ECo 602 - Analysis of Environmental Data

Intro to Bayesian Perspective and Conditional Probability

Michael France Nelson

Eco 602 – University of Massachusetts, Amherst
Michael France Nelson

Relationship between model and data

Frequentist	Bayesian

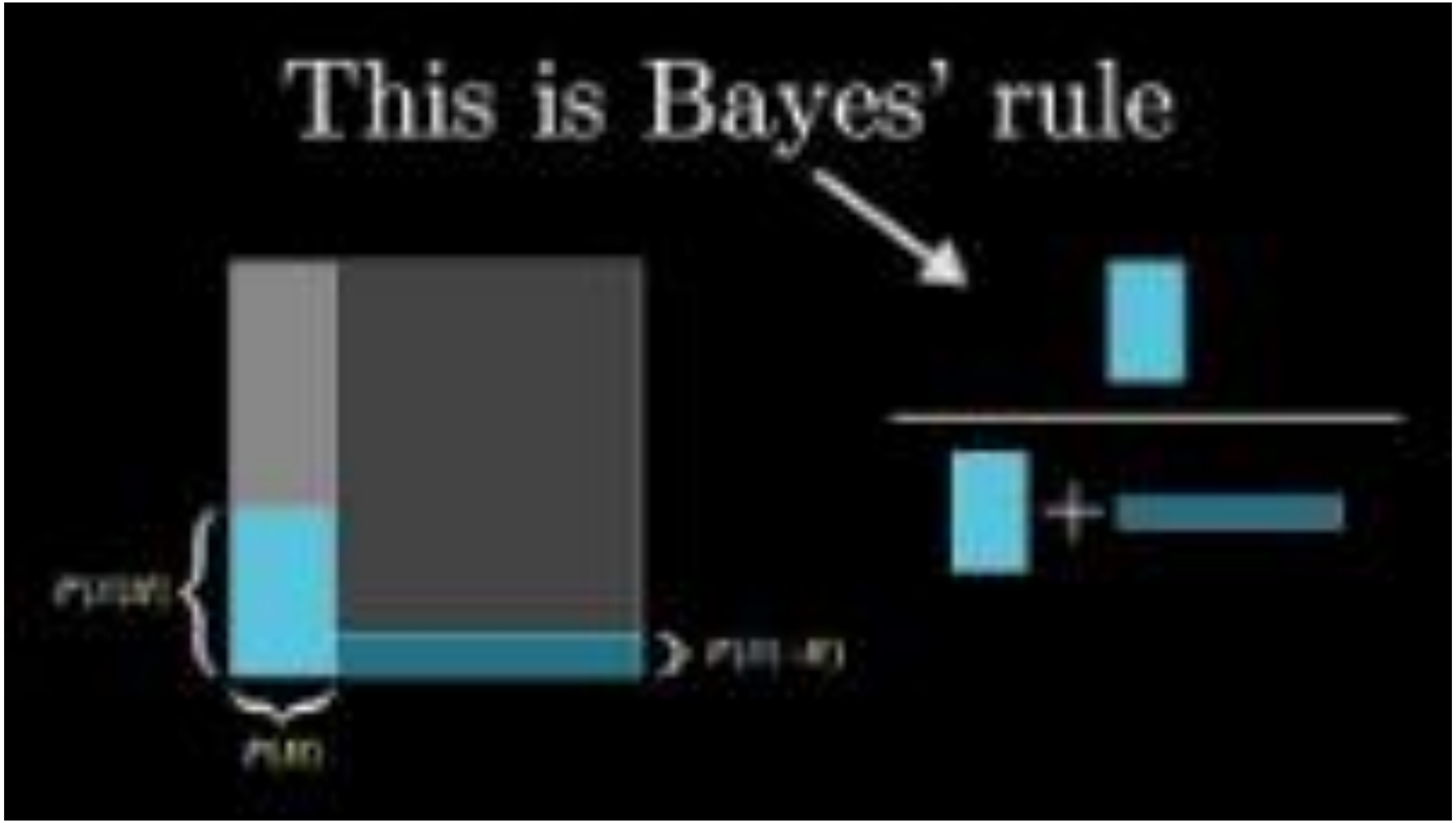
• Data are one realization of a stochastic sampling process.
• The One True Model exists and is unknowable.
    • The model is out there and we'll design a procedure that would approximate the real model if we could do many replications.
• Focus is on repeatedly sampling new data.
	• We know that our data exist, they are not random.
	• The model is a random variable that we will estimate from our fixed data.
	    • The model is out there and we'll use our data to put bounds on what we think it is.
	• Focus is on estimating the distributions of possible model parameter values from data.

Preview of Bayesian Thinking

https://youtu.be/7GgLSnQ48os

Another Cool Video

https://youtu.be/HZGCoVF3YvM



Uncertainty about the model

Frequentist	Bayesian

• True model parameters are unknowable but fixed.
• Model parameters have no distributions, they simply exist!
  • Our estimates of model parameters have probability distributions.
  • We estimate sampling distributions to characterize the distributions of our estimates.
	• Model parameters are random, they have probability distributions.
	  • We estimate these from our fixed data.

Confidence and Credibility

Frequentist 95% confidence interval	Bayesian 95% credible 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. Our interval either does or does not contain the value, but we don't know for sure.	Given our data, we are 95% certain that our particular interval contains the real parameter value. We've used our current data, as well as prior knowledge to construct an interval that we're pretty sure contains the true value.

Inference Procedures: Frequentist

Frequentist	Bayesian
1. Estimate model 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.	1. Estimate probability distributions of the model 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

Conditional Probability	Hypothesis and Data
$Pr(A\|B)$: What is the probability of A given that we know B occurred? $Pr(H\|D)$: What is the probability of our hypothesis (H) given that we have observed the data (D)	**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

Four important probabilities/distributions

1. $Pr(Y)$: the probability or likelihood of our observed data
2. $Pr(\Phi_m)$: 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?
3. $Pr(Y|\Phi_m)$: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. $Pr(\Phi_m|Y)$: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!

Bayesian Intuition Example

- Imagine a brown creeper presence/absence study
- Response is the number of sites with a presence.
- We want to infer the probability of a presence.
- Let's say this is a new study, so we assume that we will observe them at 50% of patches. This is our prior belief.
  - This could be based on expert opinion, prior studies, etc.
- If you observed presences at 3 out of 3 sites, would you change your prior belief very much?
- What if you observed presences at 300 out of 300 sites?
- https://seeing-theory.brown.edu/bayesian-inference/index.html#section3

Conditional Probability

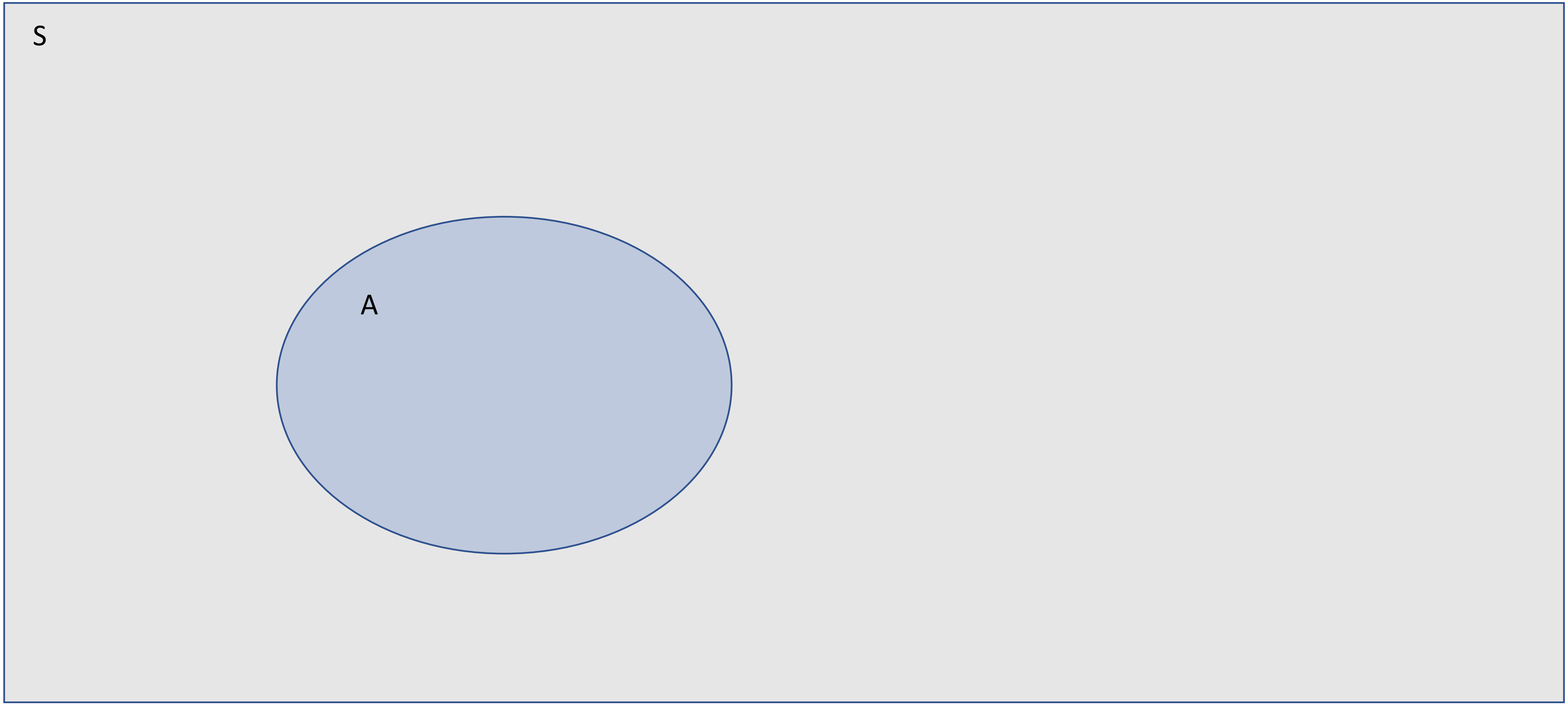
Conditional Probability: The Sample Space

S

Sample Space Properties

- Total area is 1.0
- The sample space contains the set of all possible events.
- $\Pr(S) = 1$

Conditional Probability: An Event

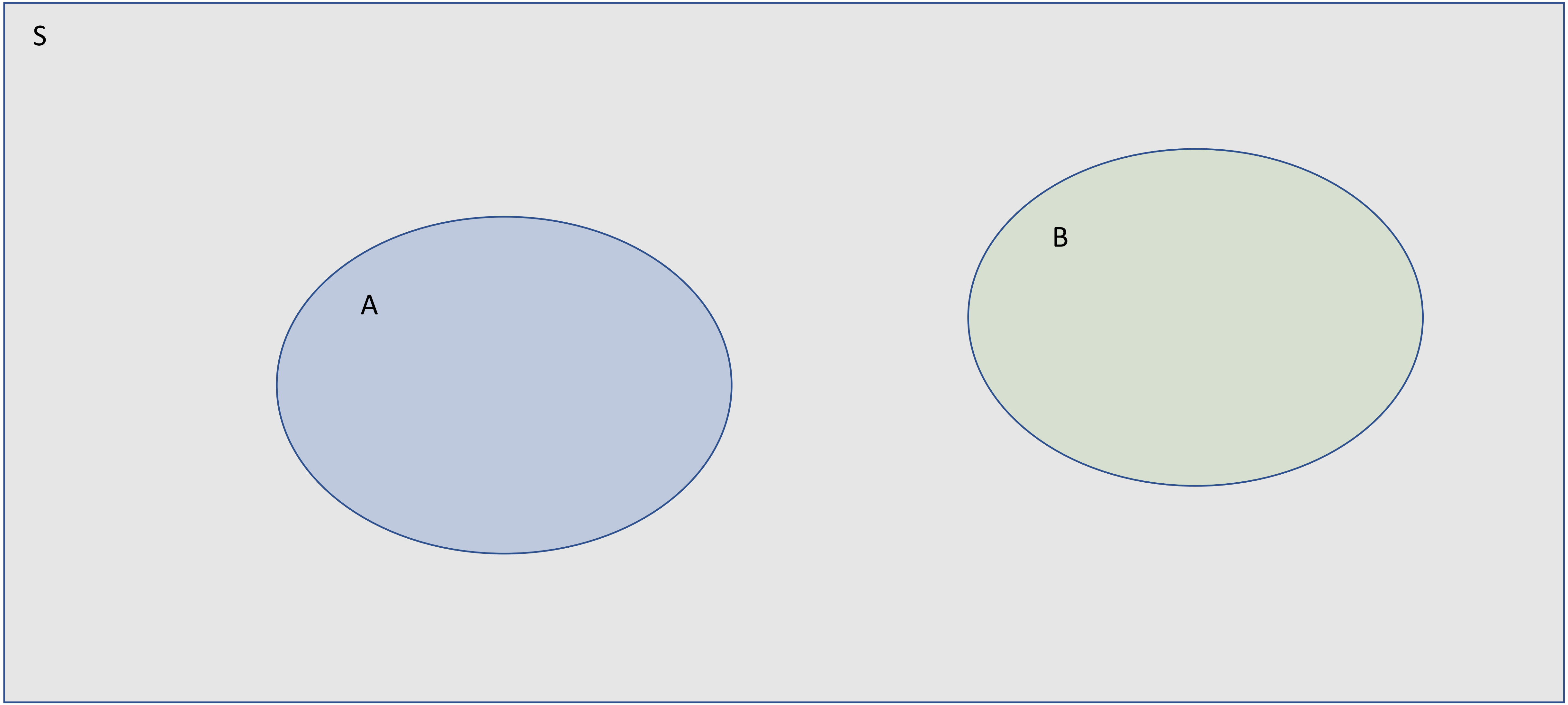


Event A Properties

- What's the probability of A?
  - We know it is equal to or less than 1.0

$$\Pr(A) = \frac{Area\ of\ A}{Area\ of\ S} = Area\ of\ A$$

Conditional Probability: Another Event



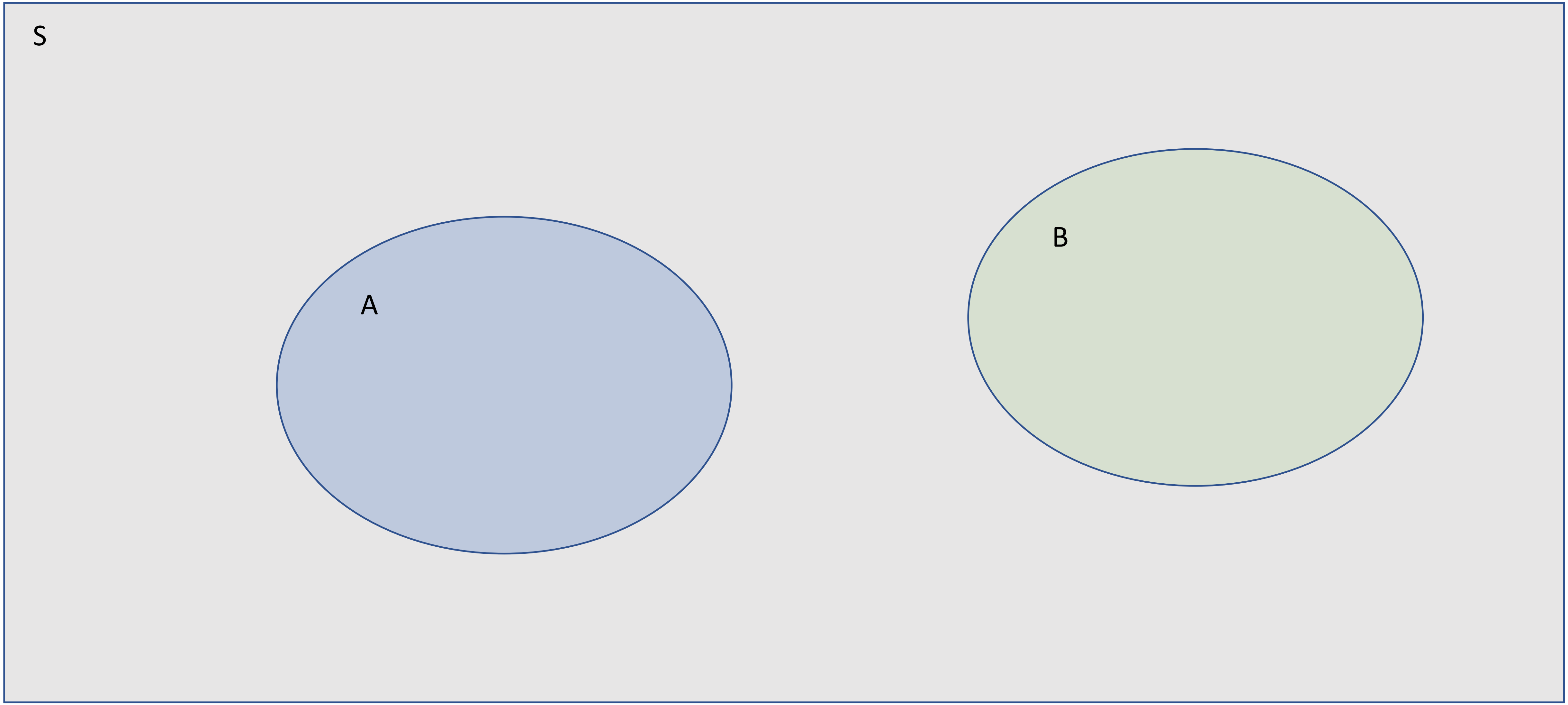
Event A Properties

- What are the event probabilities?

$$\Pr(A) = \frac{Area\ of\ A}{Area\ of\ S}$$

$$\Pr(B) = \frac{Area\ of\ B}{Area\ of\ S}$$

Exclusive Events



Exclusive Events

Some different perspectives:

- If event A occurs, B cannot occur
- There is no overlap between A and B
- The conditional probability of A given B is zero.
  - $\Pr(A|B) = 0$
- The conditional probability of B given A is zero.
  - $\Pr(A|B) = 0$

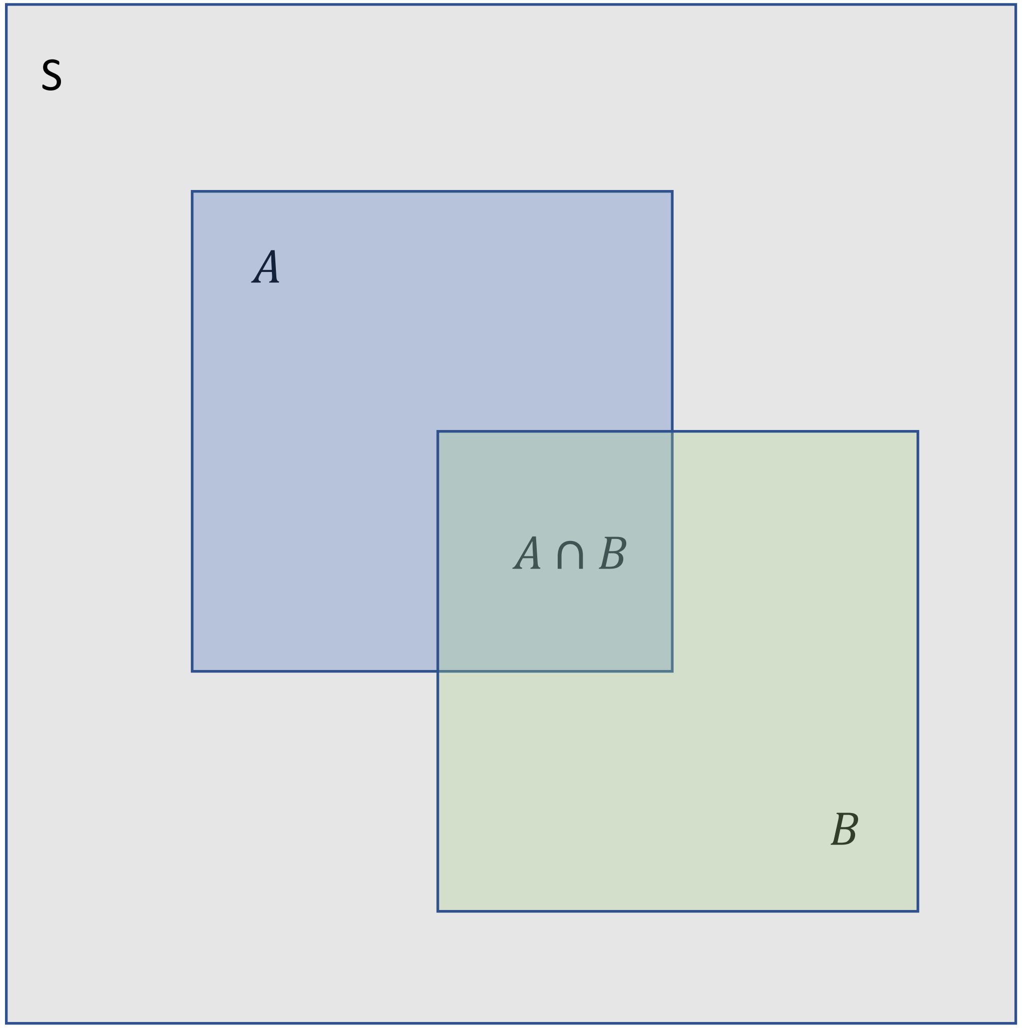
Complementary Events

Complementary Events

- Complementary events are exclusive.
  - Something can be in A or B, but not both.
  - $\Pr(A|B) = 0, \Pr(B|A) = 0$
- Complementary events fill the sample space
  - If something is not in A, then it is in B
  - $\Pr(A) + \Pr(B) = 1.0$

Overlapping Events: Conditional Probability

- We want to know: $\Pr(B|A)$



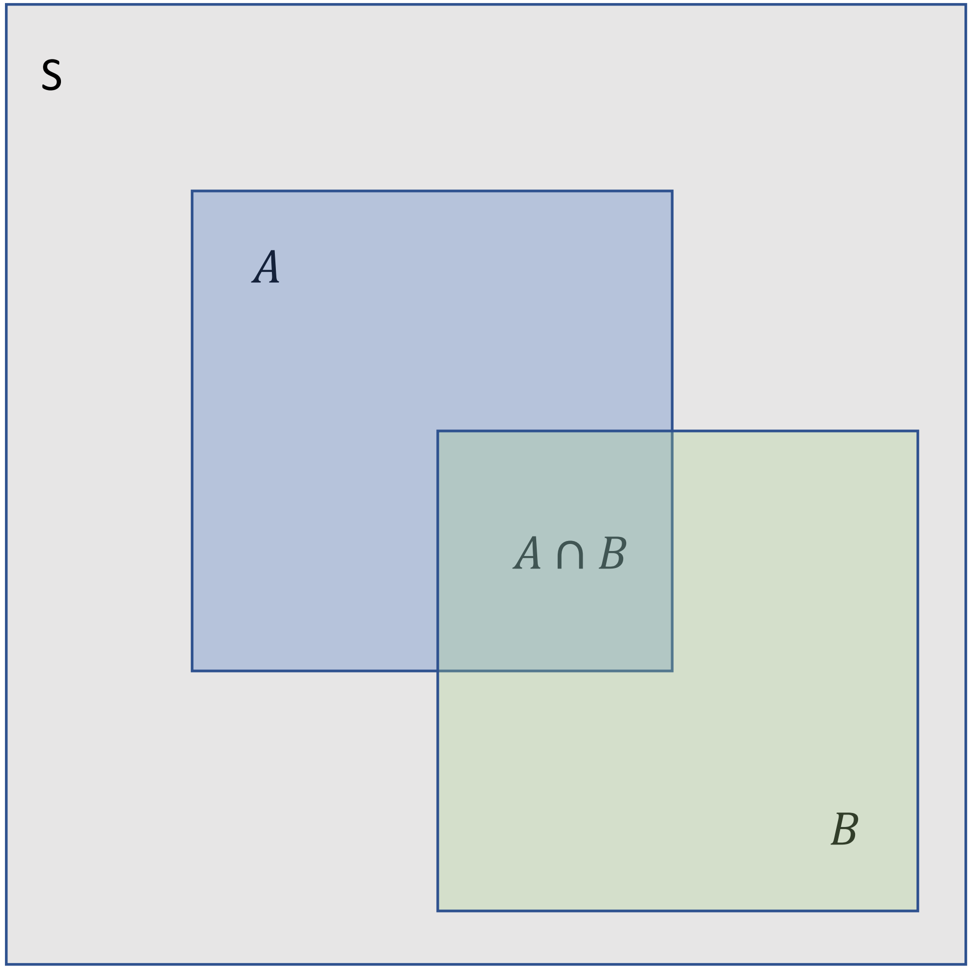
Overlapping Events: Conditional Probability

Conditional probability: We want to know how likely we are to observe an event in B if we've already observed an event in A.

In symbols:

$$\Pr(B|A)$$

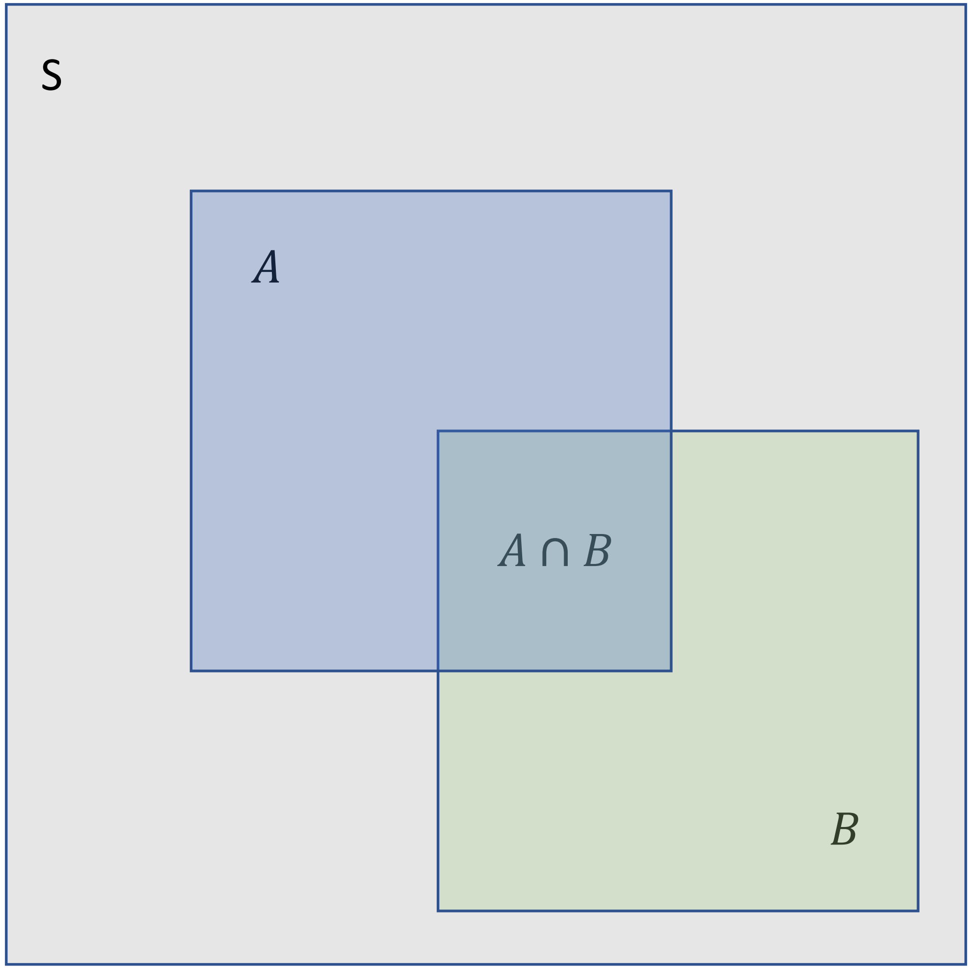
You can read this as "Probability of B given A".



Overlapping Events: Conditional Probability

What happens when we observe event A?

Our sample space changes…

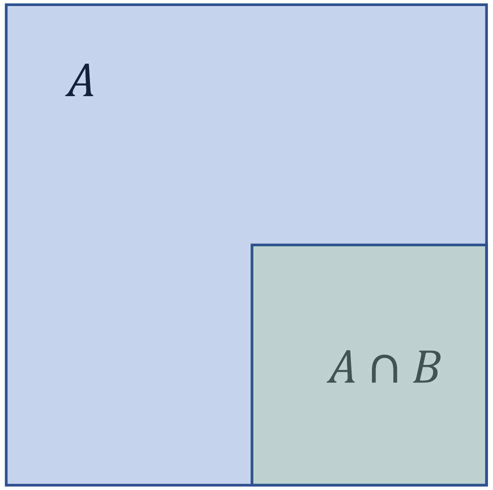


Overlapping Events: Conditional Probability

What happens when we observe an event in A?

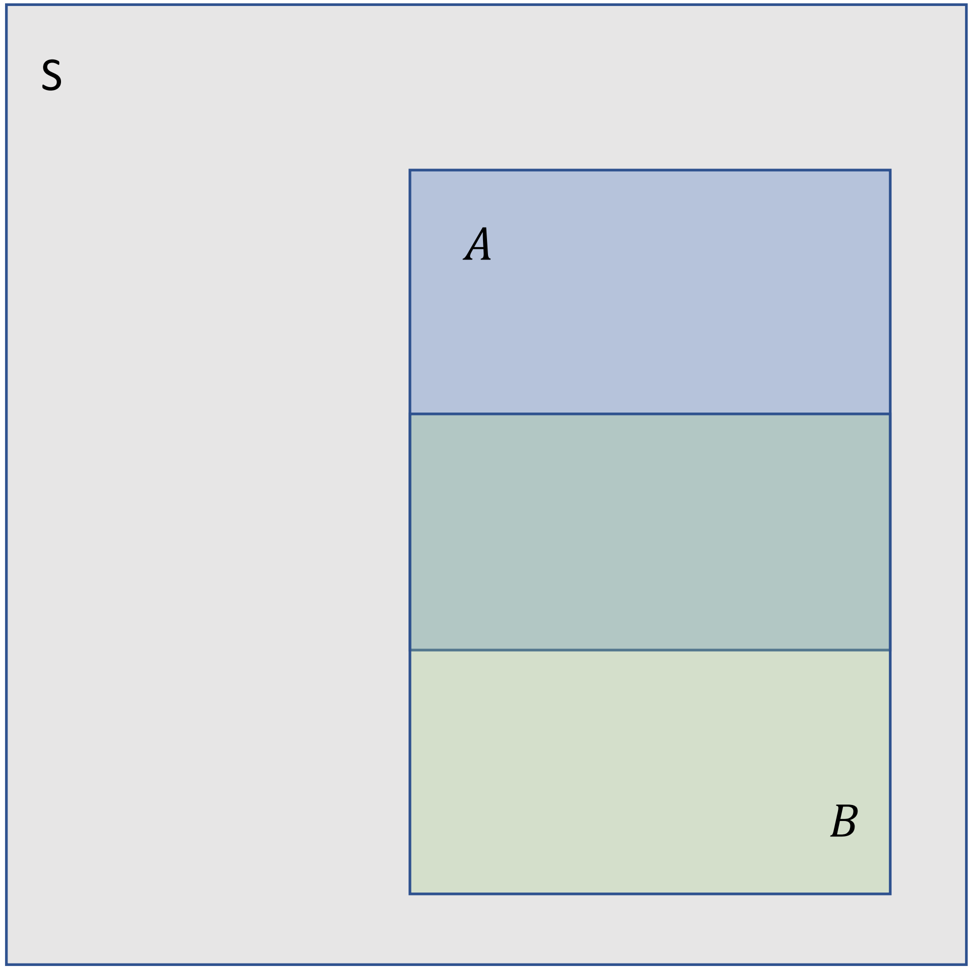
Our sample space changes… It collapses into A.

Since A contained part of B, $\Pr(B|A)$ is just $\frac{\Pr(A \cap B)}{\Pr(A)}$



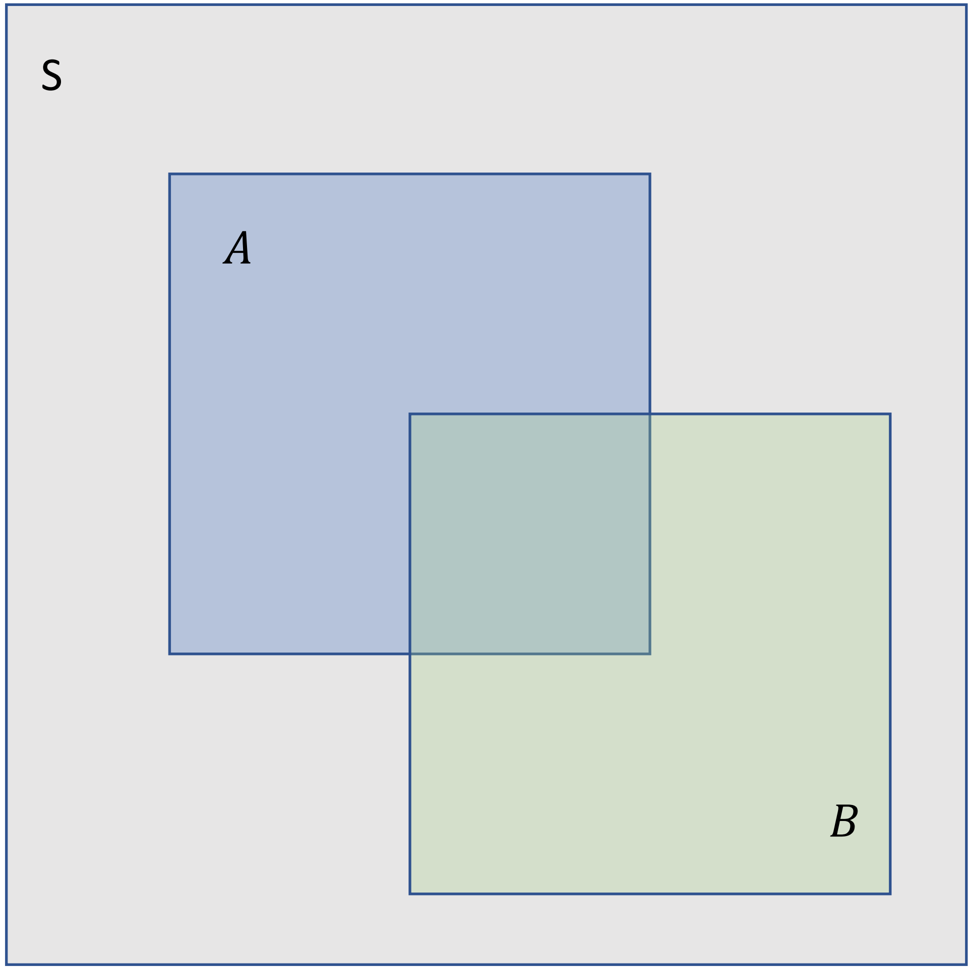
Overlapping Events: Conditional Probability

Try to guess the value of $\Pr(B|A)$



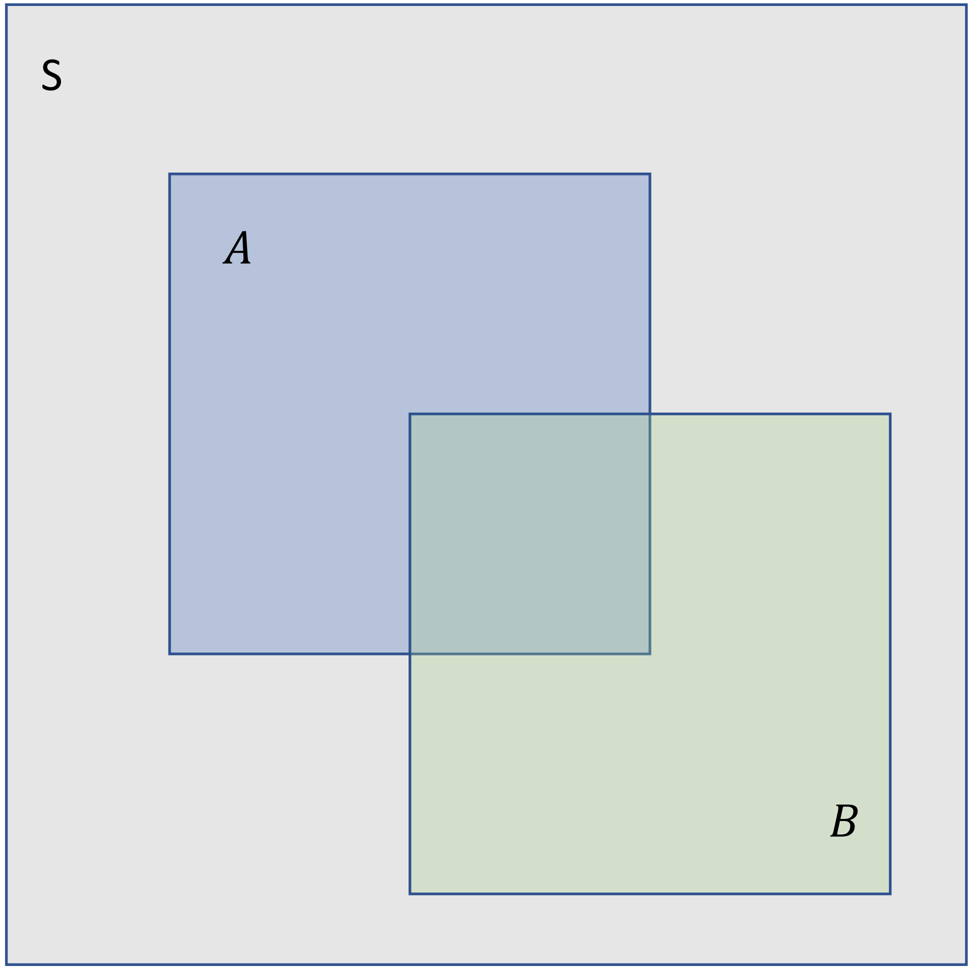
Overlapping Events: Conditional Probability

Try to guess the value of $\Pr(B|A)$



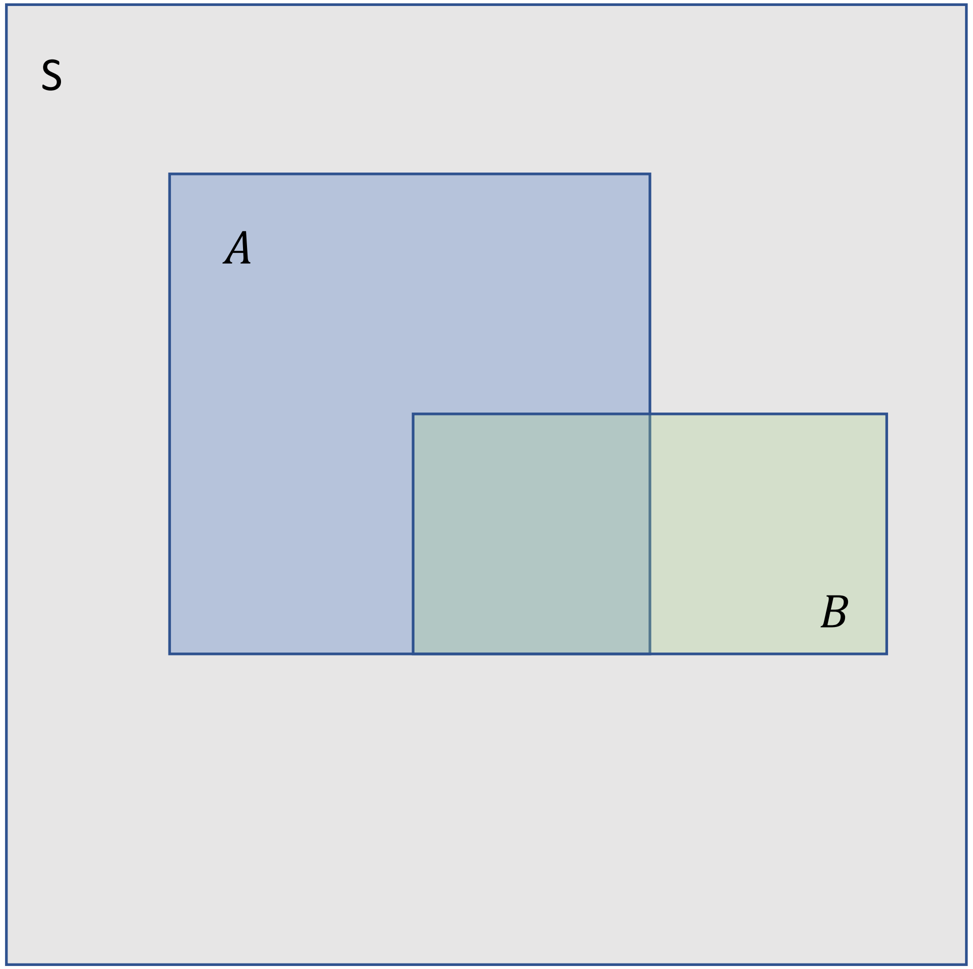
Overlapping Events: Conditional Probability

Try to guess the value of $\Pr(A|B)$



Overlapping Events: Conditional Probability

Try to guess the value of $\Pr(A|B)$



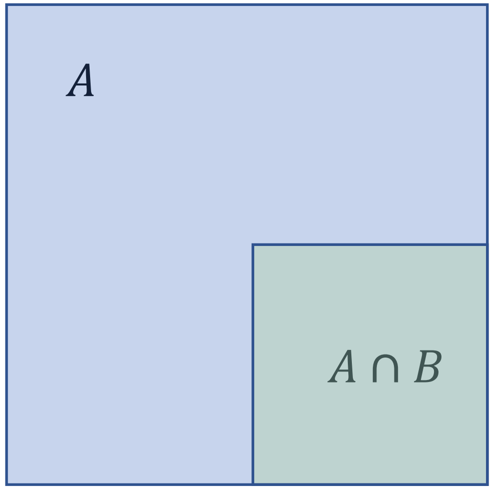
Overlapping Events: Independent Events

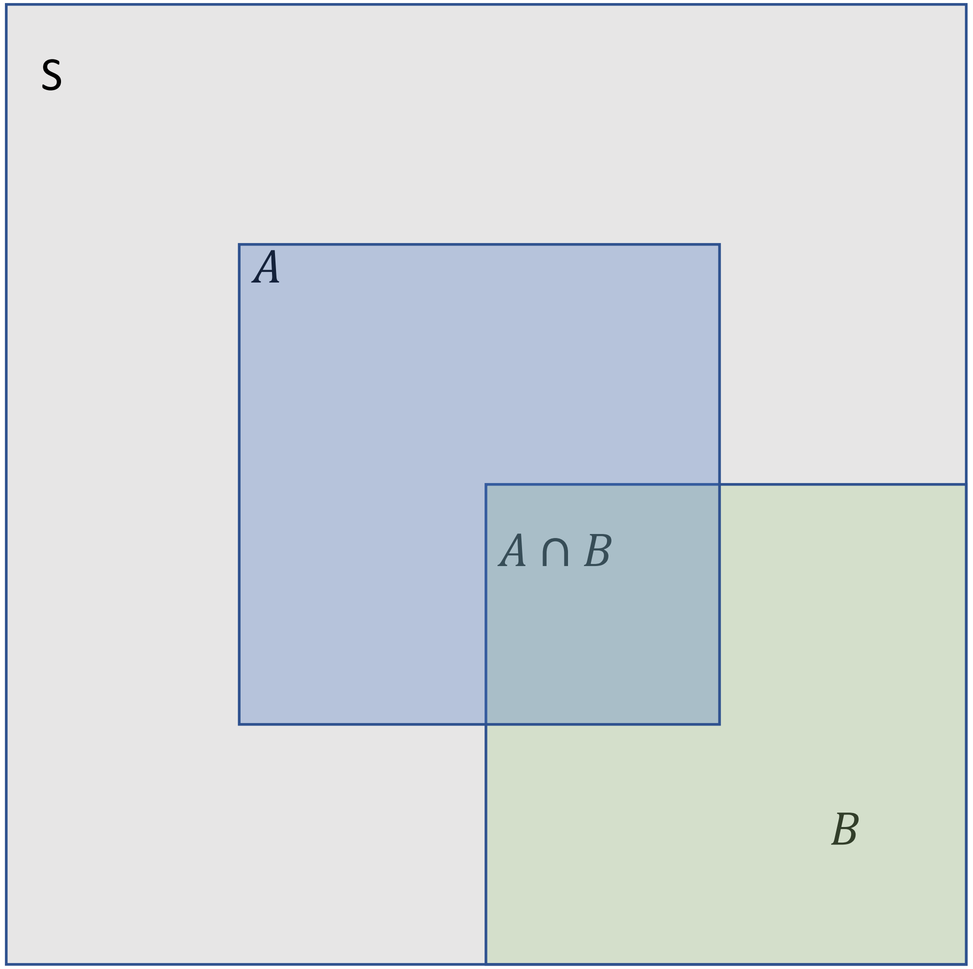
If events are independent, then:

$$\Pr(B|A) = \Pr(B)$$

In the figure, Pr(B) is 0.25

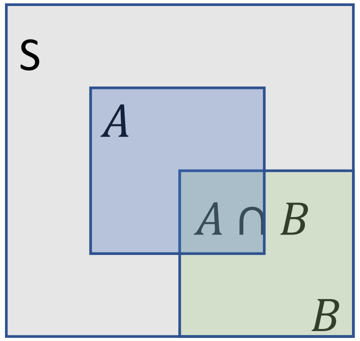
- A is ¼ of the sample space
- B is ¼ of the sample space
- $A \cap B$ is ¼ of A
- $A \cap B$ is 1/16 of the sample space

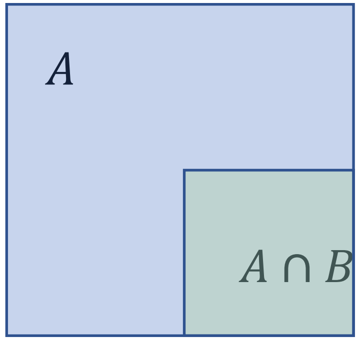




Conditional Probability: Key Points

- Sample Space
- Complementary Events
- Exclusive Events
- Overlapping Events
- To calculate a conditional probability, the sample space changes.
  - You 'collapse' the sample space into the conditioned event's space





Relationship between model and data

Frequentist	Bayesian
• Data are one realization of a stochastic sampling process. • The One True Model exists and is unknowable.	• 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	Bayesian

• True model parameters are unknowable but fixed.
• Model parameters have no distributions, they simply exist!
    • Our estimates of model parameters have probability distributions.
    • We estimate sampling distributions to characterize the distributions of our estimates.
	• Model parameters are random, they have probability distributions.
	    • We estimate these from our fixed data.

Confidence and Credibility

Frequentist 95% confidence interval	Bayesian 95% credible 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. Our interval either does or does not contain the value, but we don’t know for sure.	Given our data, we are 95% certain that our particular interval contains the real parameter value. We’ve used our current data, as well as prior knowledge to construct an interval that we’re pretty sure contains the true value.

Inference Procedures: Frequentist

Frequentist	**Bayesian**
1. Estimate model 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.	1. Estimate probability distributions of the model 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

Conditional Probability	Hypothesis and Data
$Pr(A\|B)$: What is the probability of A given that we know B occurred? $Pr(H\|D)$: What is the probability of our hypothesis (H) given that we have observed the data (D)	**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

Four important probabilities/distributions

1. $Pr(Y)$: the probability or likelihood of our observed data
2. $Pr(\Phi_m)$: 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?
3. $Pr(Y|\Phi_m)$: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. $Pr(\Phi_m|Y)$: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!

Key Concepts

- Conditional probability
- Bayes' Rule
- Differences between Frequentist and Bayesian perspectives