ECO 602
Analysis of Environmental Data

FALL 2019 – UNIVERSITY OF MASSACHUSETTS

DR. MICHAEL NELSON

Today's Agenda

Some answers and clarifications

Likelihood

Group quiz/activity: Likelihood Functions

Conditional Probabilities

Bayesian Intro

Discrete Bayesian Inference Example

Correction about Student’s t-test

I mistakenly said that R. A. Fischer first published the t-distribution. It was William Sealy GossettP

“The t distributions were discovered by William S. Gosset in 1908. Gosset was a statistician employed by the Guinness brewing company which had stipulated that he not publish under his own name. He therefore wrote under the pen name ``Student.“”

http://statweb.stanford.edu/~naras/jsm/TDensity/TDensity.html

Clarifications about Bootstrapping from last lecture

Why don’t we always use bootstrapping (or other resampling techniques), instead of estimating parameters?

What are the pitfalls of bootstrapping?

What are advantages of bootstrapping?

Bootstrapping uses

Estimate standard errors. Remember SE is a parameter of the **sampling distribution** of a **statistic**, not the population distribution.

Estimate **confidence intervals.**

Helpful when we don’t want to assume a theoretical distribution of the population, or when we don’t want to rely on the Central Limit Theorem.

Bootstrap pitfalls

Bootstrap sampling distributions tend to be too narrow: narrowness bias.

Bootstrap distributions won’t fix **nonrepresentative** or too-**small** samples.

Bootstrap estimates of the **median** can be problematic.

May be computationally intensive.

Bootstrap advantages

Conceptually simple, easy to implement, may be more intuitive than formulas for calculating standard errors:

- **SE of the mean** calculation is simple; SE of other statistics are **much** more complicated.
- Formulas for >1 predictor or response can be very complicated!

Bootstrap advantages

Can be used to illustrate concrete examples of theoretical principles:

- Bootstrapping is a good way to visualize theoretical sampling distributions

Likelihood: The scenario

Main question: How **likely** am I to have **observed** the data I collected under my **proposed** model?

Likelihood can help if you have:

1. Data
2. A proposed a distribution or model of the data
3. A set of candidate distribution/model parameters

Likelihood: independent samples

Since you are a whiz at designing experiments, you know that all of your samples are **independent!**

What do we already know about the joint **probability** of **multiple**, **independent** events?

Likelihood: independent samples

The **joint probability** of observing multiple independent events is the **product** of the probabilities of the **individual** events.

Likelihood is an estimate of how probable your **particular data** are given a **model** and a set of model **parameters**.

The **likelihood** is the **product** of the **probabilities** of each observation given your model/parameters!

Likelihood: data and model

How do we calculate the probability for a **specific event or observation** if we have a theoretical distribution?

What about when we don't have or can't use a theoretical distribution?

Likelihood: procedure

1. Collect data
2. Propose model and candidate parameter values
3. Calculate the probability density of each observation given your model and parameter values:
   - From a theoretical distribution.
   - From an empirical/resampled/simulated distribution

Likelihood: procedure

4. Multiply the densities.
   - In practice we calculate the logarithm of the densities and add them together.
   - Why might this be better than multiplying probabilties?
5. Voilà: your likelihood value for your data Y given your proposed model and parameter values: Φ_m
   - In symbols $L(\mathbf{Y} \mid \mathbf{\Phi}_\mathbf{m})$

Likelihood: maximizing

For inference, it might seem reasonable to try to find the parameter values that make our observed data most likely.

In Maximum Likelihood inference we want to **maximize the likelihood** of the parameters.

Likelihood calculations

Wouldn’t it be nice if we had a simple formula?

Sometimes we can find a formula and then find it’s minima/maxima via calculus.

Frequently such formulas don’t exist.

Likelihood visualization activity

Before we start, do you remember the **Bernoulli** distribution?

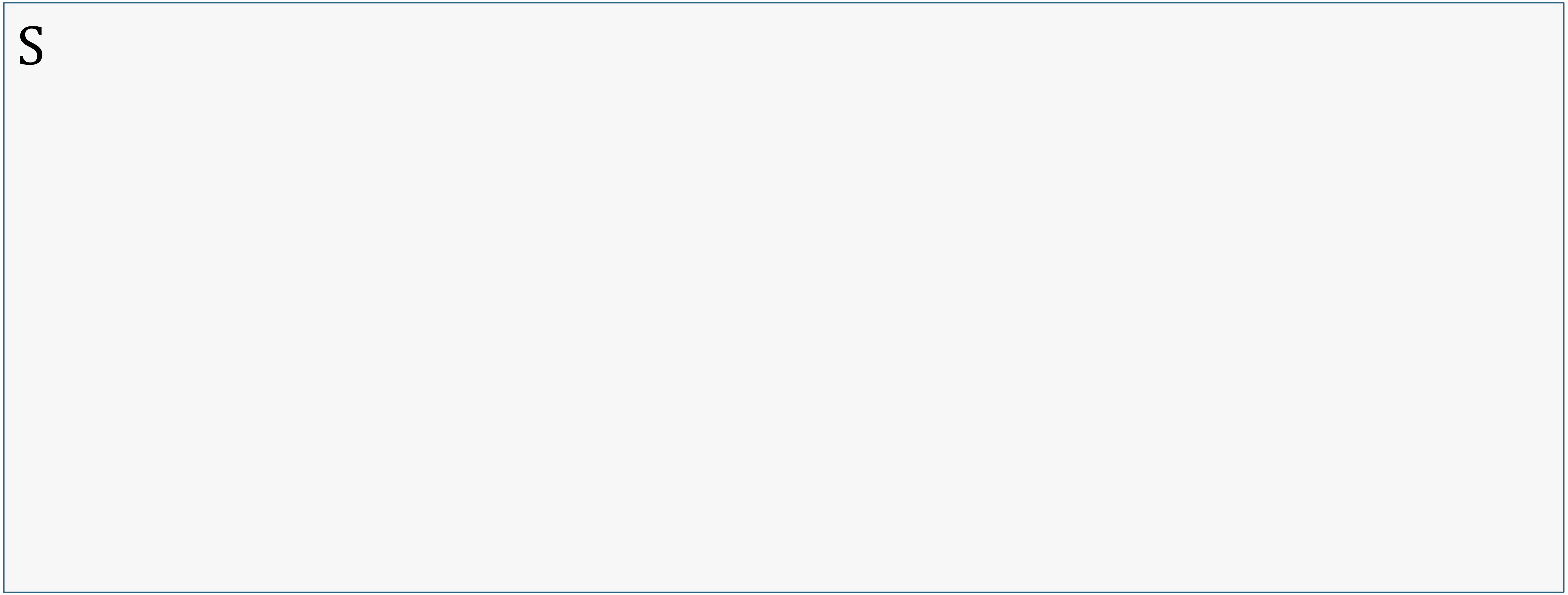
- What does it describe?
- What are its parameters?

Visit the [Seeing Theory likelihood page]

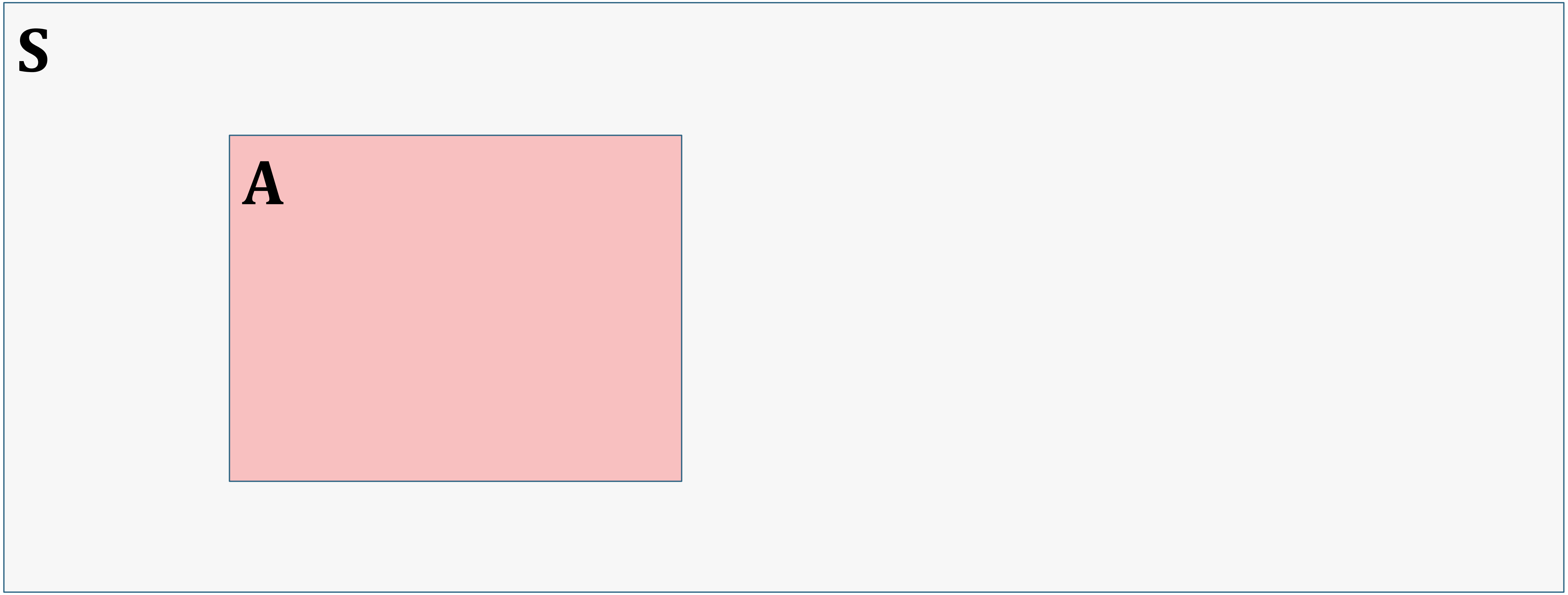
Instructions and link are on Moodle

Conditional Probability

Entire sample space: S

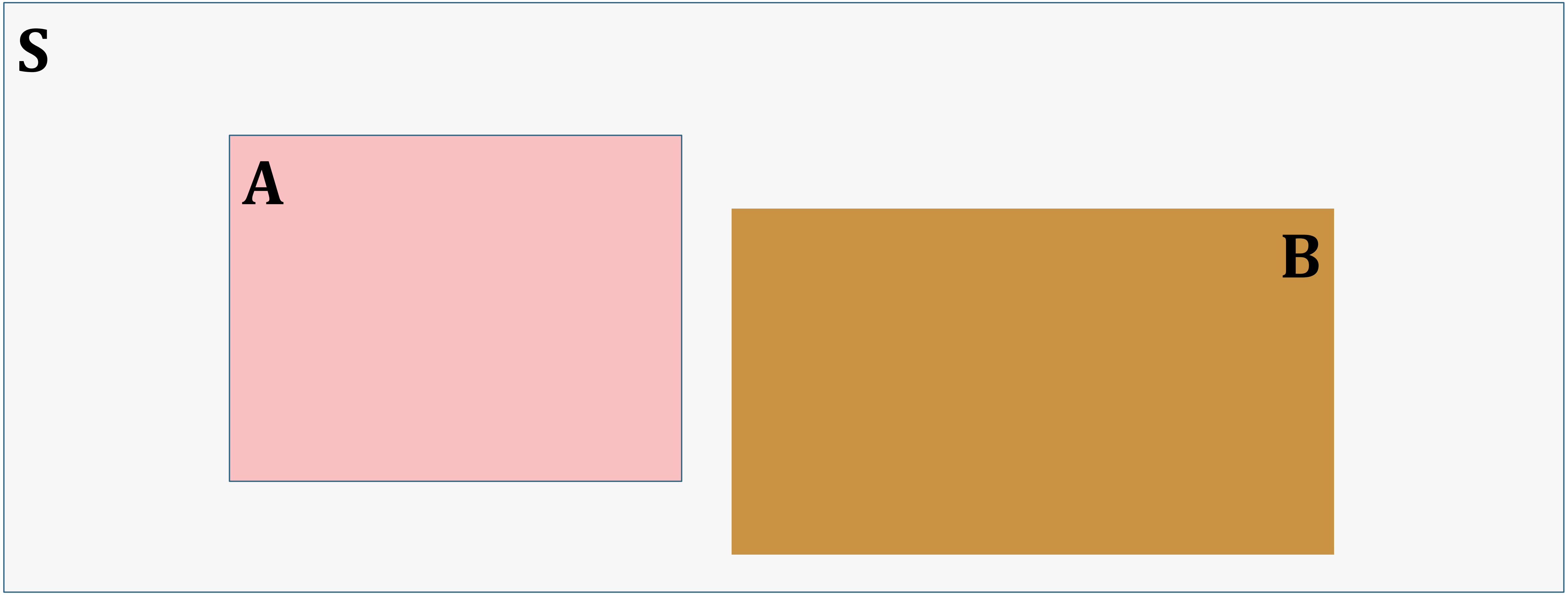


Event A

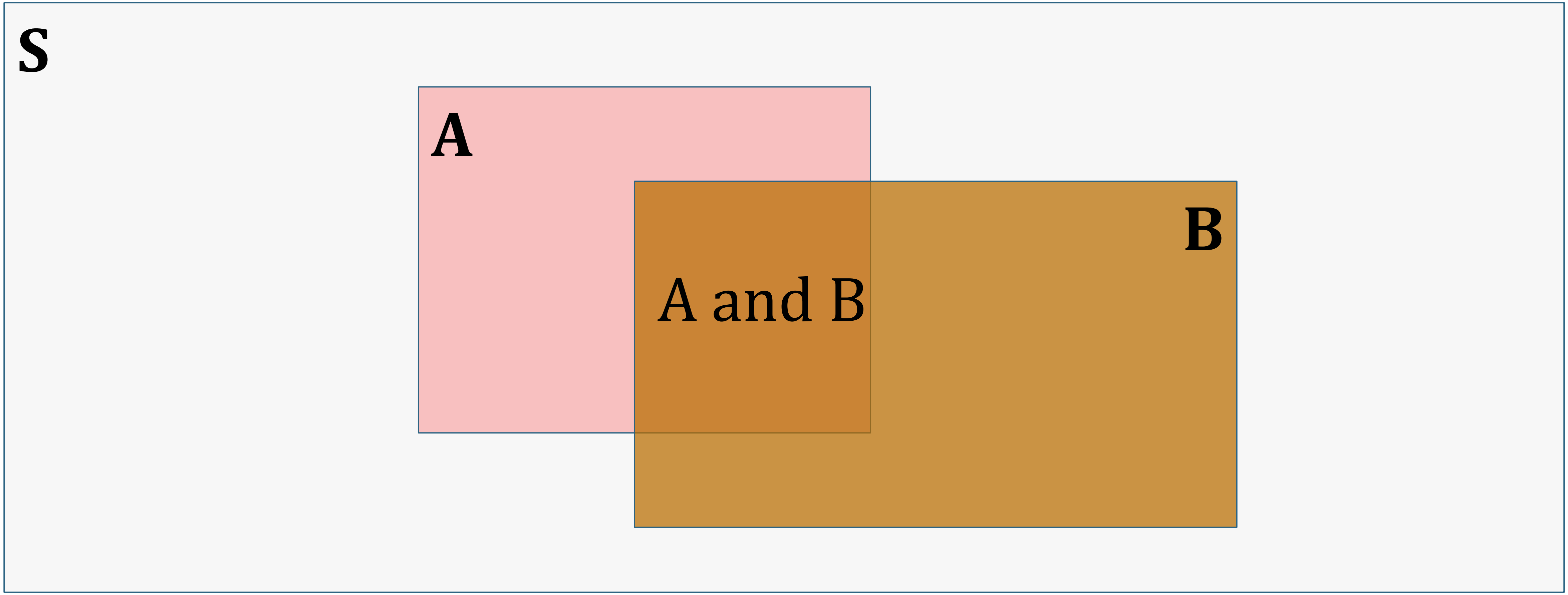


Event B

Both Events, exclusive

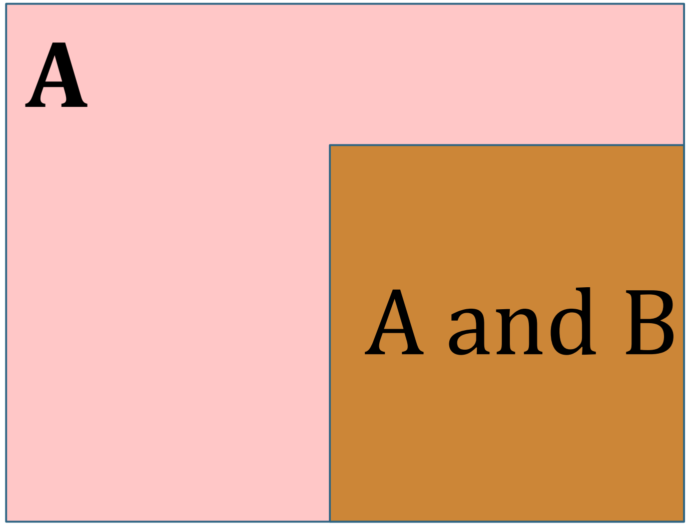


Both events, overlap



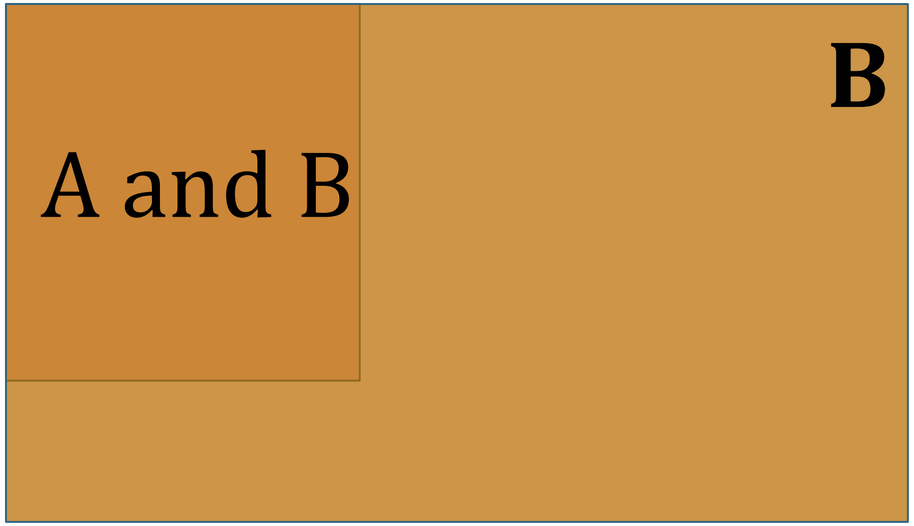
Pr(B | A)

- A has already happened
- Imagine S collapsing to A
- Pr(B | A) < Pr(A)
- Pr(B | A) = **Pr(A and B)** / Pr(A)
- Pr(B | A) Pr(A) = **Pr(A and B)**



Pr(A | B)

- B has already happened
- Imagine S collapsing to B
- Pr(A | B) < Pr(B)
- Pr(A | B) = **Pr(A and B)** / Pr(B)
- Pr(A | B) Pr(B) = **Pr(A and B)**



Common element in both conditionals: **Pr(A and B)**

Pr(A | B) Pr(B) = **Pr(A and B)**

Pr(B | A) Pr(A) = **Pr(A and B)**

This means:

Pr(A | B) Pr(B) = Pr(B | A) Pr(A)

Common element in both conditionals: **Pr(A and B)**

$Pr(A \mid B)\, Pr(B) = Pr(B \mid A)\, Pr(A)$

We can rearrange as 2 ratios:

$Pr(A) / Pr(B) = Pr(A \mid B) / Pr(B \mid A)$

Ratio of unconditioned probabilities is equal to ratio of the conditionals!

But what happens if there is no overlap?

What if Pr(A and B) is zero?

Conditionals become 0 and we have

Pr(A | B) / Pr(B | A) = 0 / 0

We can't do division by 0!

Bayes' rule to the rescue:

P(A | B) = P(B | A) P(A) / P(B)

Intro to Bayesian: Bayes' rule

Frequentist and Bayesian contrasts

Relationship between model and data

Frequentist:

- Data are one **realization** of a stochastic sampling process
- The **One True Model** exists and is unknowable

Bayesian:

- We know that our data exist, they **are not** random
- The model is a **random** variable that we will estimate from our fixed data

Uncertainty about the model

Frequentist:

- True model parameters are unknowable but **fixed**.
- Model parameters have no distributions, they simply exist!

Bayesian:

- Model parameters are random, they have **probability distributions** that we estimate from our **fixed data**.

Confidence and Credibility

Frequentist 95% **confidence** interval:

- We are confident that our **process** would produce intervals containing the true value 95% of the time.
- Certainty about whether a particular interval contains the true value is tricky.

Bayesian 95% **credible** interval:

- Given our data, we are 95% certain that our particular interval contains the **real parameter value**.

Inference: Frequentist

1. Estimate parameters that make our data most likely, under the assumption that they are one of infinite possible samples.
2. Express our parameter estimates in terms of a confidence intervals and p-values.
   - The CI either contains the param value or not. We can't know for a particular CI.

Inference: Bayesian

1. Estimate **probability distributions** of the parameters that are most likely given our data, and previous data/knowledge.
   - Conditional probability is key
2. Express our estimates in terms of credible intervals. P-values aren't as important.

Bayesian symbols and notation

Follows the format of conditional probability:

$Pr(A \mid B)$: What is the probability of A given that we know B occurred?

$Pr(H \mid D)$: What is the probability of our hypothesis (H) given that we have observed the data (D)

Bayesian symbols:

Hypothesis comprises our proposed model and a set of model parameter values

- Often denoted H or φ_m

Data comprises our **current** and **previous** data or knowledge

- Denoted D or **Y**

There are 4 important probabilities/distributions

Four important probabilities/distributions

1. Pr(**Y**): the probability or likelihood of our observed data
2. Pr(φ_m): The probability distribution of our model and parameters before data are observed
   - How could we possibly know this before we start?
   - Prior probability from previous data, maybe?

Four important probabilities/distributions

3. $\Pr(\mathbf{Y} \mid \varphi_m)$: Probability of observing the current data given our estimated model and the previous data.
   - Likelihood function of the model parameters: we want to maximize this function
4. $\Pr(\varphi_m \mid \mathbf{Y})$: Probability distribution of our estimated model parameters after the data are observed.
   - This is what we want to infer!
   - Posterior probability.

Bayesian: what do we need to proceed?

1. Pr(H): Prior unconditional distribution of the probability of our model params
2. Pr(D): Unconditional probability of observing the current data:
   - This is difficult…but we don't have to know it directly.
3. Pr(D | H): Conditional probability of observing our data given the model parameters.
   - Estimated is from the likelihood function.
   - Remember likelihood functions aren't trivial to find/define!

Discrete Bayesian Hypotheses

What if we have a set of exclusive and exhaustive hypotheses?

Bayesian Parametric Inference...

Estimate model parameters: Bayes' method

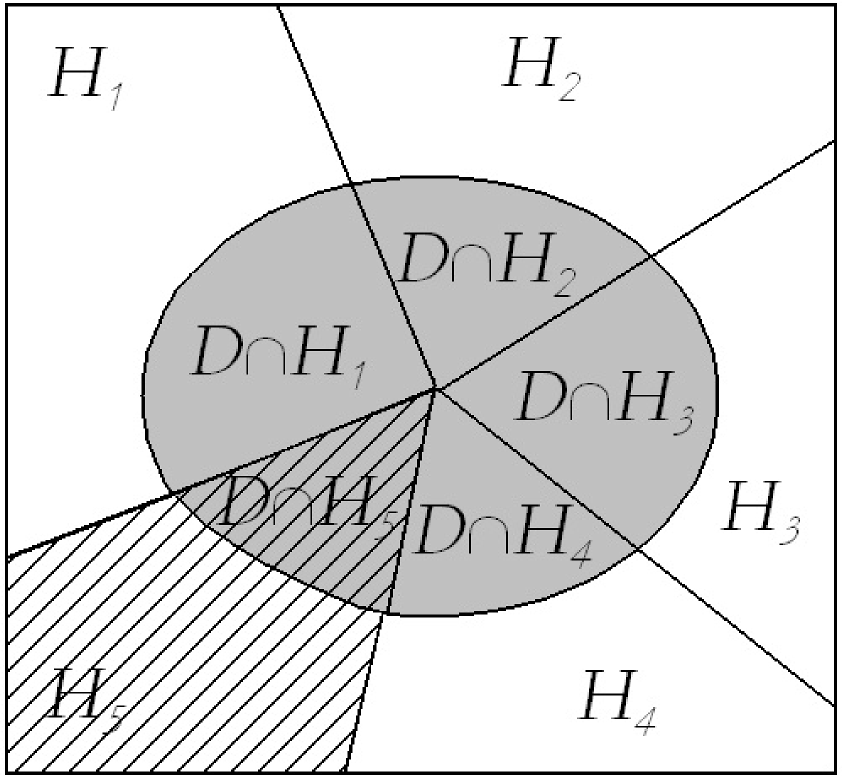
4. Major *technical* challenge to Bayes' Rule:

$$\Pr\{H|D\} = \frac{\Pr\{D|H\}\Pr\{H\}}{\boxed{\Pr\{D\}}}$$

(1) Discrete hypotheses:

$$\Pr\{D\} = \sum_{j=1}^{N} \Pr\{D \cap H_j\} = \sum_{j=1}^{N} \Pr\{D|H_j\}\Pr\{H_j\}$$

$$\Pr\{H_i|D\} = \frac{\Pr\{D|H_i\}\Pr\{H_i\}}{\sum_{j=1}^{N} \Pr\{D|H_j\}\Pr\{H_j\}}$$



From Bolker (2008)

For next time:

Continue with Bayesian.

Announce question set 4 and final projects.