Review Session 1

API-201, 9.10.21
Sophie Hill

About me

- PhD student in the Gov department
- Researching the politics of the climate crisis
- I love teaching statistics!! (seriously!)
- I have an extremely cute cat

About me

- PhD student in the Gov department
- Researching the politics of the climate crisis
- I love teaching statistics!! (seriously!)
- I have an extremely cute cat

Introductions

Tell me:

- your name
- what you think I should name my cat

Review Session norms

The main thing I ask of you in these Review Sessions is to **help me help you!**

Review Session norms

More concretely:

- **Let me know** which concepts you are finding difficult
  - DM me on Slack, or come to Office Hours

- **Provide feedback** on how these sessions could be more helpful to you
  - Use the anonymous feedback form (link is on Canvas home page too)

Problem Set 1

Problem Set 1

PS1 is posted on Canvas!

Due: **Tuesday** 9/14 at **8:30am**

From the syllabus:

> You are free (and encouraged) to discuss problem sets with your classmates. However, you must hand in your own unique written work. Any copy/paste of another's work is plagiarism. In other words, you can work with your classmate(s), sitting side-by-side and going through the problem set question-by-question, but **you must each type your own answers**. Your answers may be similar but they must not be identical, or even identical-ish. … If you have questions about the degree of collaboration allowed or about any other aspect of the Academic Code, please come to see us.

Problem Set 1: Preview

Q1) Teleworking

Conditional probabilities

Q2) COVID testing

Conditional probabilities;

Bayes rule

Q3) Estimating probabilities

Principles of thinking probabilistically

Q4) Mexico pensions

Leakage / undercoverage

Problem Set 1: Preview

Q1) Teleworking

Conditional probabilities

Q2) COVID testing

Conditional probabilities;

Bayes rule

Heads up: this question is the longest and (in my opinion) the hardest...

Q3) Estimating probabilities

Principles of thinking probabilistically

Q4) Mexico pensions

Leakage / undercoverage

Plan for today

Review:

- Types of probability
- Probability rules
- Bayes rule

Practice problems!

Probability Review

Probability Review

$P(A)$

$P(A \,\&\, B)$

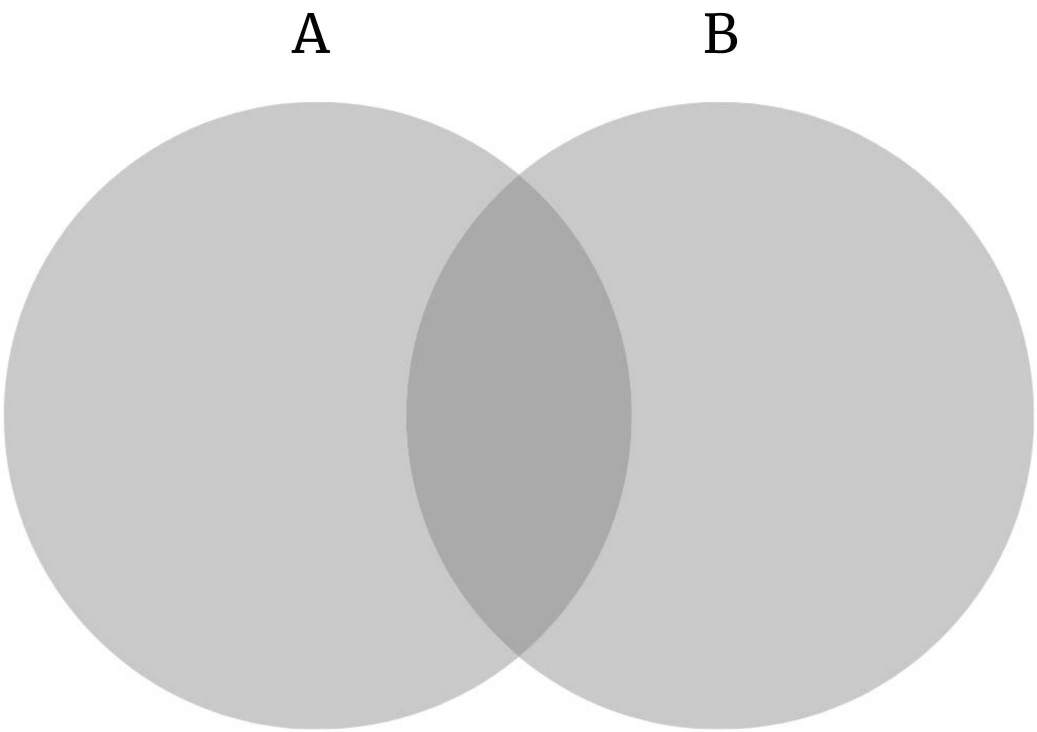
$P(A \mid B)$

Probability Review

Marginal probability	P(A)	"Probability of A"
Joint probability	P(A & B)	"Probability of A and B"
Conditional probability	P(A | B)	"Probability of A given B"

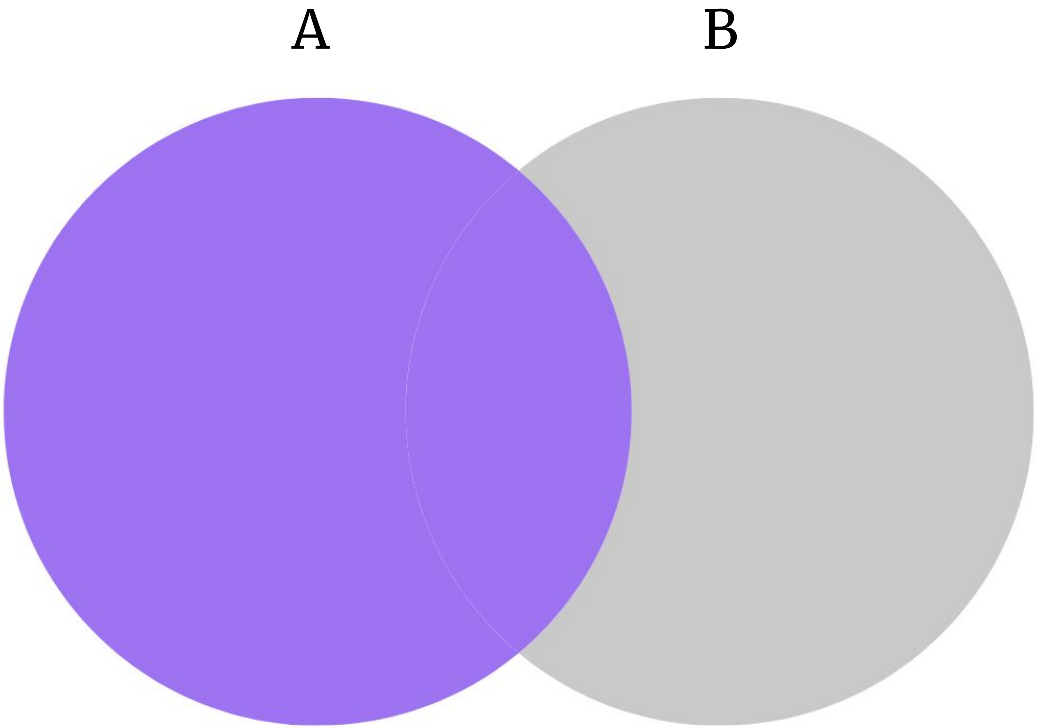
Intuitions

Which area represents P(A)?



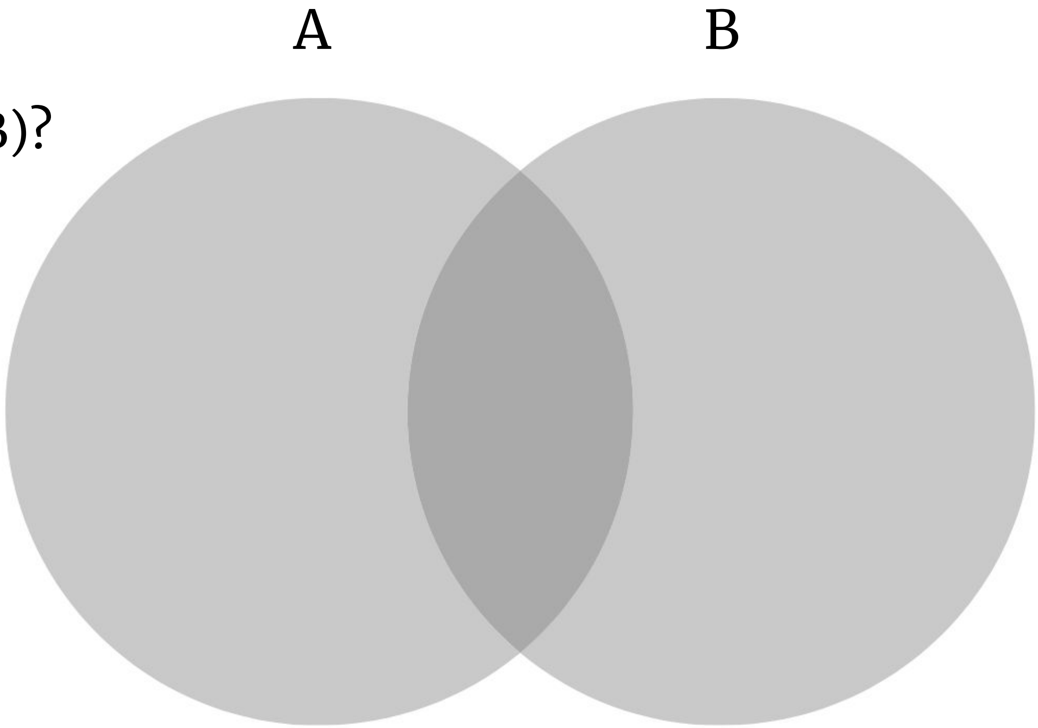
Intuitions

Which area represents P(A)?



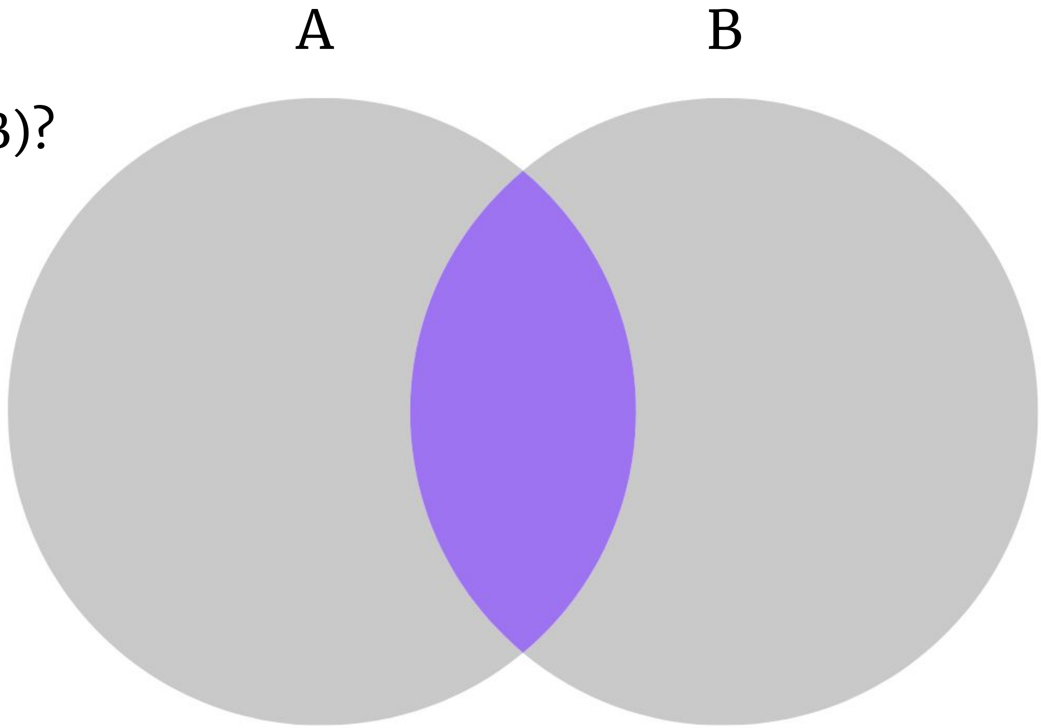
Intuitions

Which area represents P(A & B)?



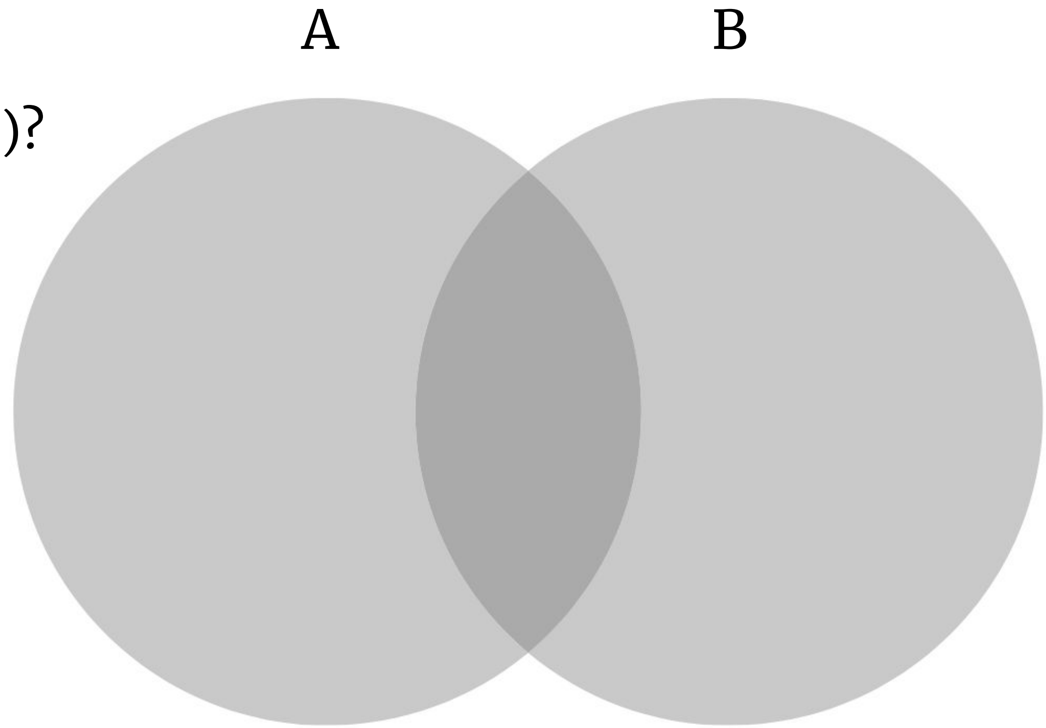
Intuitions

Which area represents P(A & B)?



Intuitions

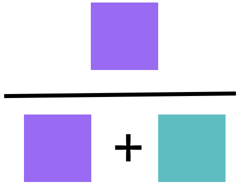
Which area represents $P(A \mid B)$?

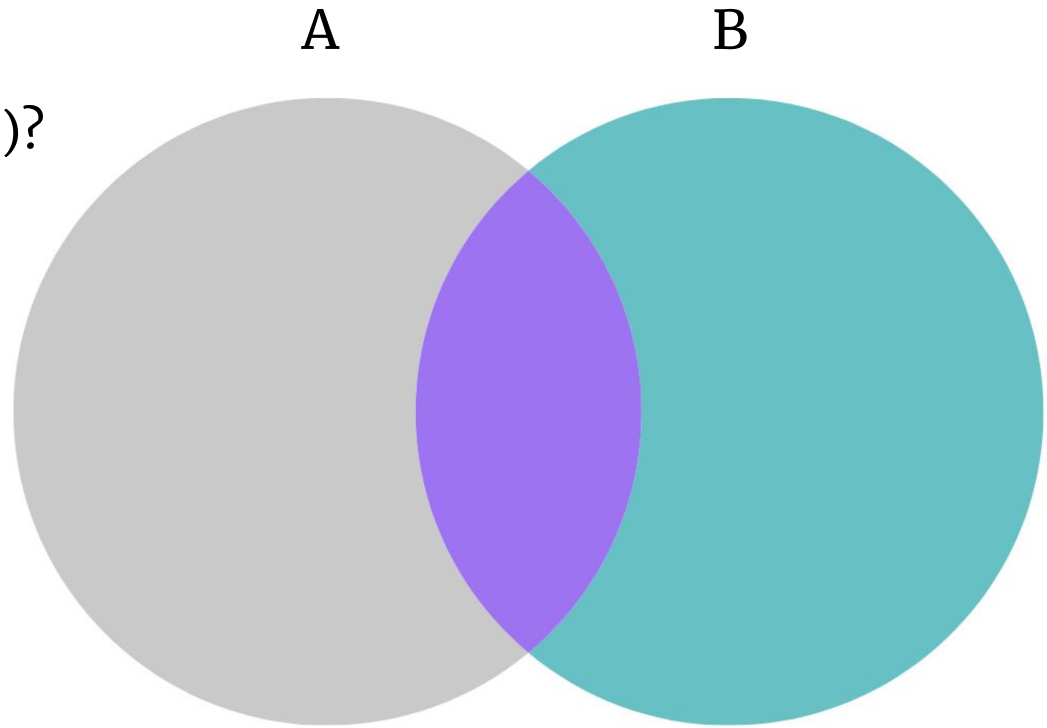


Intuitions

Which area represents P(A | B)?

P (A | B) =





A note on notation...

All of these are equivalent:

$P(\text{not } A)$

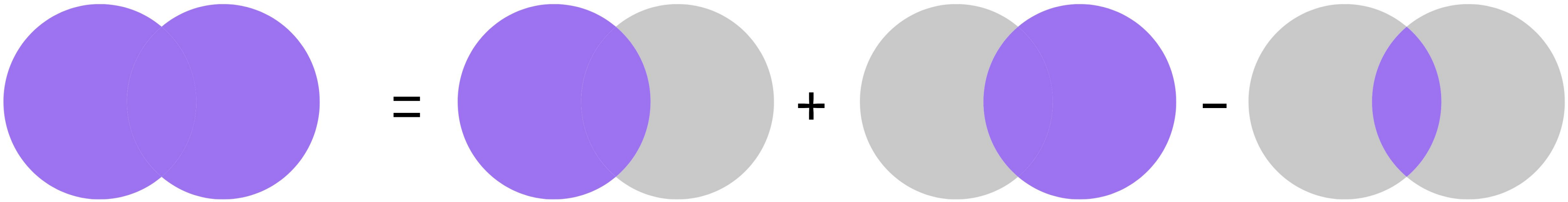
$P(\neg A)$

$P(A^c)$

Probability Rules

Addition rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \,\&\, B)$$



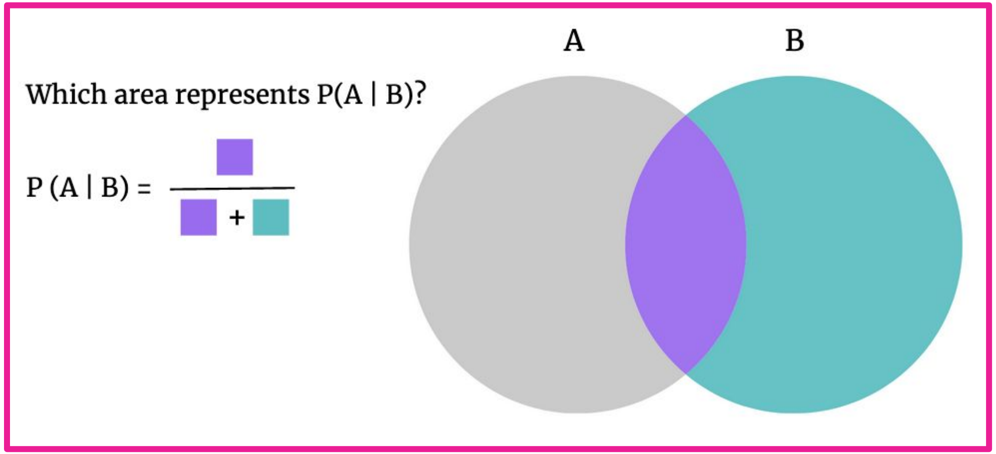
Complement rule

$$P(A^c) = 1 - P(A)$$

Multiplication rule

$$P(A \,\&\, B) = P(A \mid B) * P(B)$$

Multiplication rule



$$P(A \,\&\, B) = P(A \mid B) * P(B)$$

Wait... that looks familiar! Let's rearrange:

$$P(A \mid B) = P(A \,\&\, B) \,/\, P(B)$$

Bayes Rule

Bayes Rule

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

Bayes Rule

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

Or, equivalently:

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

Bayes Rule: Example 1

$$
\begin{aligned}
P(COVID|cough) &= \frac{P(cough|COVID) \cdot P(COVID)}{P(cough)} \\
&= \frac{0.6 \cdot 0.01}{0.05} \\
&= 0.12
\end{aligned}
$$

Bayes Rule: Example 1

$$P(COVID|cough) = \frac{P(cough|COVID) \cdot P(COVID)}{P(cough)}$$
$$= \frac{0.6 \cdot 0.01}{0.05}$$
$$= 0.12$$

What if we did the same calculation for loss of smell instead of a cough?

You can assume P(no smell | COVID) is also 0.6.

Bayes Rule: Example 2

$$
\begin{aligned}
P(COVID \mid no\ smell) &= \frac{P(no\ smell \mid COVID) * P(COVID)}{P(no\ smell)} \\
&= \frac{0.6 \cdot 0.01}{0.008} \\
&= 0.75
\end{aligned}
$$

Bayes Rule: Example 2

$$
\begin{aligned}
P(\mathit{COVID} \mid \mathit{no\ smell}) &= \frac{P(\mathit{no\ smell} \mid \mathit{COVID}) * P(\mathit{COVID})}{P(\mathit{no\ smell})} \\
&= \frac{0.6 \cdot 0.01}{0.008} \\
&= 0.75
\end{aligned}
$$

What is the intuitive explanation for the difference between P(COVID|cough) and P(COVID|no smell)?

3 clues that you might need to use Bayes Rule!!

1. You have **P(B|A)** but the question is asking for **P(A|B)**

2. The question talks about "**updating**" a subjective probability to incorporate new information

3. The question mentions the "**base rate fallacy**"

Practice Problems

Practice Problem 1: The Energy Grid

We have a diversified energy grid with wind, solar, and hydroelectric power.

- The probability that the wind plant is operating is 0.75
- The probability that the solar plant is operating is 0.5
- The probability that hydroelectric plant is operating is 0.8

Each power source operates independently of the others.

(A) What is the probability that we will experience a black out?

Practice Problem 1: The Energy Grid

We have a diversified energy grid with wind, solar, and hydroelectric power.

- The probability that the wind plant is operating is 0.75 $P(W) = 0.75$
- The probability that the solar plant is operating is 0.5 $P(S) = 0.5$
- The probability that hydroelectric plant is operating is 0.8 $P(H) = 0.8$

Each power source operates independently of the others.

(A) What is the probability that we will experience a black out?

Practice Problem 1: The Energy Grid

P(Blackout)

$= P(\neg W \,\&\, \neg S \,\&\, \neg H)$

$= P(\neg W) * P(\neg S) * P(\neg H)$

$= (1\text{-}0.75) * (1\text{-}0.5) * (1\text{-}0.8)$

$= 0.25 * 0.5 * 0.2$

$= 0.025$

N.B.: the assumption that the failure of each power source is **independent** from the others is what allows us to do this!

Practice Problem 1: The Energy Grid

(B) We see that the grid has power. What is the probability that the solar plant is operational?

Practice Problem 1: The Energy Grid

(B) We see that the grid has power. What is the probability that the solar plant is operational?

$P(\text{Solar} \mid \text{Power})$

$= P(\text{Solar} \,\&\, \text{Power}) \,/\, P(\text{Power})$

$= P(\text{Solar}) \,/\, P(\text{Power})$

$= 0.5 \,/\, 0.975 \approx 0.513$

Practice Problem 1: The Energy Grid

(C) What is the probability that both the wind and hydroelectric plants are operational? (The power is still on.)

Practice Problem 1: The Energy Grid

(C) What is the probability that both the wind and hydroelectric plants are operational? (The power is still on.)

$P(W \,\&\, H \mid \text{Power})$

$= P(W \,\&\, H) \,/\, P(\text{Power})$

$= P(W) \times P(H) \,/\, P(\text{Power})$

$= 0.75 \times 0.8 \,/\, 0.975 \approx 0.615$

Practice Problem 2: The Sally Hayes case

Sally Hayes was convicted of killing 2 of her children in 1999.

At trial, an expert witness said that the chance of 1 infant dying of natural causes was 1/8300, so the chance of 2 infants dying of natural causes was 1/8300 * 1/8300, or about 1 in 73 million.

The prosecution argued that this probability is so small that it is virtually impossible, and so Sally Hayes must be guilty.

Mother given life for baby murders

Practice Problem 2: The Sally Hayes case

Sally Hayes was convicted of killing 2 of her children in 1999.

At trial, an expert witness said that the chance of 1 infant dying of natural causes was 1/8300, so the chance of 2 infants dying of natural causes was 1/8300 * 1/8300, or about 1 in 73 million.

The prosecution argued that this probability is so small that it is virtually impossible, and so Sally Hayes must be guilty.

Mother given life for baby murders

Do you agree with the expert witness's calculations?

Practice Problem 2: The Sally Hayes case

Let us suppose that the expert witness was correct about the chance of one infant dying of natural causes: P(first child SIDS) = 1/8300

The witness said that P(2 SIDS in same family) = 1/8300 * 1/8300

Practice Problem 2: The Sally Hayes case

Let us suppose that the expert witness was correct about the chance of one infant dying of natural causes: P(first child SIDS) = 1/8300

The witness said that P(2 SIDS in same family) = 1/8300 * 1/8300

But this is only true if the probability of the first and second child dying of natural causes are **independent**. But researchers believe there may be a genetic component to SIDS, so that P(2 SIDS in same family) is *much less* than 1/8300 * 1/8300, perhaps closer to 1/8300 * 1/100.

Practice Problem 2: The Sally Hayes case

Sally Hayes was convicted of killing 2 of her children in 1999.

At trial, an expert witness said that the chance of 1 infant dying of natural causes was 1/8300, so the chance of 2 infants dying of natural causes was 1/8300 * 1/8300, or about 1 in 73 million.

The prosecution argued that this probability is so small that it is virtually impossible, and so Sally Hayes must be guilty.

Mother given life for baby murders

Do you agree with the prosecution's argument?

Practice Problem 2: The Sally Hayes case

The prosecution focused on **P(2 SIDS).** But the probability that Hayes was innocent is actually **P(2 SIDS | 2 infant deaths).** Let's use Bayes Rule:

$$P(SIDS|2deaths) = \frac{P(2deaths|SIDS) \cdot P(SIDS)}{P(2deaths|SIDS) \cdot P(SIDS) + P(2deaths|\neg SIDS) \cdot P(\neg SIDS)}$$

$$= \frac{1 \cdot (1/830000)}{1 \cdot (1/830000) + P(2deaths|\neg SIDS) \cdot (1 - 1/830000)}$$

Practice Problem 2: The Sally Hayes case

The prosecution focused on **P(2 SIDS).** But the probability that Hayes was innocent is actually **P(2 SIDS | 2 infant deaths).** Let's use Bayes Rule:

$$P(SIDS|2deaths) = \frac{P(2deaths|SIDS) \cdot P(SIDS)}{P(2deaths|SIDS) \cdot P(SIDS) + P(2deaths|\neg SIDS) \cdot P(\neg SIDS)}$$

$$= \frac{1 \cdot (1/830000)}{1 \cdot (1/830000) + P(2deaths|\neg SIDS) \cdot (1 - 1/830000)}$$

$$= \frac{1 \cdot (1/830000)}{1 \cdot (1/830000) + 0.0000046 \cdot (1 - 1/830000)}$$

$$= 0.21$$

Practice Problem 2: The Sally Hayes case

The prosecution focused on **P(2 SIDS).** But the probability that Hayes was innocent is actually **P(2 SIDS | 2 infant deaths).** Let's use Bayes Rule:

$$P(SIDS|2deaths) = \frac{P(2deaths|SIDS) \cdot P(SIDS)}{P(2deaths|SIDS) \cdot P(SIDS) + P(2deaths|\neg SIDS) \cdot P(\neg SIDS)}$$

$$= \frac{1 \cdot (1/830000)}{1 \cdot (1/830000) + P(2deaths|\neg SIDS) \cdot (1 - 1/830000)}$$

$$= \frac{1 \cdot (1/830000)}{1 \cdot (1/830000) + 0.0000046 \cdot (1 - 1/830000)}$$

$$= 0.21$$

The prosecution implied that the probability of Hayes being innocent was **1 in 73 million**, but our calculation says it is actually **1 in 5**!

Nice work everyone!!