# Intro to Bayesian Statistics

GENERAL ASSEMBLY

# Agenda

- Review: Conditional Probability & Probability Rules
- Bayes Theorem
- Bayesian Inference

# Learning Objectives

- Understand Bayes Theorem and how it's derived.
- Describe questions that can be answered by Bayesian inference.
- Compare and contrast Bayesian and Frequentist inference.
- Describe the role of the prior distribution and likelihood on the posterior distribution.

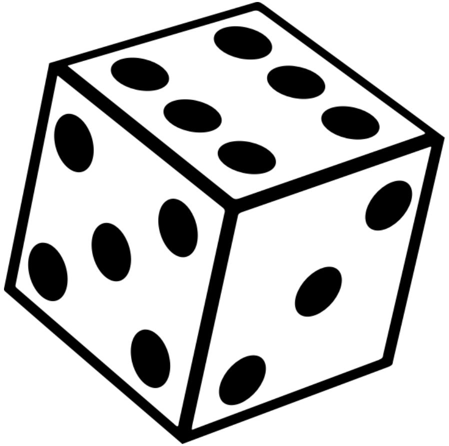
Bayesian Statistics

# Review: Conditional Probability & Probability Rules

# Example 1

Let's start with an example.

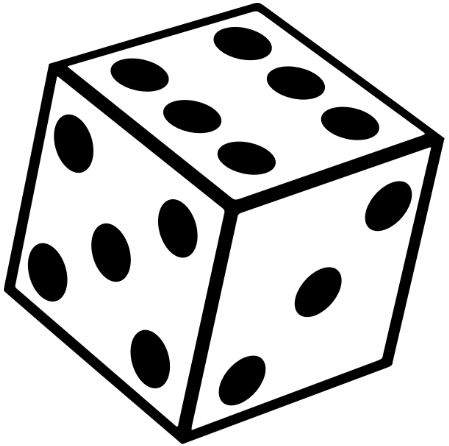
We roll one standard, six-sided die. What is the probability of rolling a 1?



# Example 1

We roll one standard, six-sided die. What is the probability of rolling a 1?

→ ⅙

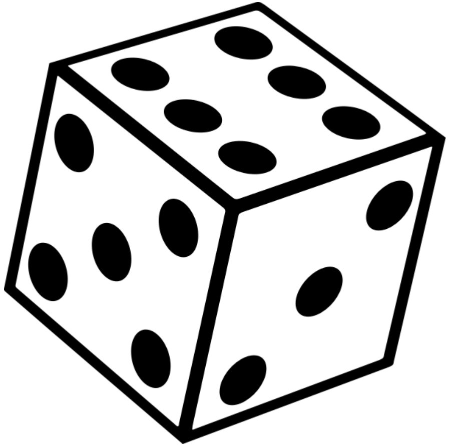


# Example 1

We roll one standard, six-sided die.

I tell you that we rolled an odd number.

Now, what is the probability of rolling a 1?



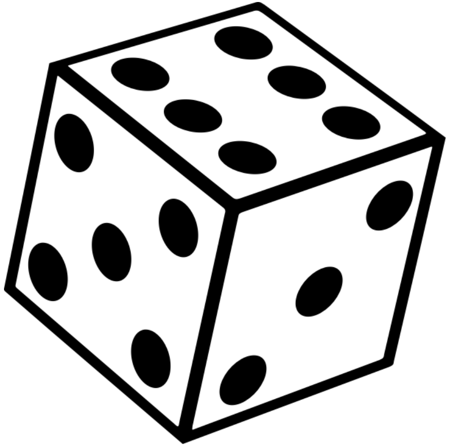
# Example 1

We roll one standard, six-sided die.

I tell you that we rolled an odd number.

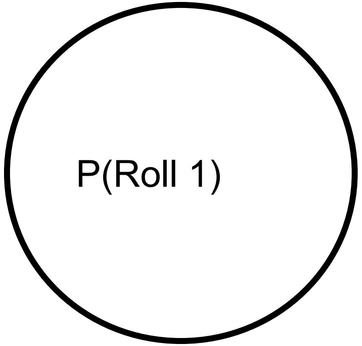
Now, what is the probability of rolling a 1?

➜ ⅓



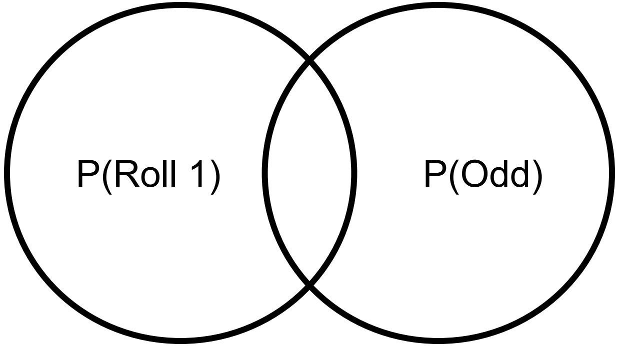
# Reminder: Joint Probability

Intersection: A ∩ B = the set of elements in A and B



# Reminder: Joint Probability

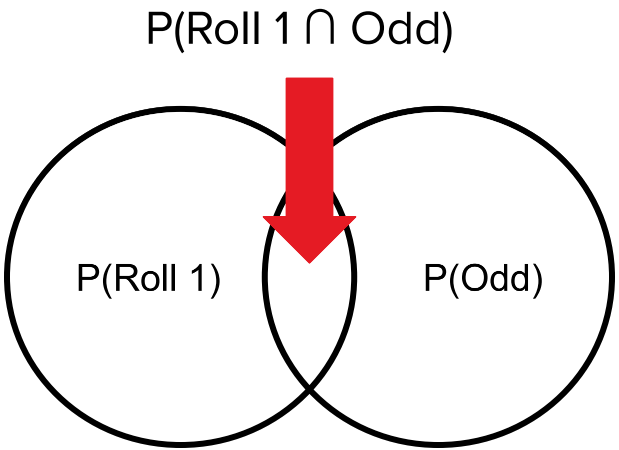
Intersection: $A \cap B$ = the set of elements in A and B




 | 

# Reminder: Joint Probability

Intersection: A ∩ B = the set of elements in A and B



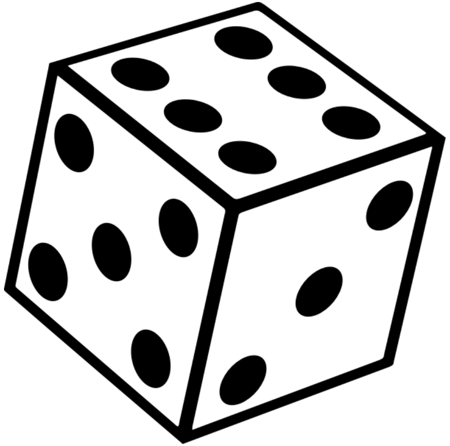
# Reminder: Probability Rules

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Note: A|B means “A given B” or “A conditional on the fact that B happens.”

# Reminder: Probability Rules

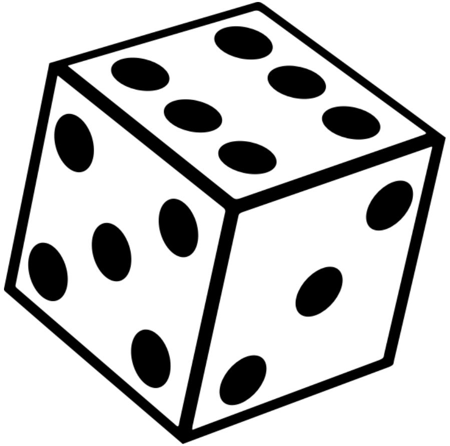
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



# Reminder: Probability Rules

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

P(Roll 1 | odd)

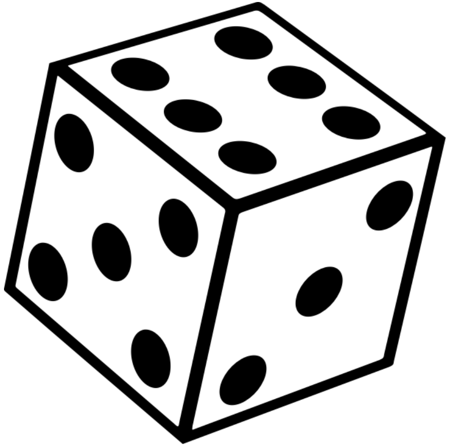


# Reminder: Probability Rules

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

P(Roll 1 ∩ odd) = ⅙

P(Roll 1 | odd)



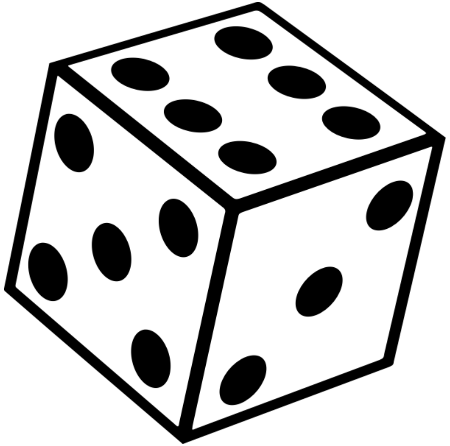
# Reminder: Probability Rules

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

P(Roll 1 ∩ odd) = ⅙

P(odd) = ½

P(Roll 1 | odd)



 | 

# Reminder: Probability Rules

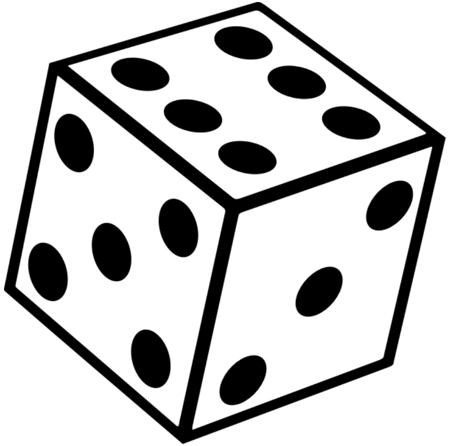
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

P(Roll 1 ∩ odd) = ⅙

P(odd) = ½

= ⅓

P(Roll 1 | odd)



# Reminder: Probability Rules

$$P(A \cap B) = P(A|B)P(B)$$

We took the last rule, multiplied both sides by P(B), and voila!

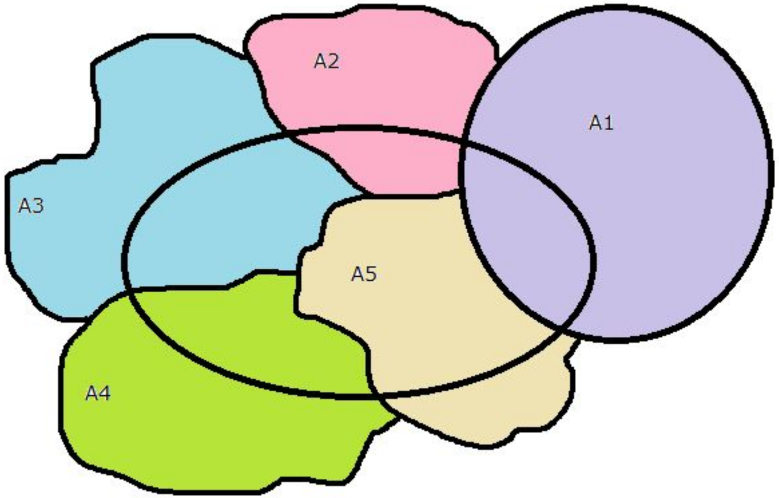
We can rearrange these as well!

$$P(B \cap A) = P(B|A)P(A)$$

# The Law of Total Probability

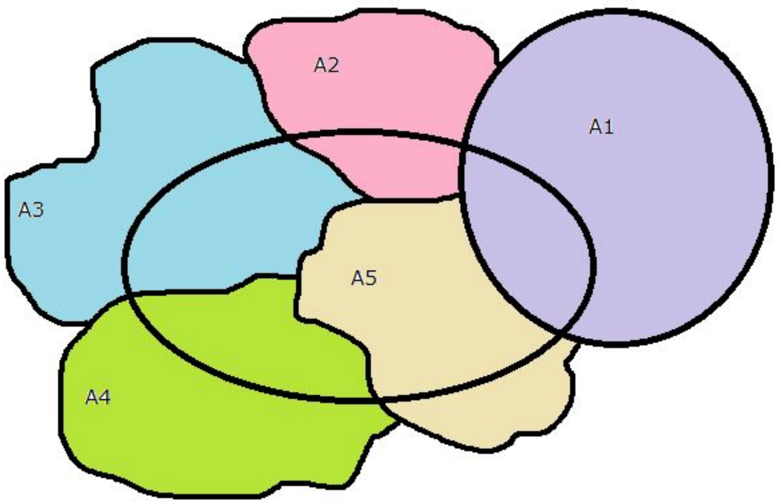
$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$



# The Law of Total Probability

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$

$$P(B \cap A) = P(B|A)P(A)$$



Bayesian Statistics

# Bayes' Theorem

# Bayes' Theorem: Purpose

Bayes theorem allows us to find conditional probability based on our prior knowledge/experience.

# Bayes’ Theorem: Analogy

You hear a car alarm go off. What do you do?

A. Call the police.
B. Look outside to see which car it is/make sure everything is okay.
C. Do nothing.

# Bayes’ Theorem: Analogy

You hear a car alarm go off. What do you do? For me:

A. Call the police.
B. Look outside to see which car it is/make sure everything is okay.
C. **Do nothing.**



https://betterexplained.com/articles/an-intuitive-and-short-explanation-of-bayes-theorem/

# Bayes' Theorem

Bayes' Theorem relates P(A|B) to P(B|A).

It allows us to use prior knowledge/evidence/beliefs to inform our decisions.

# Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- P(A) is the probability that A occurs given no supplemental information.
- P(B|A) is the likelihood of seeing evidence (data) B assuming that A is true.
- P(B) is the probability that B occurs given no supplemental information.

# Bayes' Theorem

Likelihood

Class Prior Probability

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Posterior Probability

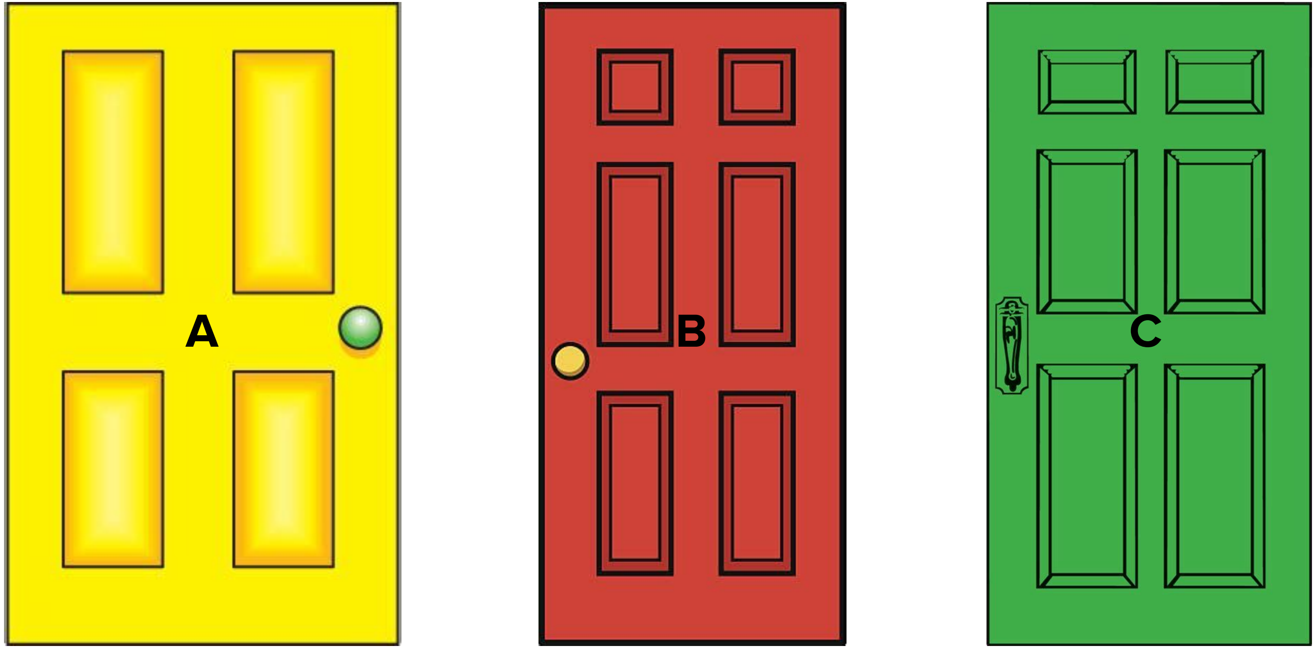
Predictor Prior Probability

# The Monty Hall Problem

Let’s play a game. You have a once-in-a-lifetime opportunity to win a sports car! All you have to do is pick which door the car is hidden behind. The other two doors have goats behind them.

# The Monty Hall Problem

Pick a door!



# The Monty Hall Problem: What should you do?

If you continue to repeat this game, you will notice that you will win more often if you **switch doors**!

How can this be? Isn't the probability of the car being behind each door just ⅓?

Let's use Bayes' Theorem to prove it.

# To the notebook!

Bayesian Statistics

# Bayesian Inference

# What is Bayesian inference?

You write a function.

How likely is it that your code doesn't contain any errors?

# What is Bayesian inference?

You write a function.

How likely is it that your code doesn’t contain any errors?

It probably depends on your confidence in your code. This could be based on prior code you’ve written or the specific problem.

# What is Bayesian inference?

You run one example through your function and it works!

Now how likely is it that your code doesn't contain any errors?

# What is Bayesian inference?

You run one example through your function and it works!

Now how likely is it that your code doesn’t contain any errors?

You’re confidence in the code has probably increased a little!

# What is Bayesian inference?

You run several more tests through your function and they all work!

At this point, you're probably feeling pretty confident in your function.



https://nbviewer.jupyter.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter1_Introduction/Ch1_Introduction_PyMC3.ipynb

# What is Bayesian inference?

This is an example of you naturally thinking like a Bayesian.

Bayesian inference allows us to update our current beliefs based on new information we get.

# Bayesian vs. Frequentist Inference

Frequentist Inference

# Bayesian vs. Frequentist Inference

Frequentist Inference

- “Classic” statistical thinking
- Probability is based on the frequency of events
- For example, what is the probability of a candidate winning a political election?
    - This is really hard for a frequentist to answer since each specific election only happens once!

# Bayesian vs. Frequentist Inference

Bayesian Inference

- Probability is interpreted as how confident we are in an event occurring
- We take our prior beliefs into account and can continue to update our beliefs as we get more data
- For example, what is the probability of a candidate winning a political election?
    - This is easier to answer from a Bayesian perspective - how confident are we that a candidate will win? We can take prior information (the polls!) into account to find this.

“

**When the facts change, I change my mind - what do you do?**

John Maynard Keynes

## Bayesian Inference

- More intuitive answers
- Works with low sample sizes
- Less common
- Harder to initially set up
- We include our personal knowledge/beliefs: interpreted as more subjective

## Frequentist Inference

- Less intuitive answers
- Requires larger sample sizes
- More common
- Simpler to get set up
- Data speaks for itself: interpreted as more objective

# How do we do Bayesian Inference?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) \propto P(B|A)P(A)$$

# How do we do Bayesian Inference?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) \propto P(B|A)P(A)$$

“Is proportional to”

# How do we do Bayesian Inference?

$$P(A|B) \propto P(B|A)P(A)$$

Instead of events/single numbers, we represent these as distributions

# How do we do Bayesian Inference?

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

# How do we do Bayesian Inference?

Our goal: Generate this!

We pick these!

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

# How do we do Bayesian Inference?

$$f(\mu|\text{data}) \propto f(\text{data}|\mu) f(\mu)$$

This is our **prior distribution**: this summarizes what we think about our parameter **before we gather any data**.

# How do we do Bayesian Inference?

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

This is our **likelihood**: this summarizes how likely it is to observe the data we observed given various values of our parameter.

# How do we do Bayesian Inference?

**Our goal: Generate this!**

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

This is our **posterior distribution**: this summarizes what we think about our parameter **after taking our data into account**. It is a combination of our prior beliefs and our data.

# Steps to Bayesian Inference

1. Select a prior distribution
2. Select a likelihood
3. Generate the posterior distribution
4. Do whatever inference we want!

# Steps to Bayesian Inference

1. Select a prior distribution
2. Select a likelihood
3. Generate the posterior distribution
4. Do whatever inference we want!

You will see an example of this done in Python this afternoon!