BUSN 5000

Bayes' Rule

Abigail Cormier

Terry College of Business

Spring 2025
(updated 19 Feb 25)

What you need to know this week

- What is the difference between a frequentist and a Bayesian?
- Why learn Bayes' rule?
- How do you set up a Bayes' rule problem geometrically?
- What is the mathematical formula for Bayes' rule?
- What are the key vocabulary terms for Bayes' rule?

Section 1

Certainly Guilty?

Collins Case Background

Facts of the case:

1. It is 1964 in Los Angeles, CA
2. Woman is thrown to the ground and purse is stolen
3. Eyewitness gives a description, but doesn't get a good look
    - Woman: white, with blond ponytail
    - Man: black, with mustache and beard
    - Couple drove a yellow car

Based on this eyewitness description: police arrest Malcolm and Janet Collins

Trial of Malcolm and Janet Collins

Prosecution brings in a mathematician to testify that based on the match to the description alone, the Collins are guilty. The mathematician provides something like the following calculation:

1. Probability a man in the U.S. is black: $\frac{1}{10}$
2. Probability a man has a beard: $\frac{1}{10}$
3. Probability a man has a mustache: $\frac{1}{5}$
4. Probability a man drives a yellow car: $\frac{1}{200}$

Trial of Malcolm and Janet Collins

Prosecution brings in a mathematician to testify that based on the match to the description alone, the Collins are guilty. The mathematician provides something like the following calculation:

1. Probability a man in the U.S. is black: $\frac{1}{10}$
2. Probability a man has a beard: $\frac{1}{10}$
3. Probability a man has a mustache: $\frac{1}{5}$
4. Probability a man drives a yellow car: $\frac{1}{200}$

Then the mathematician calculated the probability of these events occurring, assuming that they were independent, and continued adding characteristics of the couple until arriving at a probability of 1 in 12 million that the couple was innocent and matched the description of the eyewitness.

Result: Jury convicts both individuals

Does this seem like good reasoning?

Section 2

Frequentism vs. Bayesian

What is frequentism again?

In short, frequentism is the framework we have been working in for the entirety of this class so far to develop ideas about statistical inference.

The frequentist definition of probability, as a refresher is:

Definition. Let A be some event and $S(N)$ the number of occurences in N random trials. Then the probability of event A is:

$$P(A) = \lim_{N \to \infty} \frac{S(N)}{N}$$

Then we discussed the idea by calling on the law of large numbers and a fair coin flip simulation.

Frequentist vs. Bayesian

A Frequentist and a Bayesian walk into a room with a large cardboard box that is filled with marbles. Illustrating the difference between these two frameworks is possible by showing what each would do when confronted with the question:

What is the probability of any given color in the box?

1. What does the Frequentist do?
2. What does the Bayesian do?

Why should you care?

If our foundation for inference is built on frequentism, why bother learning Bayes' rule at all?

Three thoughts:

1. We are all Bayesians
2. Importance of asking the right question
3. Practically, it will be on the (next) test

Section 3

Bayes' Rule

Motivating Example

Meet Ali:

- Quiet and studious person
- Enjoys economics
- Has a passion for detail
- Loves learning about econometrics and the law of demand

Motivating Example

Meet Ali:

- Quiet and studious person
- Enjoys economics
- Has a passion for detail
- Loves learning about econometrics and the law of demand

Which is more likely: Ali is an economist or Ali is a data scientist?

Thinking about context

How common is it to be an economist or a data scientist?

According to the Census bureau:

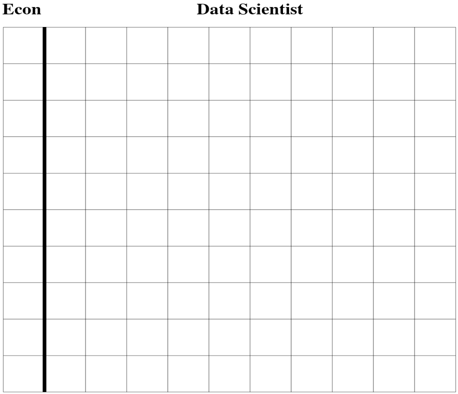
- Approximately 16,000 economists in the U.S.
- Approximately 160,000 data scientists in the U.S.

This means that for every 1 economist, there are approximately 10 data scientists. How might this ratio matter in the context of the question I just asked you?

Sidenote: this is not a comment on knowing obscure facts like the number of economists vs. the number of data scientists—I did not know it before writing this lecture. **It is about whether you thought to ask the question.**

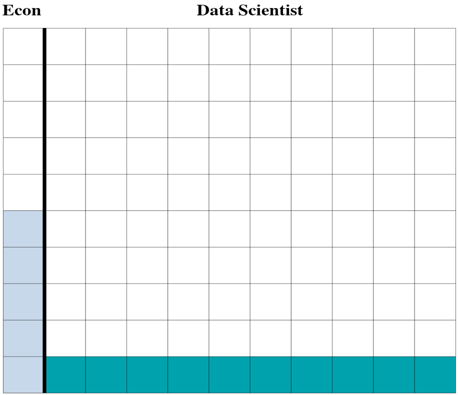
Walking through an example

So the ratio of data scientists to economists is 10 to 1. Let's use this fact to reason through the question of Ali's likely professional identity.



Add what we know about Ali

Let's add in the description (our evidence) and say about $\approx 50\%$ of economists fit this description and $\approx 10\%$ of data scientists fit it:

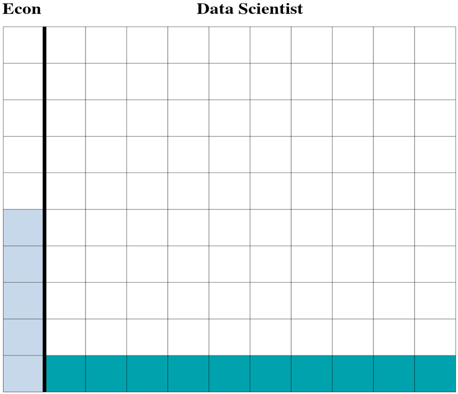


Knowing what we know now: **What is the probability that Ali is an economist given the description I provided about him?**

How many of each type fit the description?

From our sample of 110 total, with 100 data scientists and 10 economists:

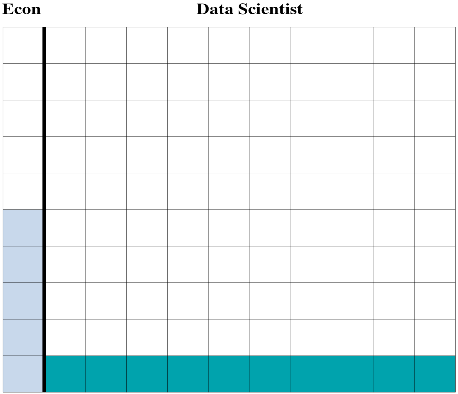
- How many individuals who are economists fit our description?
- How many individuals that are data scientists fit our description?



$P(Economist|Description)$

What is the probability that Ali is an economist given the description I provided about him or what is $P(Economist|Description)$?

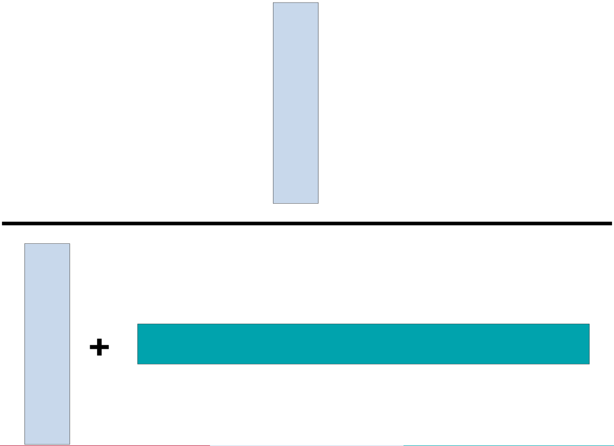
- To answer: we think about taking our prior images and pulling out the needed sections:



$P(Economist|Description)$

What is the probability that Ali is an economist given the description I provided about him or what is $P(Economist|Description)$?

- To answer: we think about taking our prior image and pulling out the needed sections:



$P(Economist|Description)$

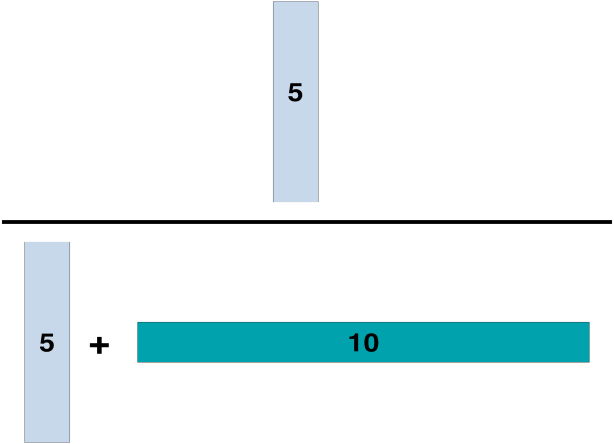
Adding in what we know about the number of people in each section:

What was that again?

$P(Economist|Description)$

Adding in what we know about the number of people in each section:

What was that again?



An answer to $P(Economist|Description)$

We can now see that the $P(Economist|Description)$ is $\approx 33\%$. This is **despite** the fact that we said an economist was 5 times as likely to fit the description compared to the alternative occupation of data scientist!

How did we get here?

- We took the new information - Ali's description (Evidence)
- Incorporated outside knowledge - the probability of being an economist (Prior)
- We used the information that the probability of fitting a description given the individual is an economist is 50% (Likelihood)

Section 4

The math behind Bayes' Rule

Adding notation to the posterior

You have already seen me incorporate conditional probability into our discussion, but let's say a bit more about what this notation about the **posterior**, or the probability we want to learn about is saying:

$$P(Economist|Description)$$

is a conditional probability that should be read as "the probability that the person is an economist *given* (|) the description provided.

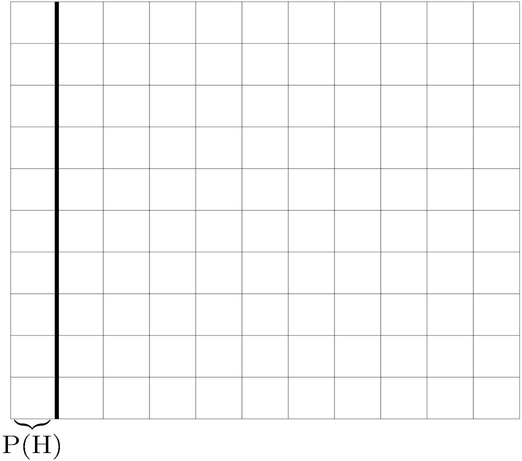
We could generalize this to a formulation that makes sense in many contexts as:

$$P(Hypothesis|Evidence) \text{ or } P(H|E)$$

which can be read as the probability the hypothesis is true given that we have seen the evidence.

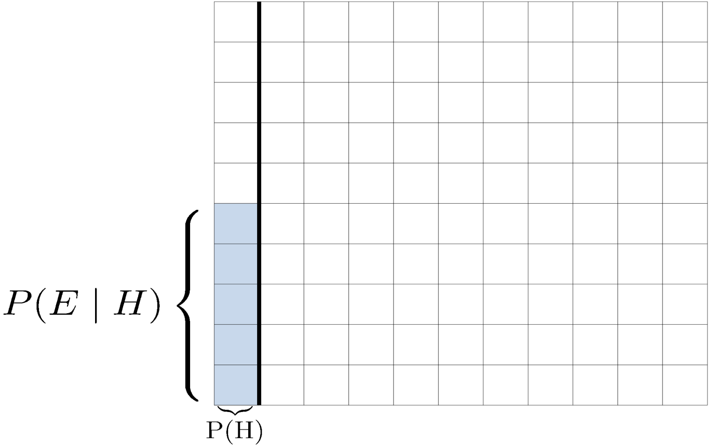
Adding notation to the prior

Let's bring up our image again and discuss the notation for the **prior** or the probability the hypothesis is true:



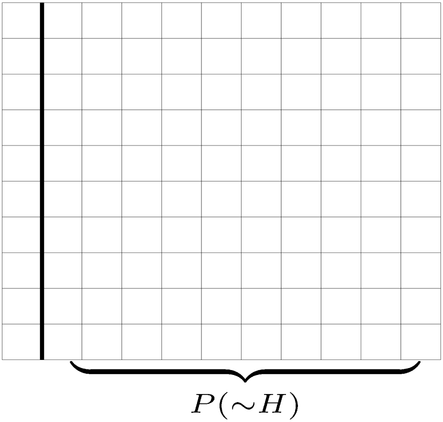
Adding notation to the likelihood

Now, let's add in the notation for the likelihood, or $P(E|H)$, which in our example was equal to the probability that a person fits the description given they are an economist.



Adding notation to the space of $P(\sim H)$

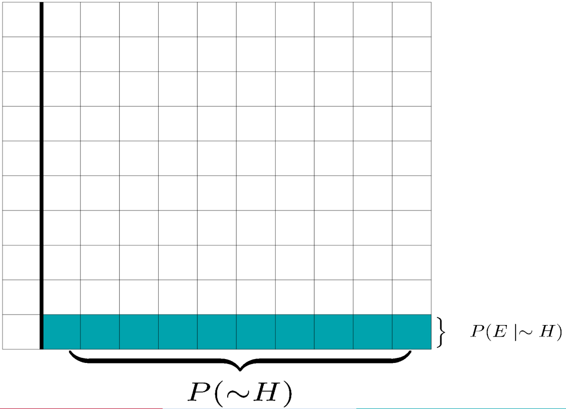
It is also important to think about the probability our hypothesis is false or $P(\sim H)$, which in this example is that Ali is a data scientist.



Adding notation to $P(E|\sim H)$

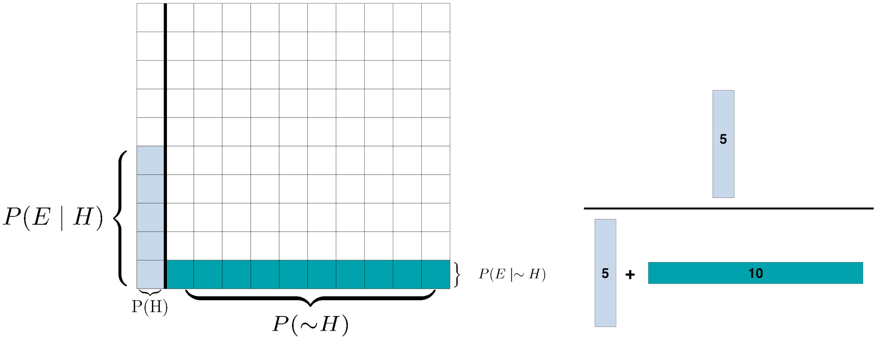
As we saw, a key part of this space is $P(Description|\sim Economist)$ or $P(E|\sim H)$. Essentially, this is the case where we see the evidence, but our prior is false.

What was the $P(Description|\sim Economist)$?

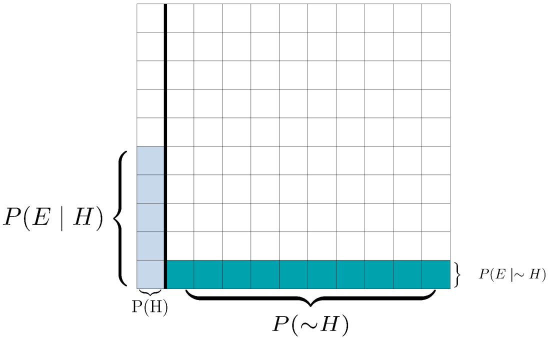


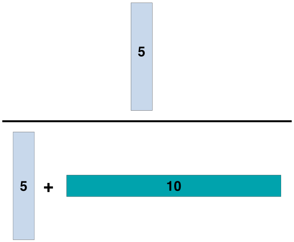
Putting it all together

Let's pull up our image where we calculated the $P(Economist|Description)$ or the **posterior** and see if we can add in the math notation we have been discussing:



Putting it all together





- Where do we get 5?
- Where do we get 15?

A formula for Bayes' Rule!

This work leaves us with the following:

$$P(H|E) = \frac{P(H)P(E|H)}{P(H)P(E|H) + P(\sim H)P(E|\sim H)}$$

You will sometimes see this simplified as:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

but for nearly every case of solving a problem using Bayes' Rule, you will need the first formula we derived.

Reviewing vocabulary of Bayes' Rule

Bayes' rule can be expressed in terms of learning about some hypothesis (H) given new evidence (E):

$$\underbrace{P(H|E)}_{\text{posterior}} = \frac{\overbrace{P(E|H)}^{\text{likelihood}} \times \overbrace{P(H)}^{\text{prior}}}{\underbrace{P(E|H)P(H) + P(E|\sim H)P(\sim H)}_{\text{marginal probabilty of the evidence}}}.$$

Posterior - your updated belief which is a function of the following components:

- your prior belief
- the likelihood of the evidence
- the strength of the evidence (determined by $P(E|H)$ and $P(E|\sim H)$)

Section 5

Returning to the Collins case

First problem: Independent probabilities

Recall that I said the mathematician took many details about the couple's characteristics and multiplied them together as *independent* probabilities and got a probability of 1 in 12 million... But does that seem like the right approach? Are these features all independent?

We might reasonably think that features like beard and mustache might be correlated, meaning, those with beards are more likely to have mustaches.

If you account for all of these relationships and correlations, let's say we get a new probability of innocence as 1 in a million, still very low – are we done?

First problem: Independent probabilities

Recall that I said the mathematician took many details about the couple's characteristics and multiplied them together as *independent* probabilities and got a probability of 1 in 12 million... But does that seem like the right approach? Are these features all independent?

We might reasonably think that features like beard and mustache might be correlated, meaning, those with beards are more likely to have mustaches.

If you account for all of these relationships and correlations, let's say we get a new probability of innocence as 1 in a million, still very low – are we done?

I hope you said no, because the **biggest issue** is that this entire analysis so far has been answering the wrong question!

Again, asking the right question and credibly answering it is crucial.

Asking the right question

What is the right question to ask?

Asking the right question

What is the right question to ask?

How about, “How likely is the Collins couple to be innocent given that they match the eyewitness description?”

The following facts will be useful:

- The population of LA in 1964 had about 2 million couples
- We said it was 1 in a million to get a match on the eyewitness description
- We know that 1 couple who matches the description is guilty

Answering the right question

Using what we have learned about Bayes' Rule we can answer:

"How likely is the Collins couple to be innocent given that they match the eyewitness description?"

And, importantly, find out how Malcolm Collins did upon his appeal to the California Supreme Court.

See Bueno de Mesquita and Fowler, ch 15

Thinking about applying Bayes' rule

To apply Bayes' rule, answer each of the following three questions:

1. What is the base rate?
2. How rare is the evidence?
3. How relevant is this evidence?

The key takeaway of Bayes' rule is new evidence "updating" prior beliefs or evidence.

Matching the description in our world today

"Updating" is not a concept that exists in isolation of sitting down to solve Bayes' Rule on your test or in this class.

The idea of "updating" is a process that repeats in the real world around us.

For example:

- Strategies for fundraising
- A/B testing in businesses
- Dating
- Medical tests & medical practice

Also—very famous examples such as the SS Central America and Alan Turing cracking the Nazi Enigma

For more examples, look here for 3Blue1Brown's videos on Bayes' Rule.

Telling the story

1. Write a sentence describing the error that the original mathematician made in the Collins case.
2. Using the Ali example, write a sentence stating the probability of fitting the description given that the individual is an economist, provide both the relevant vocabulary term and how that is different from the posterior.
3. Write a sentence describing in words the posterior for the Ali example.