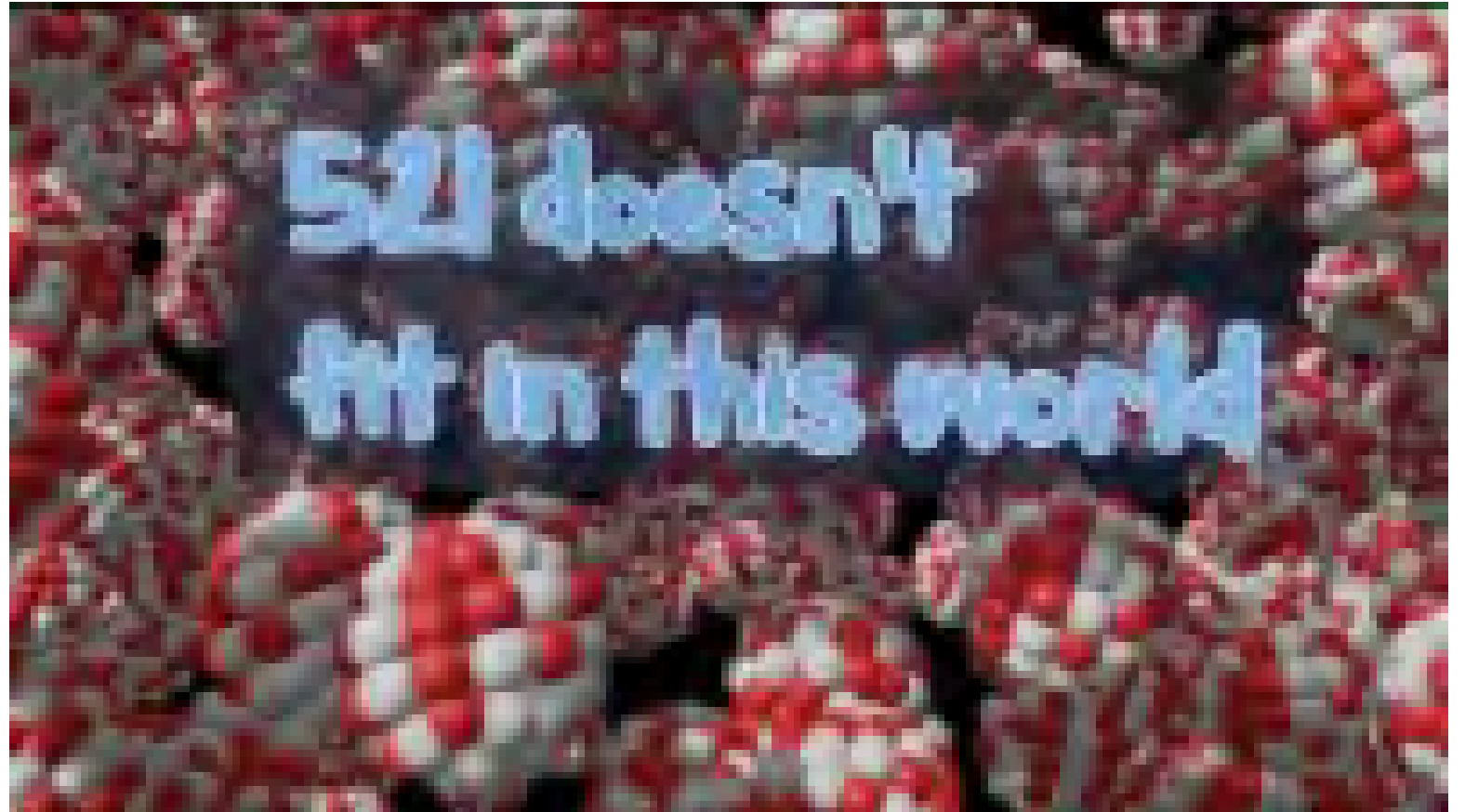
- How many combinations are there for a deck of 52 cards?
  - A trick question, there is only one combination since order doesn't matter!
- How many permutations are there for a deck of 52 cards?
  - Let's calculate:
    - There are 52 possibilities for card 1
    - 51 possibilities for card 2
    - 50 for card 3
    - And so on...
    - That's  $52 * 51 * 50 * \dots * 3 * 2 * 1$
  - There are  $52!$ , that is 52 factorial, permutations of a deck of cards. That's a huge number
- The card deck permutation question is like sampling without replacement.

# Combinations, Permutations, and Four-Letter Words

- How many four-letter words are possible using the English alphabet?
  - Let's calculate:
    - There are 26 possibilities for letter 1
    - 26 for letter 2
    - 26 for letter 3, etc.
  - $26^4 = 456976$
- How many five-letter words are there?
  - By what factor does the number increase for each extra letter?
- The word question is like sampling with replacement.

# How big is 52!

- It doesn't fit in this world!



# For Next Tuesday

- Make sure you've read the acorn/probability questions. We'll work on them in class using real acorns.



➤ Slide Deck 4 Issues – There were many missing and/or duplicated slides.

- Deck is fixed, please obtain corrected pdf version from GitHub.

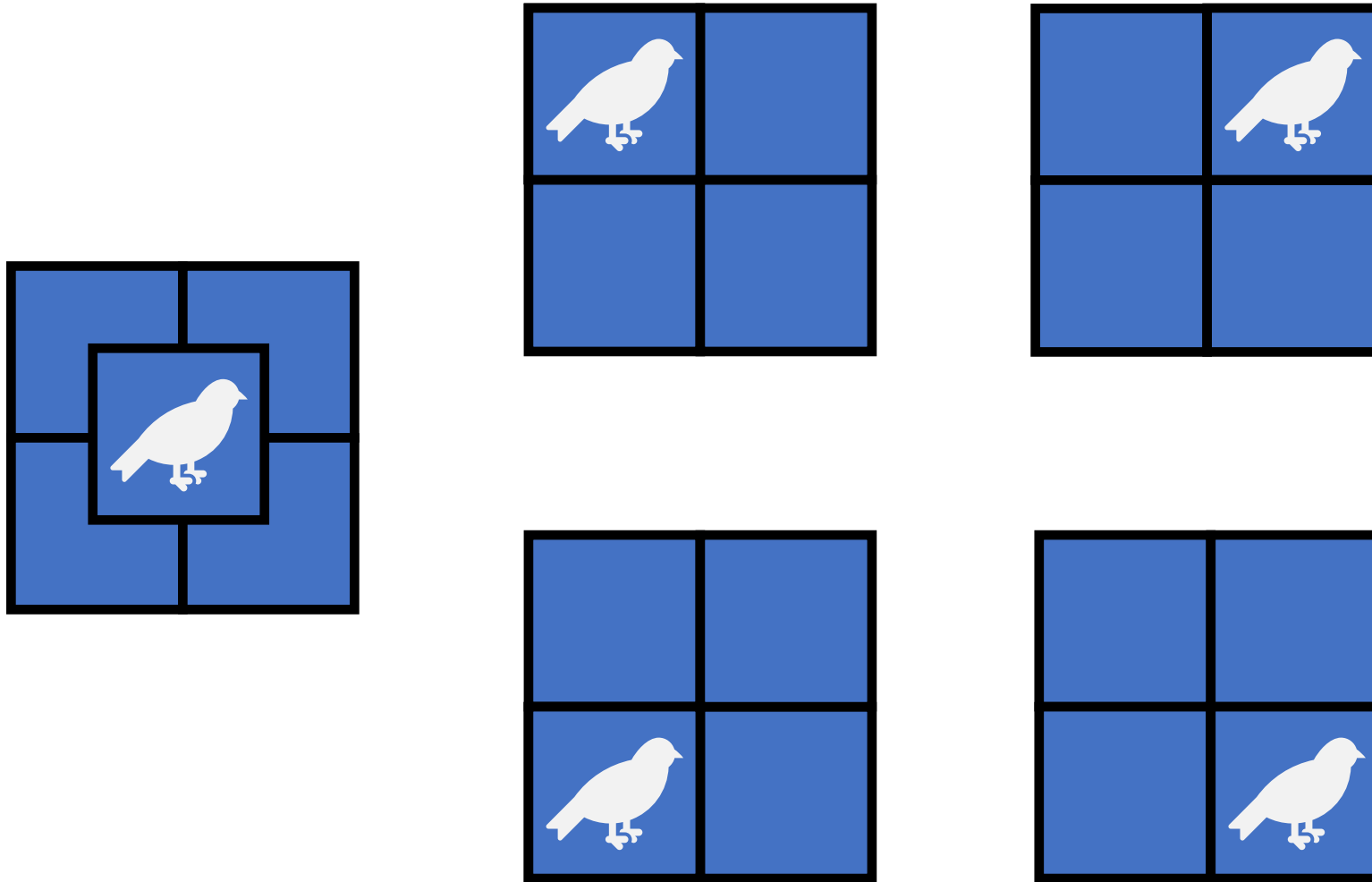
➤ Group Assignments

- Make sure all group members have access to common R code
- We'll be starting a new RNotebook assignment – you'll need all of your group code!

# Tuesday Oct 5

- For today
  - Finish this deck:
    - Combinations, permutations, sample spaces, events, and probabilities
    - Distribution functions
    - Continuous distributions
  - Start in-class probability/distributions activity
- For Thursday
  - Start deck 5: Frequentist concepts
  - Finish in-class probability/distributions activity
  - Start in-class probability calculations activity (if time allows)

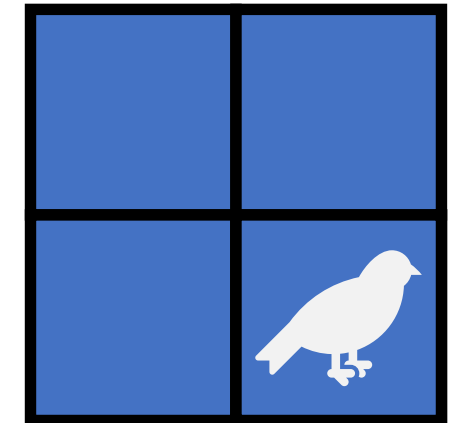
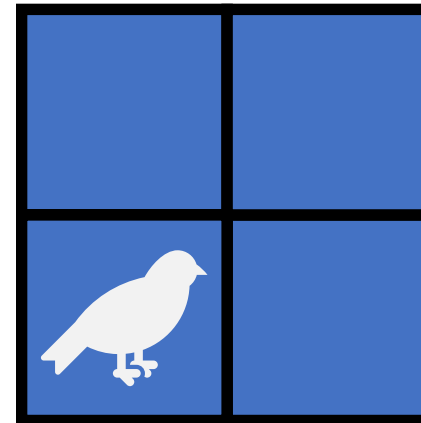
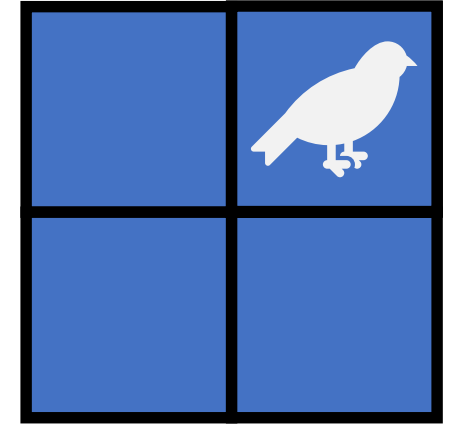
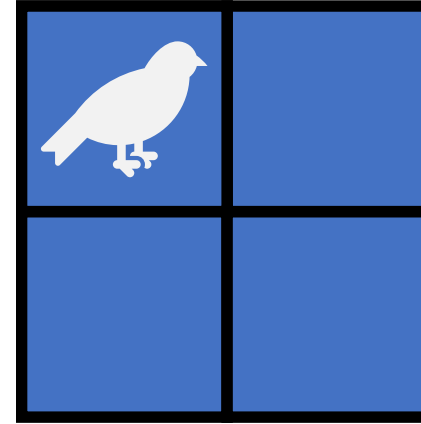
# Events, Probabilities, Combinations, and Permutations



# Events, Probabilities, Combinations, and Permutations

We generally use **combinations** to enumerate **events** and populate **sample spaces** when **order** or arrangement **doesn't matter**.

We need **permutations** to figure out **probabilities**. We use permutations to figure out events and sample spaces when **order or arrangement is important**.



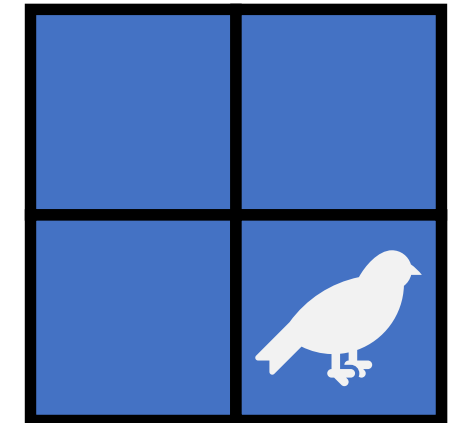
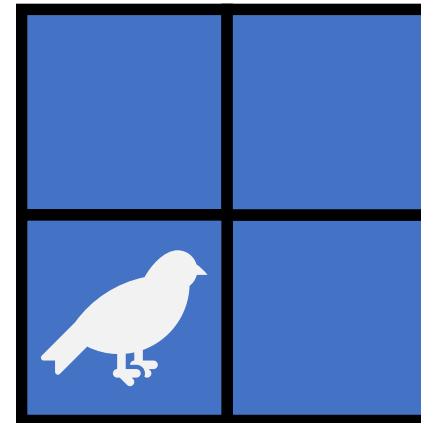
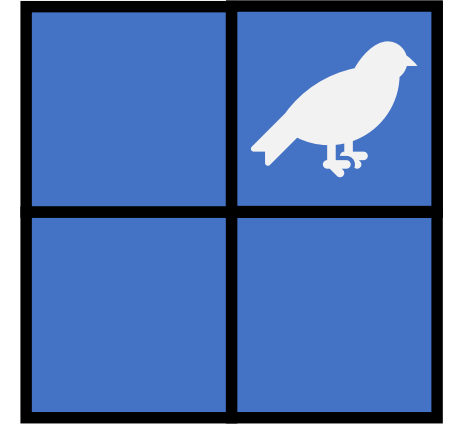
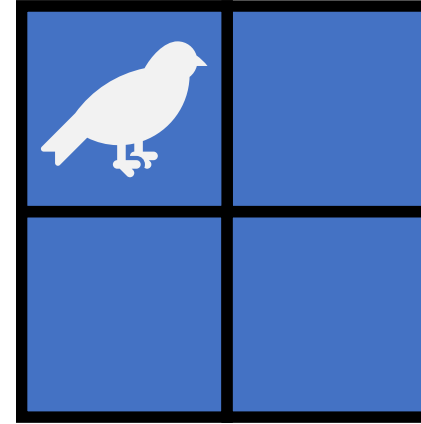
# Events, Probabilities, Combinations, and Permutations

We generally use **combinations** to enumerate **events**.

- How many sites had bird presences?

We need **permutations** to figure out **probabilities**.

- How many ways could a single site have a bird presence?

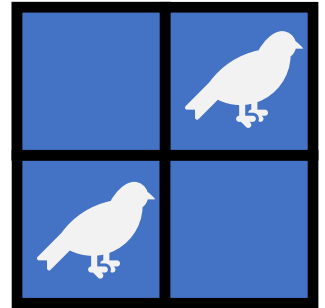
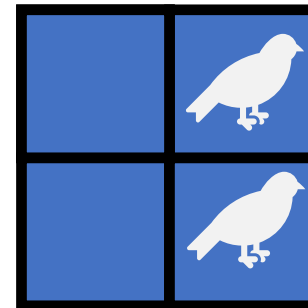
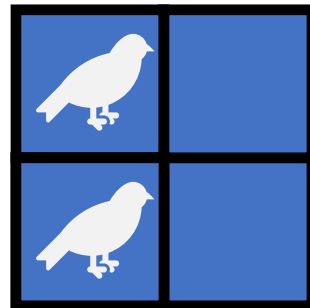
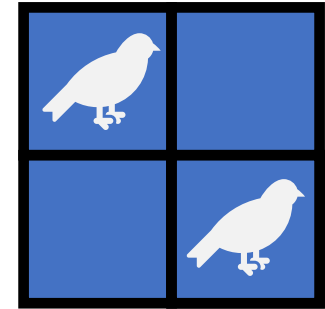
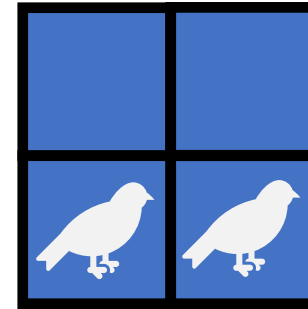
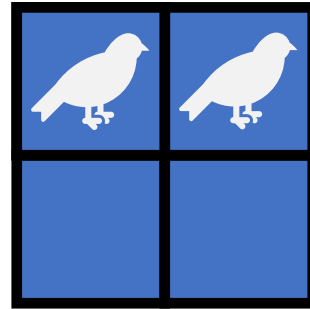


# One Combination, Six Permutations

## Combinations

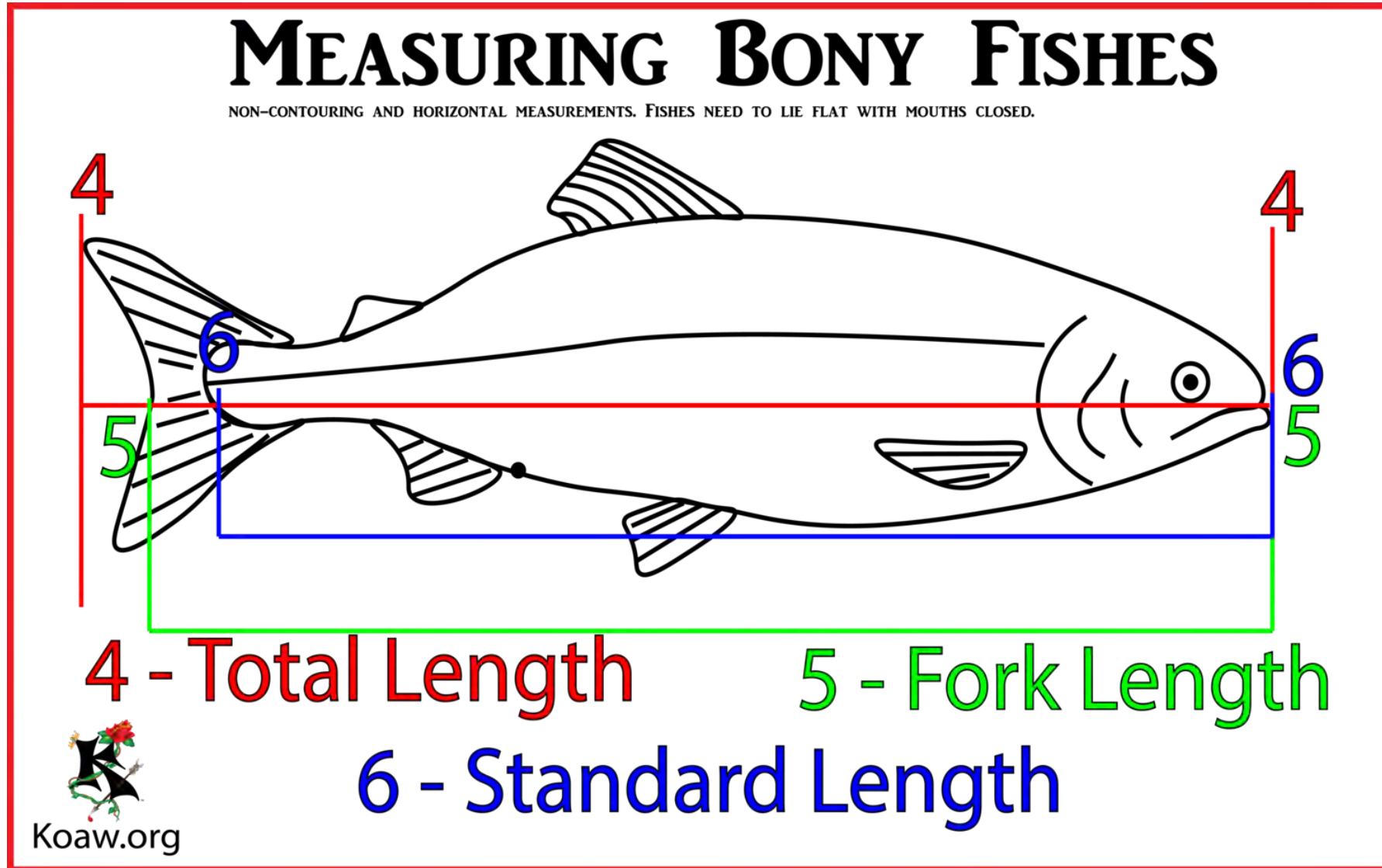
- One event: Two sites with bird presence
- The event can be permuted six ways!

## Permutations



# Probability Distribution Functions

# Probability Distribution Functions





# What questions should a distribution function answer?

- Am I more likely to observe a fish that is 20cm, or a fish that is 11cm?
  - Probability Density Function: relative likelihood of  $x$
- What is the probability that a fish is longer than 20cm?
  - Likelihood of  $x$  or smaller: cumulative Density Function
- How long is a fish in the 90th percentile?
  - Quantile Function

# Probability Density/Mass Functions

## **Probability density or mass functions answer the questions:**

- Am I more likely to catch a fish that measures 6cm or 14.5cm?
- What is the probability that I collect *exactly* two acorns of Red Oak out of a mixture of Red Oak and Bur Oak acorns?

## **They associate an event with a measure of likelihood**

- This is the probability of the event for discrete distributions
- For continuous distributions it is more complicated, but you can think of it as relative likelihood.

# Probability Density/Mass Functions

**They are maps of events in the sample space to probabilities.**

- **Probability Density Functions** for continuous distributions
- **Probability Mass Functions** for discrete distributions.

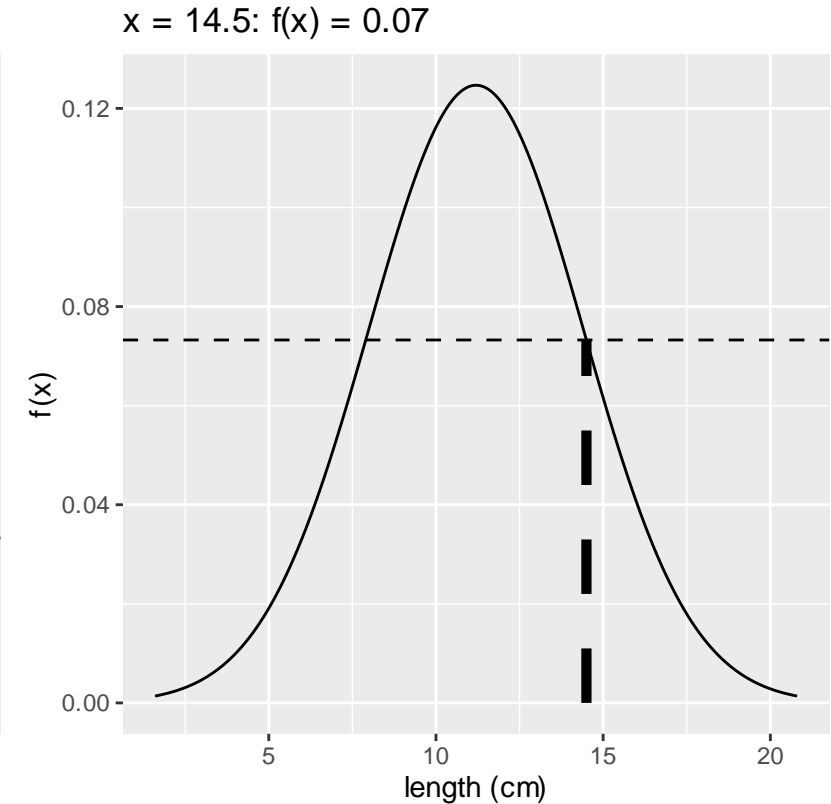
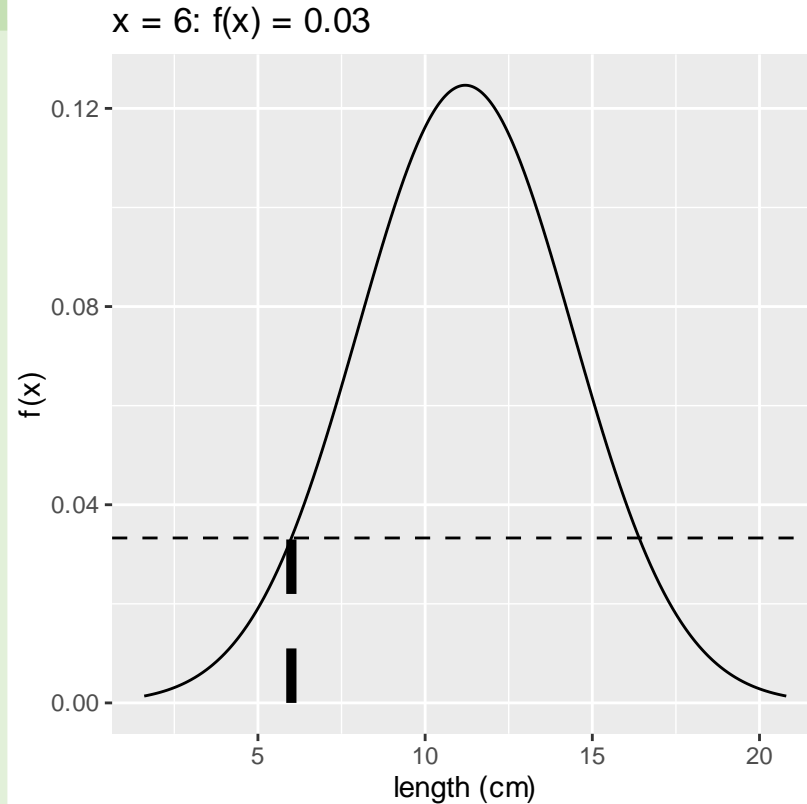
**Probabilities are always between zero and 1:**

- The values of PDFs and PMFs are always non-negative.

# Probability Density Function

**PDFs tell us about relative likelihoods**

**Am I more likely to catch a fish that measures 6cm or 14.5cm?**



# Cumulative Probability Functions: CDFs & CMFs

## The CDF/CMF answers:

- What is the probability that I catch a fish that weighs 153g or less?
- What is the probability that *at least* 3 of the acorns are Bur oak?

Cumulative density is the **accumulated area under the density curve** to the left of  $x$ .

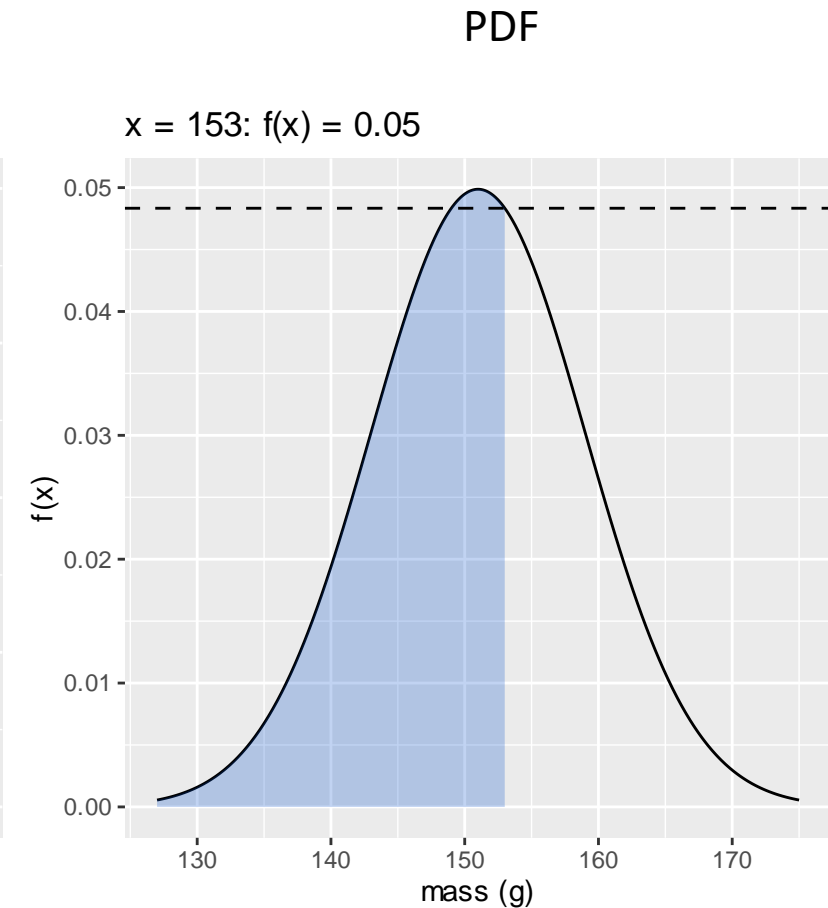
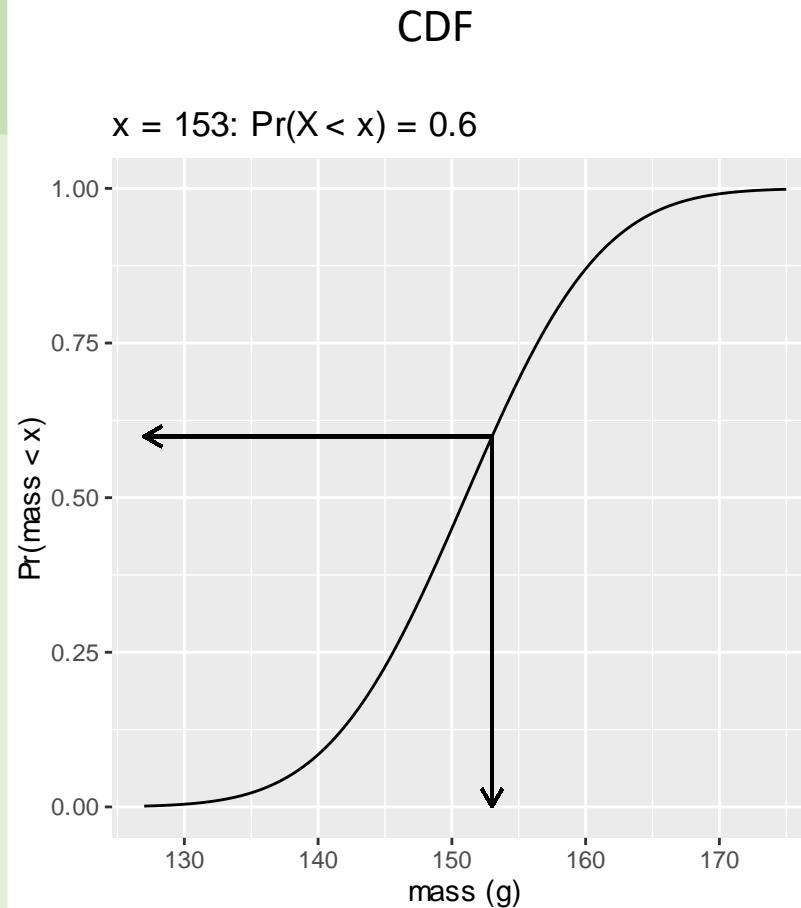
- It's an integral (or a sum for discrete distributions).
- It is the probability of observing a value equal to or less than  $x$ .
- The  $n$ th percentile (quantile).

# Cumulative Probability Functions: CDFs & CMFs

**CDFs tell us the probability of an event:**

**What is the probability that I catch a fish that weighs 153g or less?**

- Read the mass on the x-axis: 153g.
- Read the corresponding probability on the left: 60%

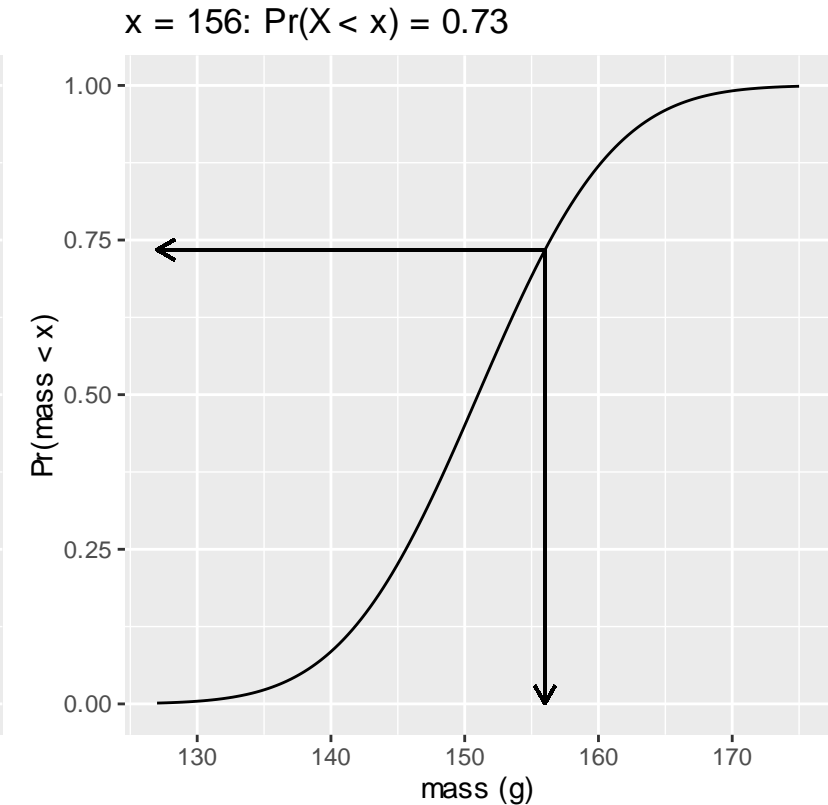
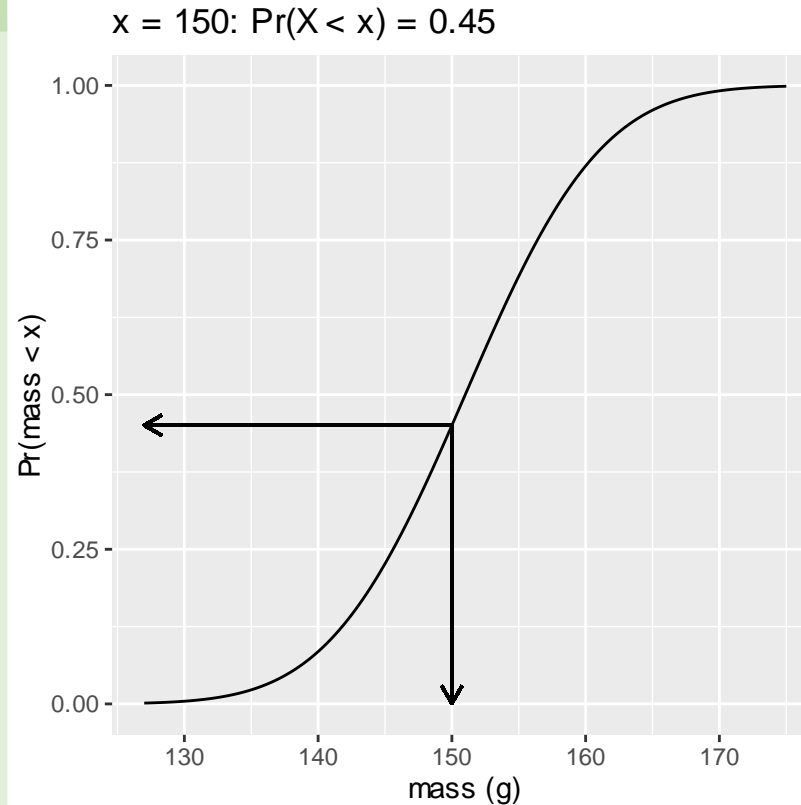


# Cumulative Probability Functions: CDFs & CMFs

**CDFs tell us the probability of an event:**

**What is the probability that I catch a fish that weighs between 150g and 156g?**

**- Take the difference of probabilities from the CDF:  $0.73 - 0.45 = 28\%$**

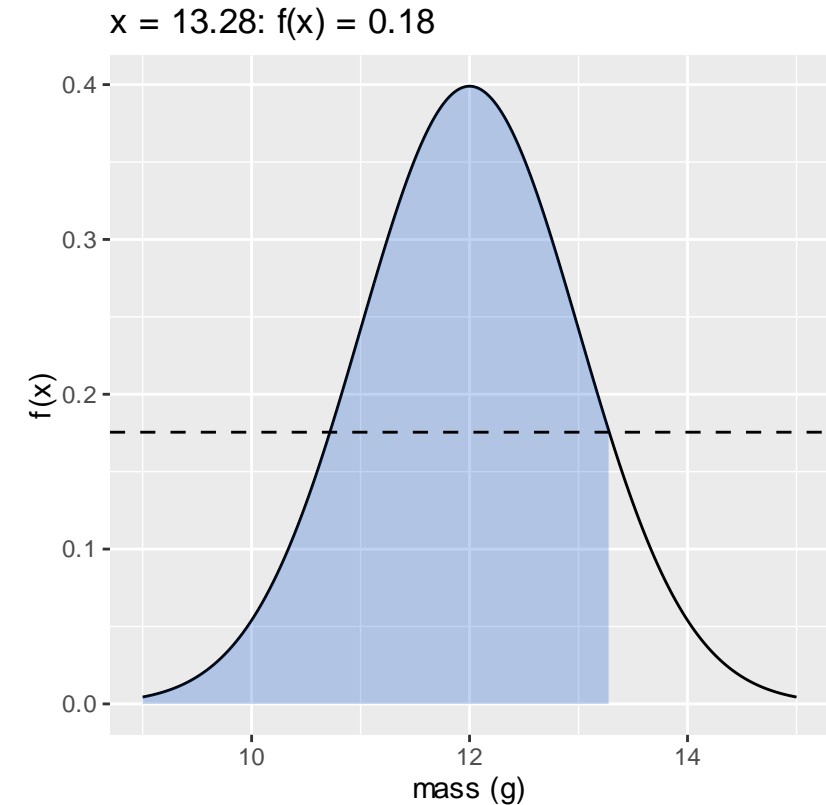
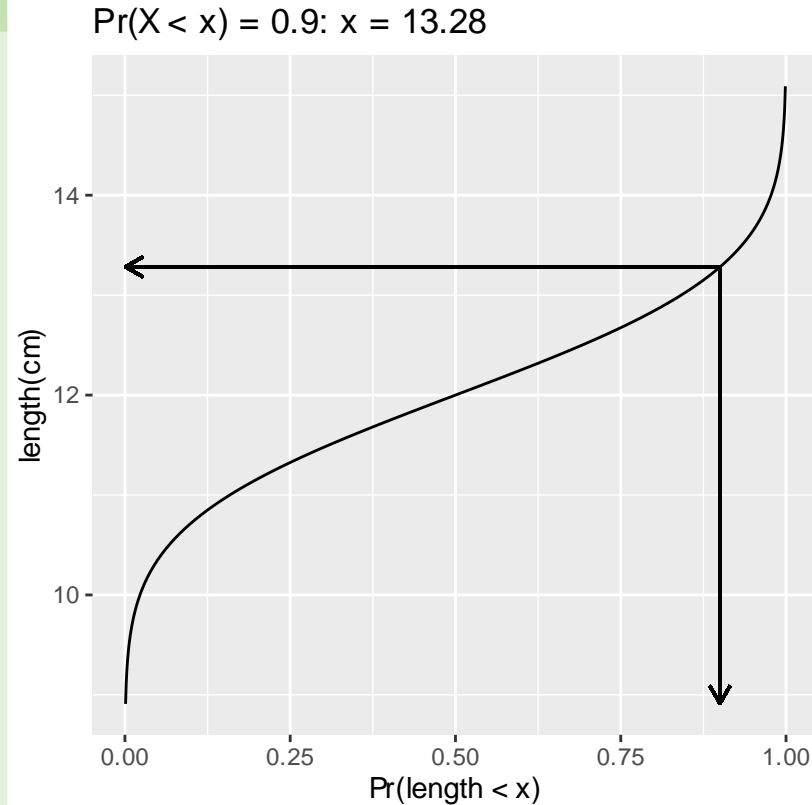


# Quantile Functions

Quantile functions tell us about percentiles:

**What length will 90% of all fishes will be shorter than?**

- Read the percentile on the x-axis.
- Read the size on the y-axis.



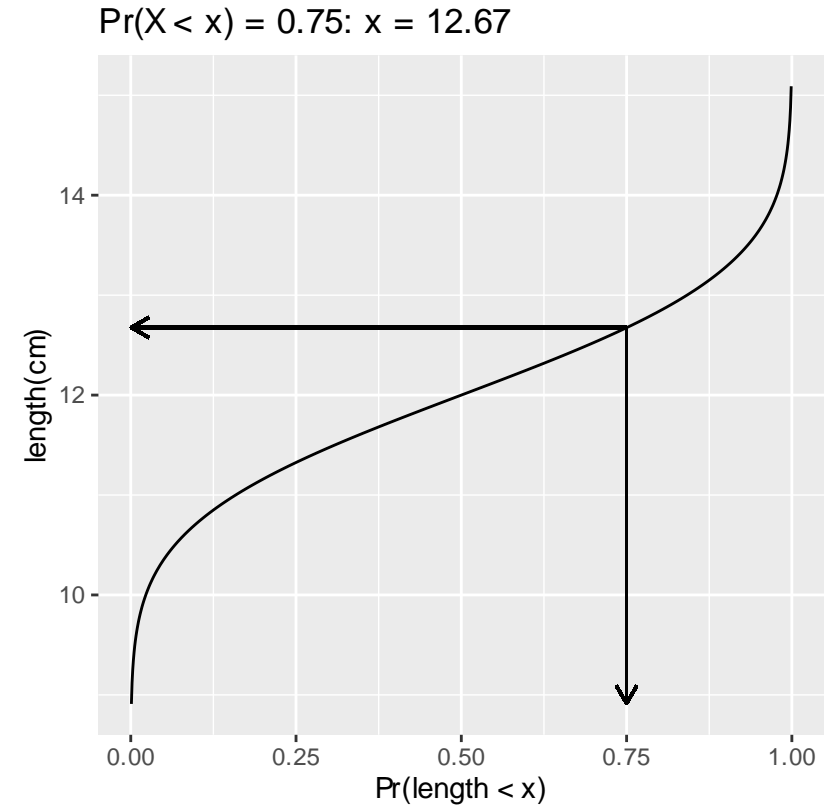
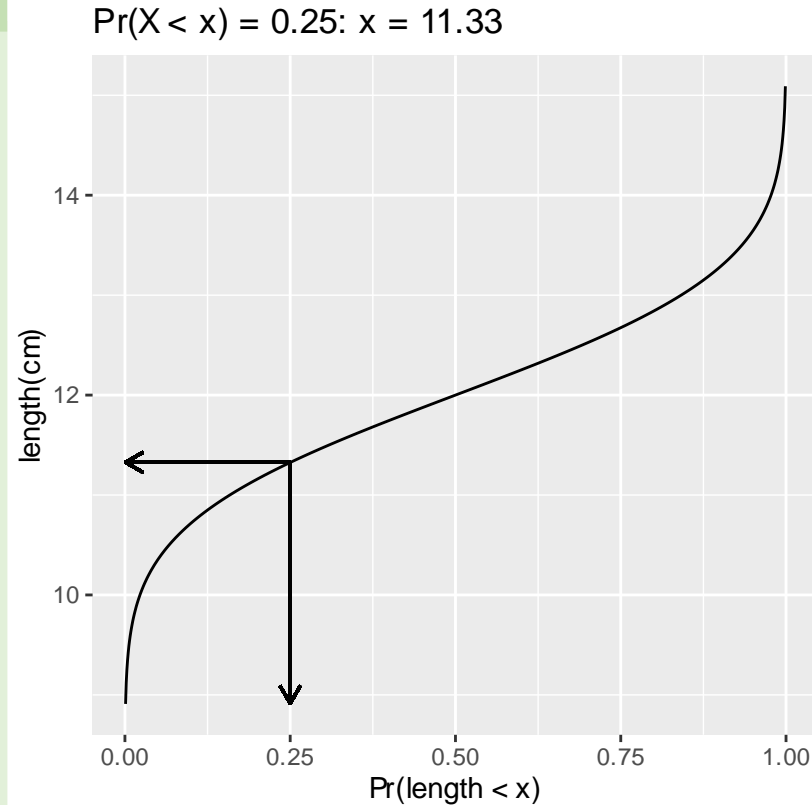


# Quantile Functions

Quantile functions tell us about percentiles:

What lengths span the middle 50% of the range?

- Read the percentiles on the x-axis.
- Read the sizes on the y-axis: 11.3cm – 12.7cm



# Parametric (Theoretical) Distributions

**Parametric distributions are defined by mathematical *functions***

- The functions have one or more *parameters* that define how probabilities are allocated to events.
- We often want to estimate the parameters from samples.

**The *binomial distribution* has two parameters:  $n, p$ .**

**The *Poisson* distribution has only one parameter:  $\lambda$**

# Empirical Distributions

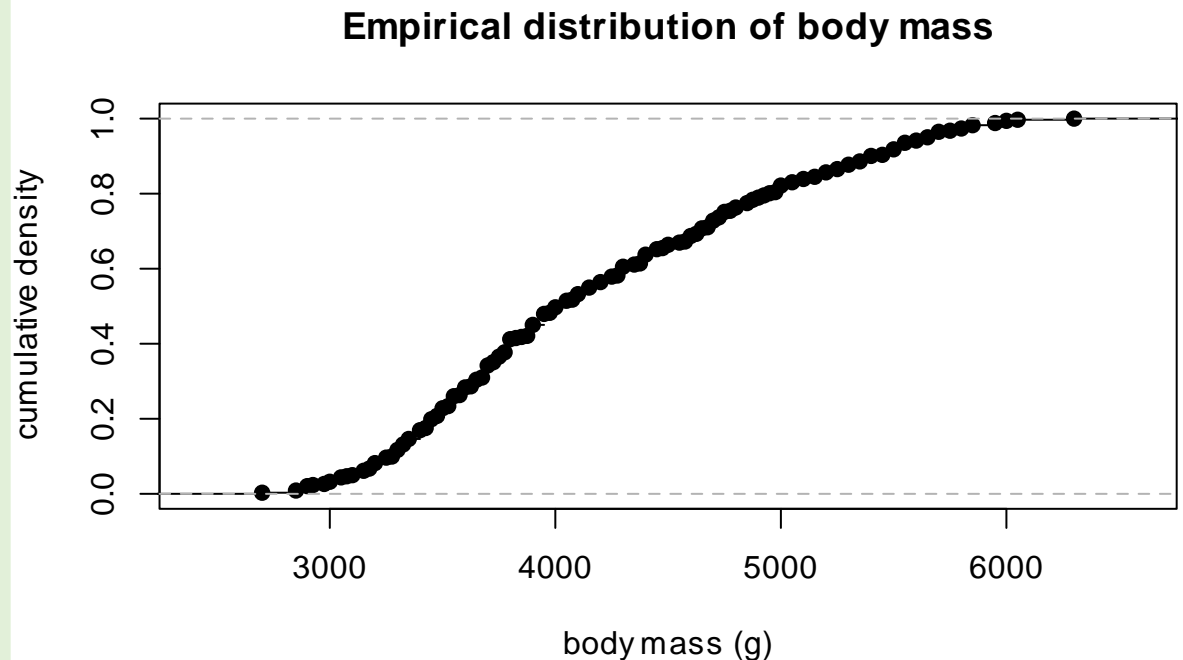
**Empirical distributions are computed from *observations*.**

- There is no analytical function: the shape is computed from data.
- We can compare empirical distributions to parametric distributions.

**Histograms are analogous to a PDF/PMF**



**Empirical cumulative distribution function are analogous to the CDF/CMF**



# Recap

- Theoretical and empirical distributions
- Parameters
- Distribution functions
  - Probability Density/Mass
  - Cumulative Density/Mass
  - Quantile Functions

# Key concepts

- Continuous sample spaces
- Normal distribution
- Exponential distribution
- T distributions

# Continuous Sample Space

## **Continuous distributions' sample space: the *real* numbers**

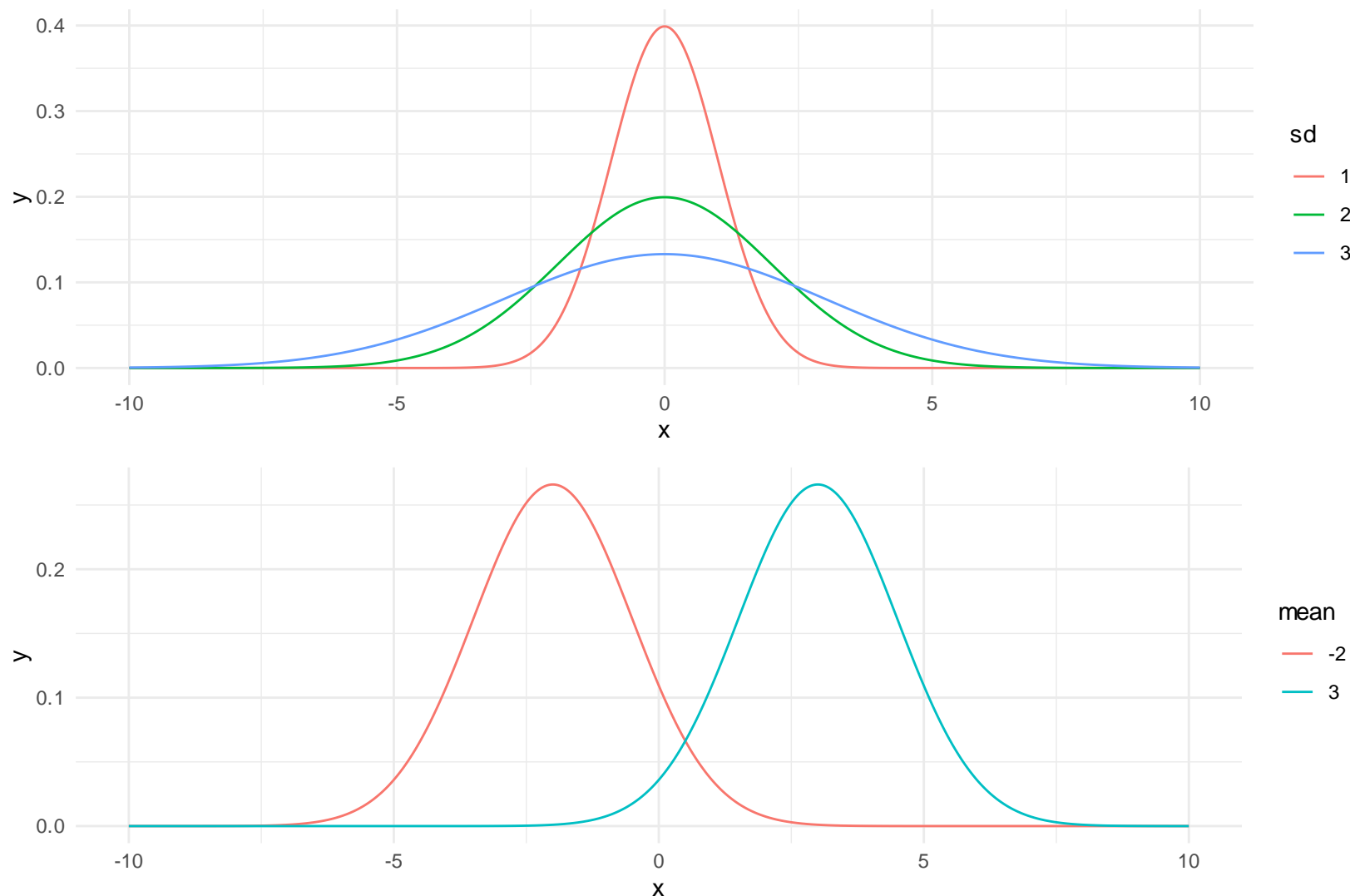
- All continuous distributions have *infinite* sample spaces.
- Continuous distributions may have *bounded* or *unbounded* sample spaces.

**PDFs are *continuous functions*.**

# Normal Distribution

The normal distribution has 2 parameters:  $\mu$  and  $\sigma$

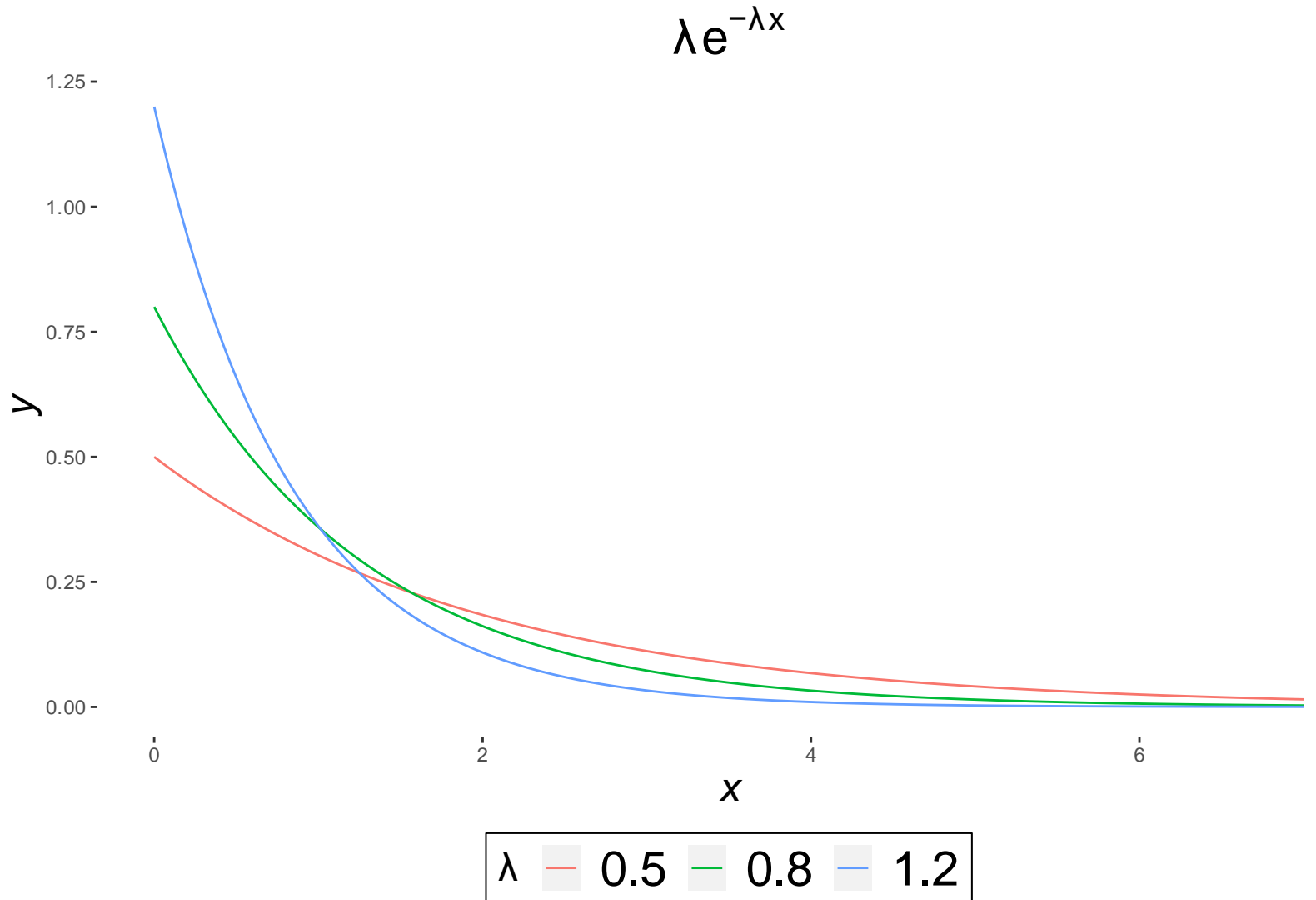
- The Standard Normal distribution has  $\mu = 0$  and  $\sigma = 1$
- The mean,  $\mu$  moves the curve left or right.
- The standard deviation  $\sigma$  controls the width.



# Exponential Distribution

**Exponential distribution models exponential decay.**

- Small observations are common, large observations are rare.

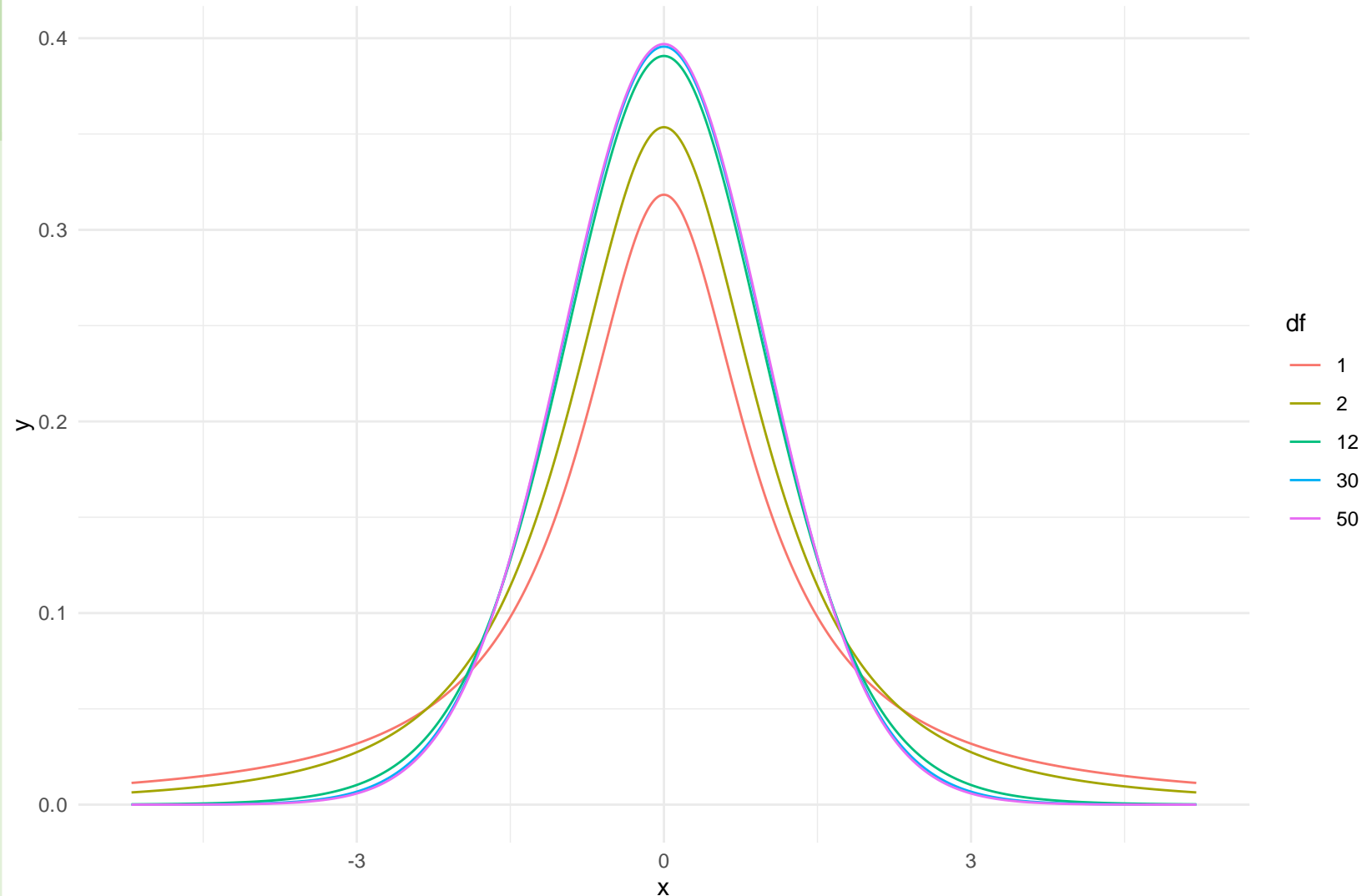




# The T Distribution

The t-distributions are like a sample-size adjusted version of the standard Normal

- The adjustment is via the *degrees of freedom* parameter
- As  $df \rightarrow \infty$  the t-distribution approaches the standard Normal



# Skew and Kurtosis

## Skew and Kurtosis are *higher-order* moments of distributions

- Mean is the 1st moment, variance is the 2nd moment
- Skew is a measure of asymmetry
- kurtosis is a measure of *pointiness*
  - Platykurtotic: flat with short tails, extreme events are less common.
  - Leptokurtotic: pointy with long tails, extreme events are more common.
- Skew and kurtosis are measured in reference to a normal distribution

# Kurtosis

- **Platykurtotic = too flat**
- **Leptokurtotic = too pointy**



# Key concepts

- Continuous sample spaces
- Normal distribution
- Exponential distribution
- T distributions