Module 2B : Probability

Frequentist and Bayesian building blocks

<u>Agenda:</u>

- Bayesian Probability
  - Structure of Bayes' Theorem: $P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[B|A]P[A]}{P[B]}$
  - The Monty Hall Problem: illustrating the philosophical difference with Frequentist camp - ability to <u>update</u> probability with new information
  - Examples:
    - Pedigree Analysis

Bayes probability definition:

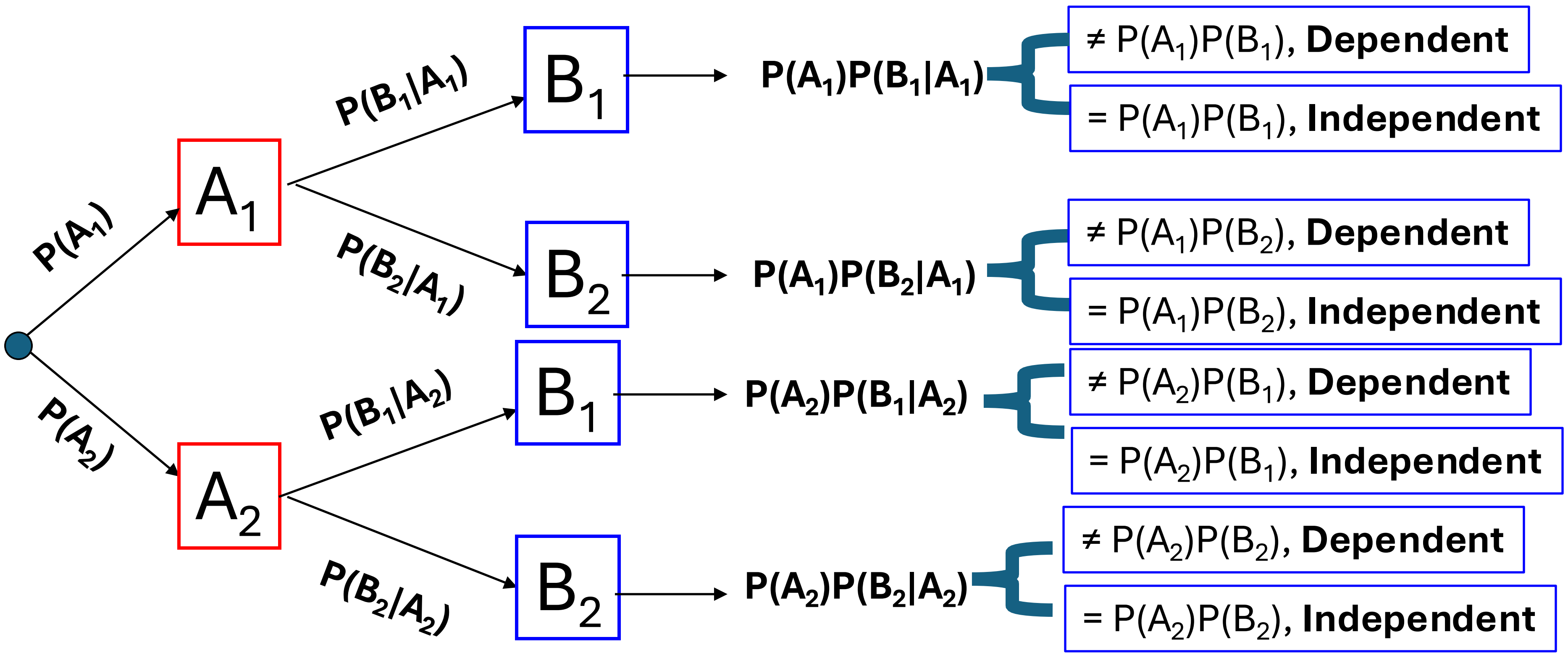
“Probability is a measure of belief associated with the occurrence of an event “

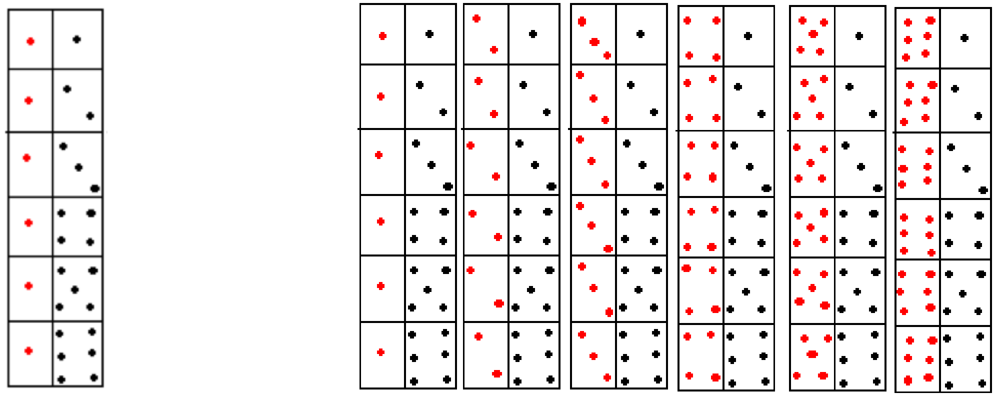
- Probability is subjective and can be **updated** when new information is available. Start with a **PRIOR** probability (or belief) and update it with a **LIKELIHOOD** and end with a **POSTERIOR** probability.

Ex. Will it rain tomorrow?

- **Frequentist Answer**: probability of rain in a place is sum of days of rain over time.
- **Bayesian Answer**: it will depend on whether it is Summer or Winter.

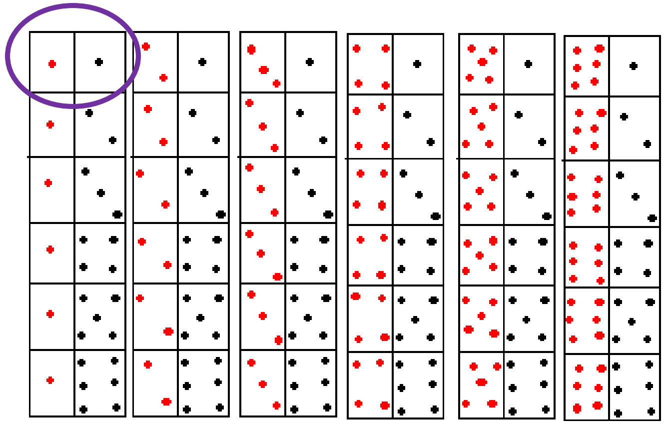
1st Variable 2nd Variable $P(A \cap B)$ **Dependent or Independent?**



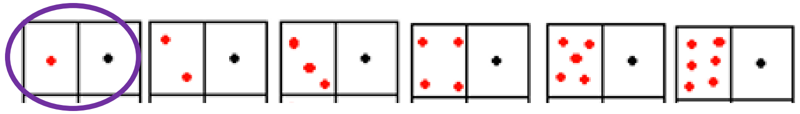


P(A∩B) <u>Versus</u> P(A|B)

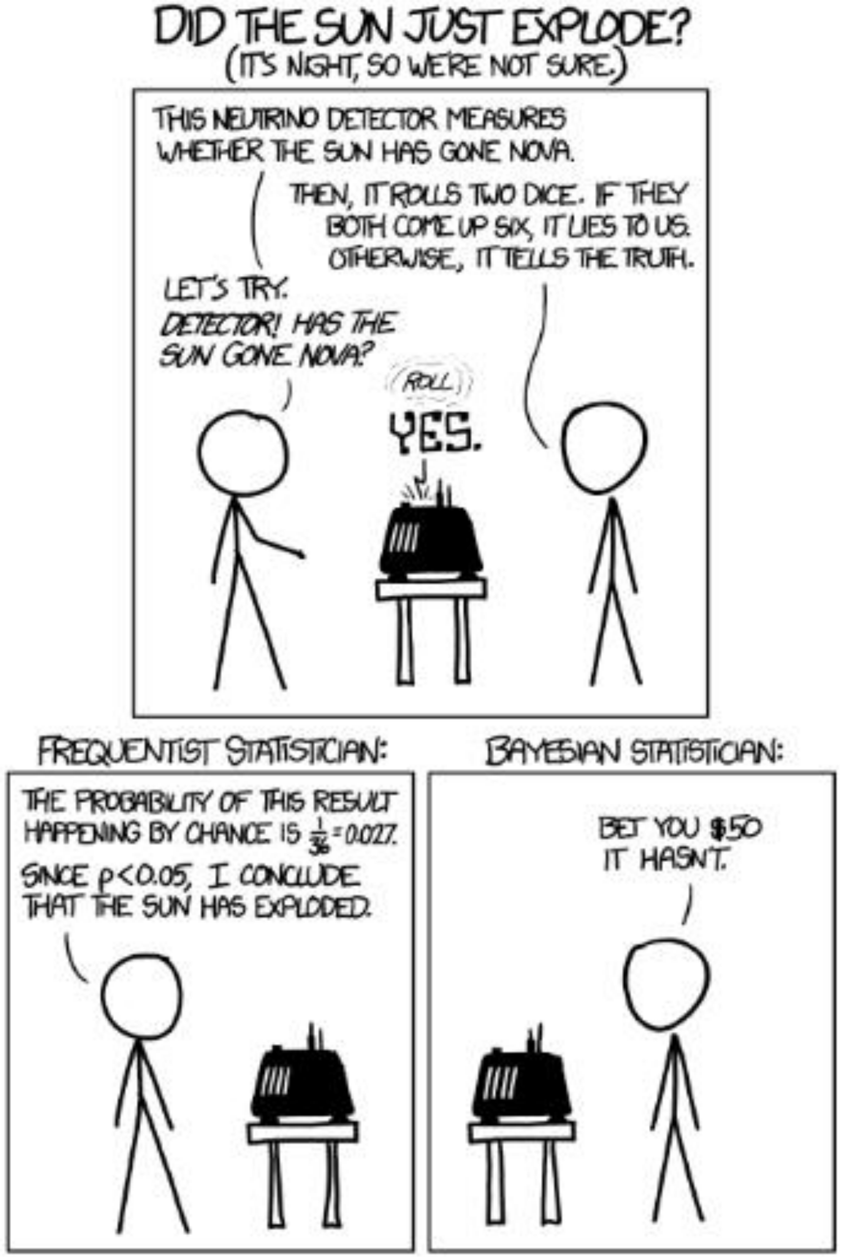
$P(red=1 \cap black=1) = 1/36$



$P(red=1|black=1) = 1/6$



- **Bayes inference is different from what we normally do**
- Probability is a way of quantifying uncertainty and assumes that with random sampling we can infer meaningfully from repeated experiments (**Frequentist)**
  - With enough information, and properly designed sampling, we can accurately estimate the **unknown but constant parameter value with increasing precision**
- Some phenomenon make sense under frequentist definition….
  - If we toss a fair coin, what is the probability of 10 heads in a row?
  - If we assign treatments randomly to subjects, what is the probability that a sample mean difference between treatments will be greater than 20%?
  - What is the probability that, given the null hypothesis is true, of obtaining data that is at least as extreme as that observed?
- Some phenomenon don't….
  - What is the probability that polar bears will be extinct in 30 years?
  - What is the probability that hippos are sister group to whales?
- There is no random trials inherent in certain meaningful questions (Hippos either ARE or ARE NOT sister groups; polar bears will either be EXTINCT or NOT in 30 years)
  - **The unknown parameter value is not a constant truth; it has inherent uncertainty** and no matter how much information you get, it will not have a constant value.
  - Probability is subjective and can be updated when new information is available. You start with a prior probability and update it with a likelihood and end with a posterior probability.



http://xkcd.com/1132/

Let's warm up our brains with a bit of a riddle (it uses a similar 'updated knowledge' logic that the Monty Hall problem uses):

Inside of a dark closet are five hats: three blue and two yellow. Knowing this, three students go into the closet, and each selects a hat in the dark and places it unseen upon their head.

Once outside the closet, no one can see their own hat. The first student looks at the other two, thinks, and says, "I cannot tell what color my hat is." The second hears this, looks at the other two, and says, "I cannot tell what color my hat is either." The third student has an eye infection and temporarily can't see. The temporarily blind student says, "Well, I know what color my hat is." What color is their hat?

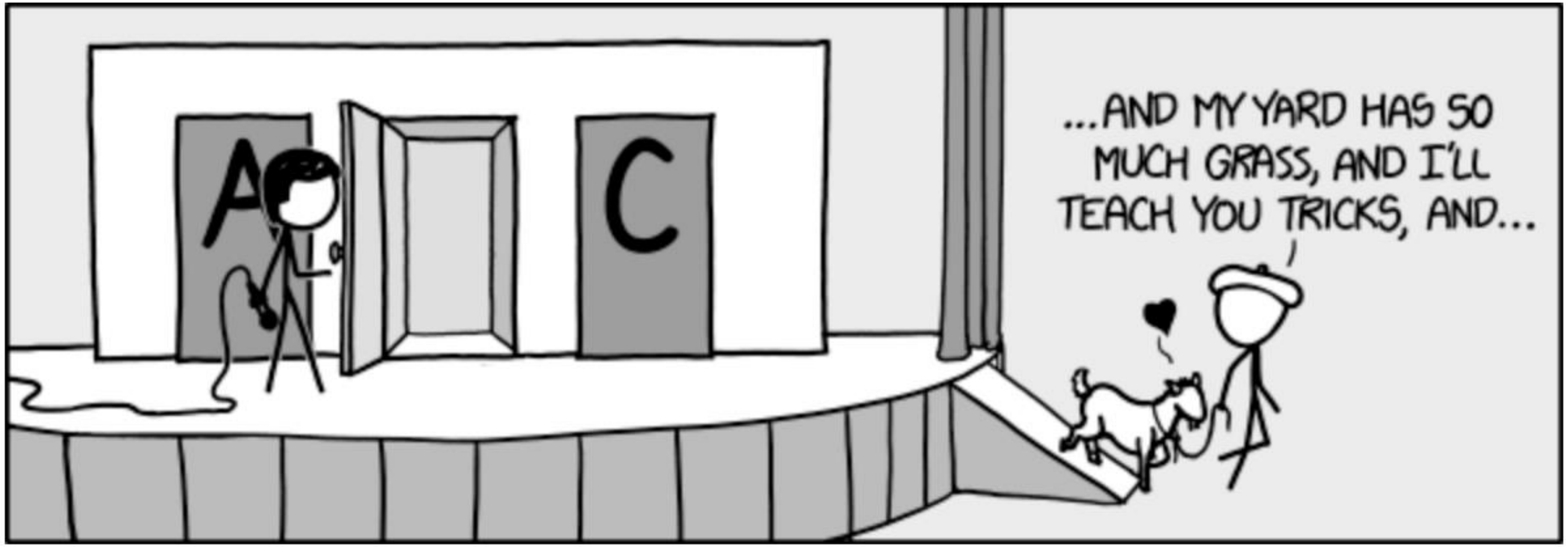
Monty Hall Problem

https://montyhall.io/

- **1,000,000. P(picking the car) = 1/1,000,000.**
- **Pick a door – two doors have goats, and one door has a car**
- **Monty Hall then opens one of the remaining doors and ask….**

"Do you want to switch your door or stay with the door your have chosen?"

Should you switch or stay?



https://xkcd.com/1282/

<u>Good overview of Monty Hall problem:</u>

- https://www.statisticshowto.com/probability-and-statistics/monty-hall-problem/
- https://www.explainxkcd.com/wiki/index.php/File:monty_hall.png

The TV show "MythBusters" have a short segment (first 15 minutes or so) on Monty Hall problem:

https://www.youtube.com/watch?v=oWWNZ_eciGI

$$\boxed{P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}}$$

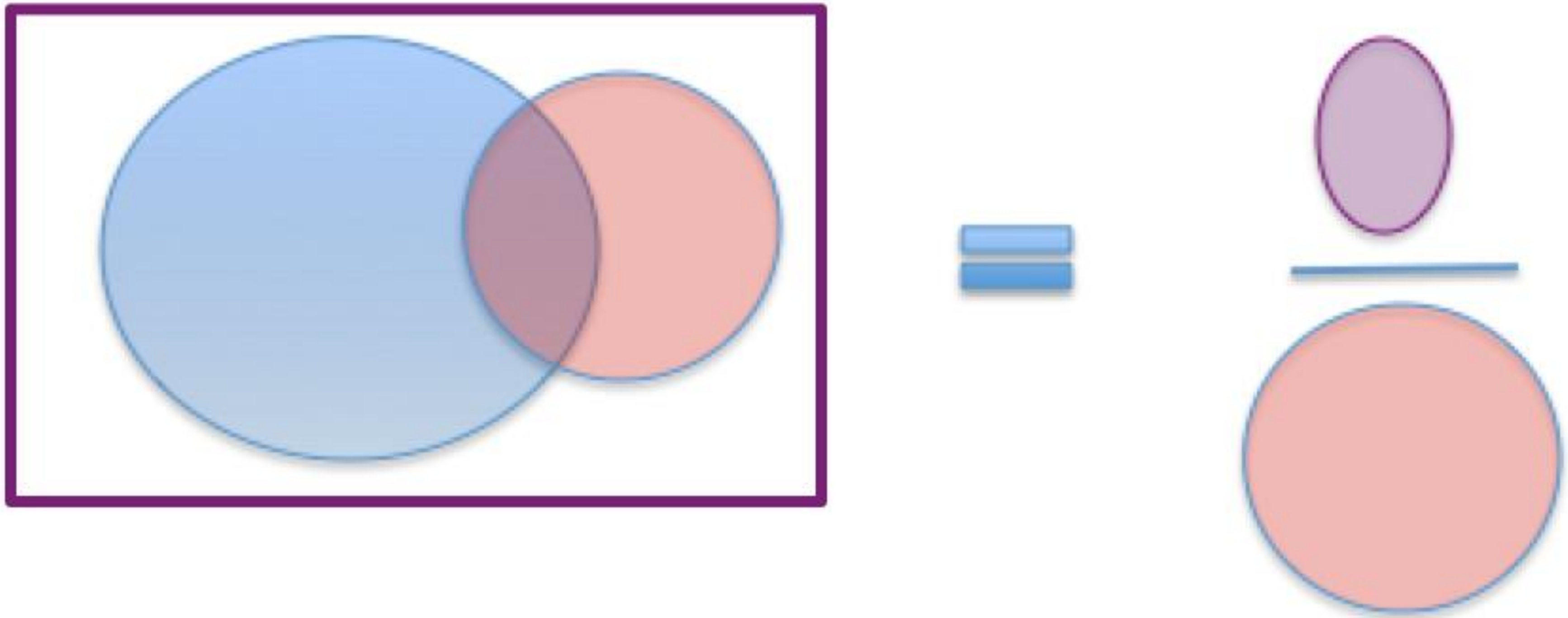
You can interpret Bayes as reducing the state space, like the following equations which use the proportion of A intersecting with B over the WHOLE universe (i.e., black die and red die both equal 1 is 1/36) and then reduce the proportion by dividing by the probability of the first event

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} \qquad P[B \mid A] = \frac{P[A \cap B]}{P[A]}$$

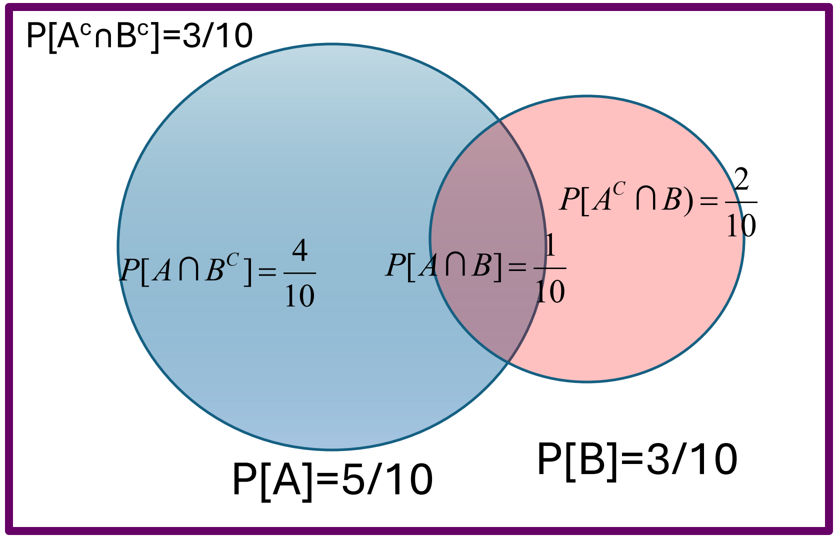
You might notice that both equations involve the SAME **numerator** whereas the **denominator** changes based on what event has happened <u>first</u> i.e. what we already know and what we still want to know.

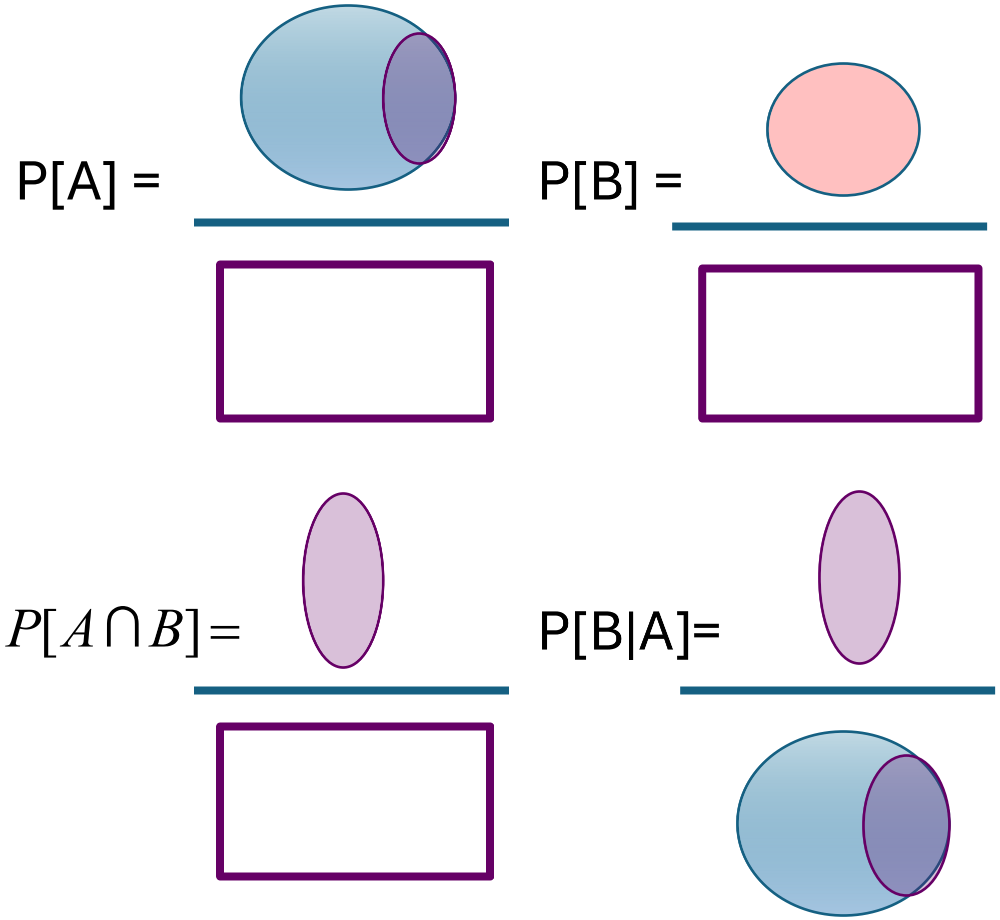
Sophisticated rearrangement of the multiplication rule…..

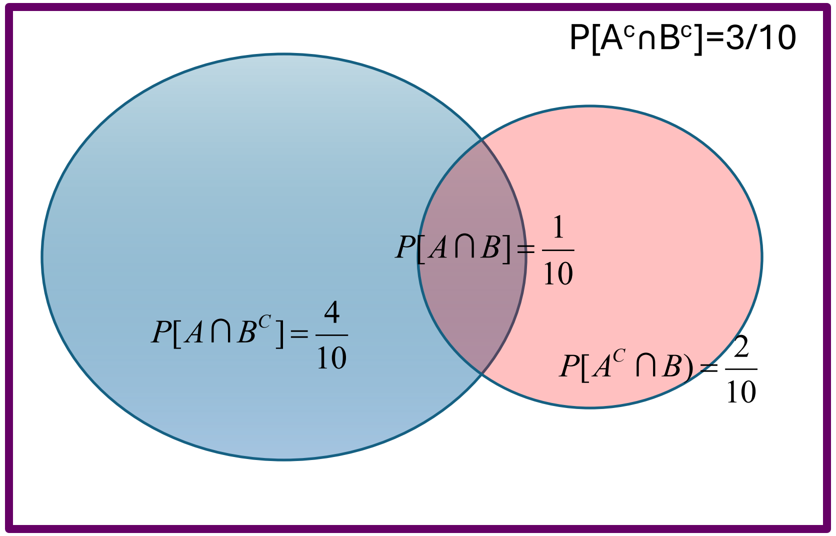
$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[B|A]P[A]}{P[B]}$$



$$P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}$$







$$\boxed{P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}}$$

P[A]=5/10

P[B] = 3/10

$$P[A \cap B] = \frac{1}{10}$$

P[A|B]= $\frac{1/10}{3/10} = \frac{1}{3}$

P[B|A]= $\frac{1/10}{5/10} = \frac{1}{5}$

<u>Bayes' theorem in words:</u> The conditional probability of **A (the hypothesis) given B (the data)** is the conditional probability of **B (data)** given **A (hypothesis)** scaled by the relative probability of **A compared to B**

P[Hypothesis|DATA] = $\dfrac{\textbf{P[Data|Hypothesis]P[Hypothesis]}}{\textbf{P[Data]}}$

The **PRIOR** hypothesis:
The original probability of the hypothesis without any additional information

The **LIKELIHOOD** interpreted as:
P(observation GIVEN the hypothesis)

$$P[A\mid B]=\frac{P[A\cap B]}{P[B]}=\frac{P[A]P[B\mid A]}{P[B]}$$

the **POSTERIOR probability** interpreted as the P(hypothesis GIVEN the observation)

The **observation/data/ Evidence** that has been observed

Example: Suppose we want to calculate the probability that someone will die of lung cancer **given** that they smoke. We study a cohort of individuals, determining who smoke and which ones do not and track them until they died. Then we could calculate the number of smokers who died of lung cancer.

There is an easier way, however....

USE BAYES

$$\boxed{P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}}$$

Specify the question: What is event 'A' and what is event 'B'? A=lung cancer death; B=smoker

Probabilities	Where/how do we get them?
P[Death due to lung cancer | Smoker]	
P[Death due to lung cancer]	estimated from death records
P[Smoker]	Polling appropriate population
P[Smoker | Death due to lung cancer]	estimated from death records

Example:

$$P[A \mid B] = \frac{P[A]P[B \mid A]}{P[B]}$$

Specify the question: What is event 'A' and what is event 'B'?

Probabilities	Where/how do we get them?
P[Death due to lung cancer | Smoker]	
P[Death due to lung cancer]	estimated from death records
P[Smoker]	Polling appropriate population
P[Smoker | Death due to lung cancer]	estimated from death records

P[Smoker] = 0.5
P[Smoker | Death due to lung cancer] = 0.9
P[Death due to lung cancer] = 0.3

P[Death due to lung cancer | Smoker] = $\frac{0.9 \times 0.3}{0.5}$ = 0.54

Note: Using Bayes, also gives us a bonus calculation: **P[Non-Smoker | Death due to lung cancer] = 0.1**
P[Death due to lung cancer | Non-smoker] = $\frac{0.1 \times 0.3}{0.5}$ = 0.06

Jim was bitten by a mosquito during his trip to South Sudan. He gets tested for Malaria. What is the probability that Jim has Malaria given a positive test result, considering the following facts: Malaria occurs in 1 in 1,000 people in South Sudan, the test for Malaria has an 85% probability of detecting Malaria, but there is also a 10% false positive rate as well.

Let's gather our information.

P[malaria]=1/1000 = 0.001
P[no malaria]=999/1000 =0.999
P[pos test|malaria] = 0.85
P[pos test|no malaria] =0.10

$$\text{P[malaria| positive test]} = \frac{\text{P[pos test|malaria]P[malaria]}}{\text{P[positive test]}} = \frac{0.85 * 0.001}{(0.85 * 0.001 + 0.10 * 0.999)} = 0.00844$$

There are two ways to have a positive test: because you have malaria or because the test is inaccurate.

P[positive test] = P[Positive|malaria]*P[malaria]+P[positive|no malaria]*P[no malaria]
= 0.85*0.001+0.10*0.999

1/1000 → 8.4/1000

The words “you have won” are sometimes used in emails (5%). Most emails are not spam, but let’s say a person has a 2% chance of receiving spam mail. <u>In addition, 80% of spam emails have the words “you have won” in them compared to 10% of non-spam emails</u>. Use Bayes’ Theorem to calculate the probability that an email with the words “you have won” is spam.

There are a handful of other probabilities and terms that are often given:

Specificity = P[negative test | don’t have condition]

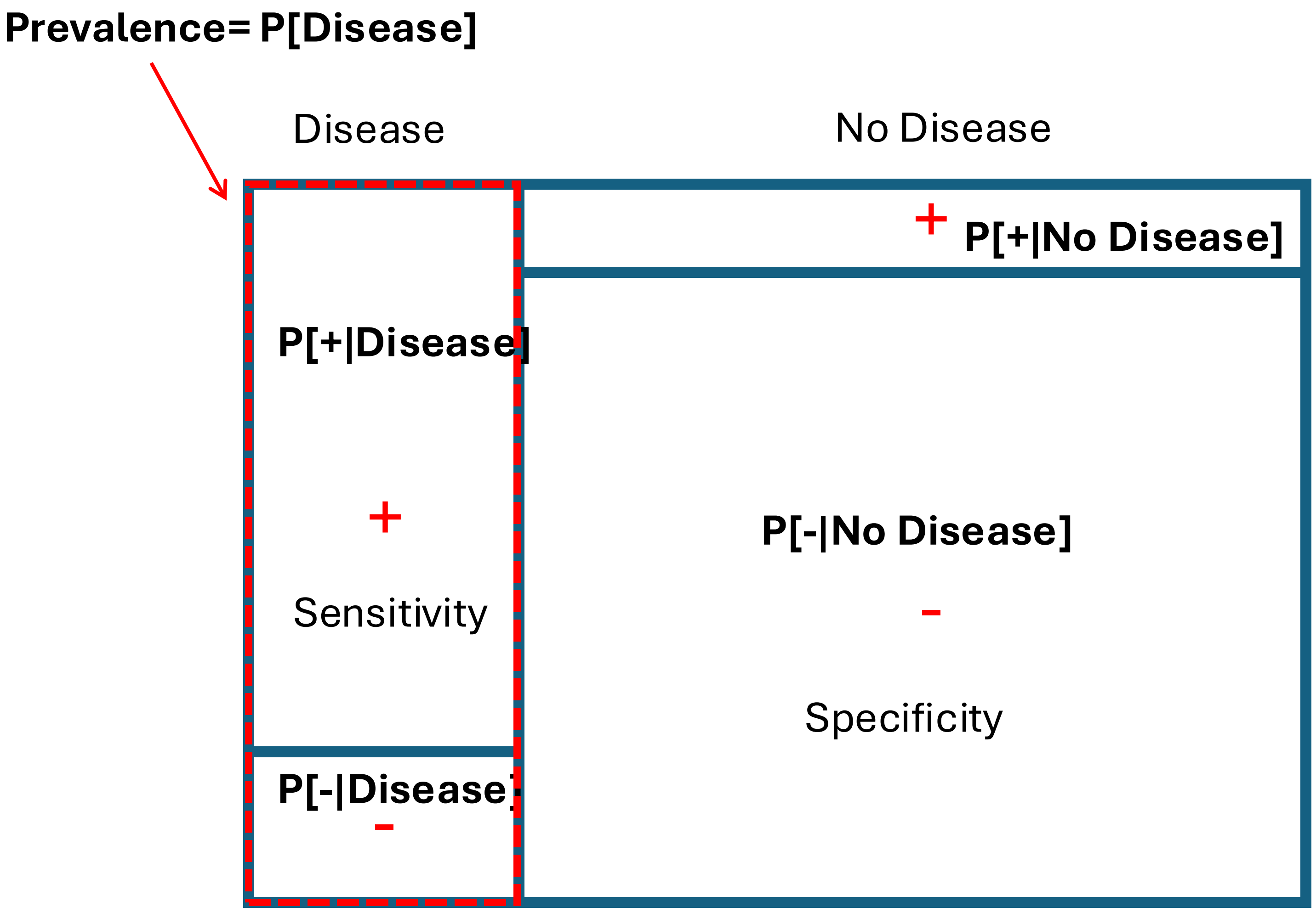
From this, you can get:

False alarms = 1-Specificity = P[positive test | don’t have condition]

<u>We have:</u>

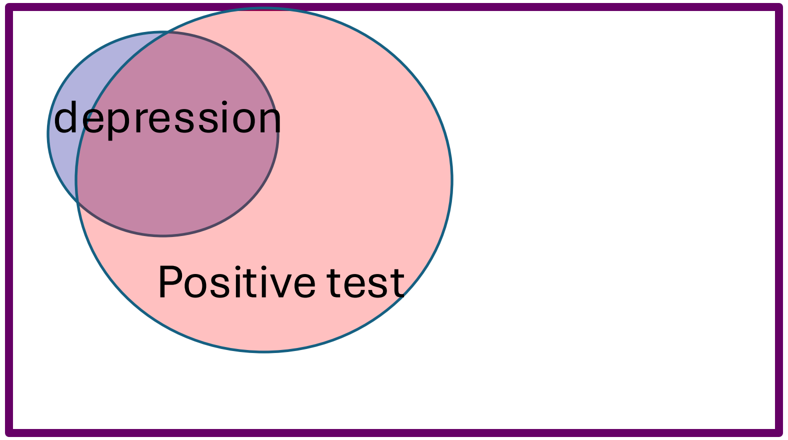
Sensitivity = Likelihood = P[positive test | have condition]

Prevalence = Prior = P[condition in general pop.]



$$P[\text{Disease}|+] = \frac{P[+|\text{Disease}]P[\text{Disease}]}{P[+]}$$

MOST POSITIVES ARE FALSE POSITIVES



Blue = depression proportion
Red= positive test for depression

P[depression]=0.1

P[negative Test|No Depression]=0.8

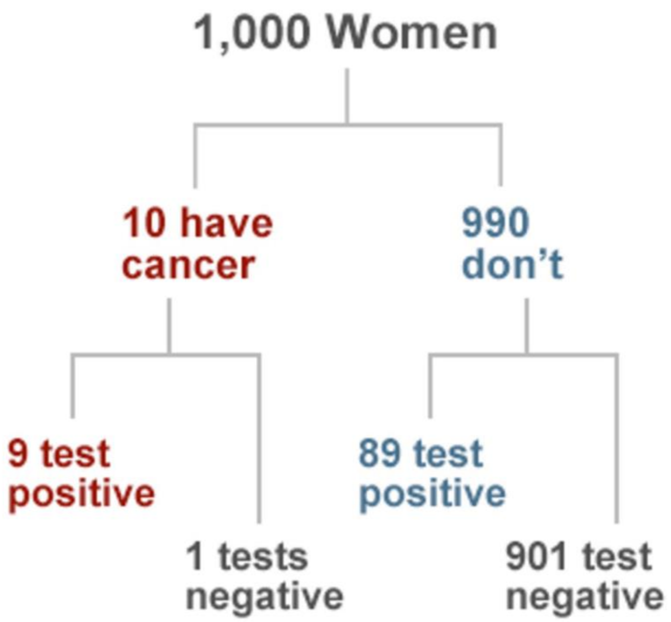
P[positive test | depression] = 0.9

P[depression | positive test] = $\frac{0.9*0.1}{(0.9*0.1 + 0.8*0.9)}$ = 0.33

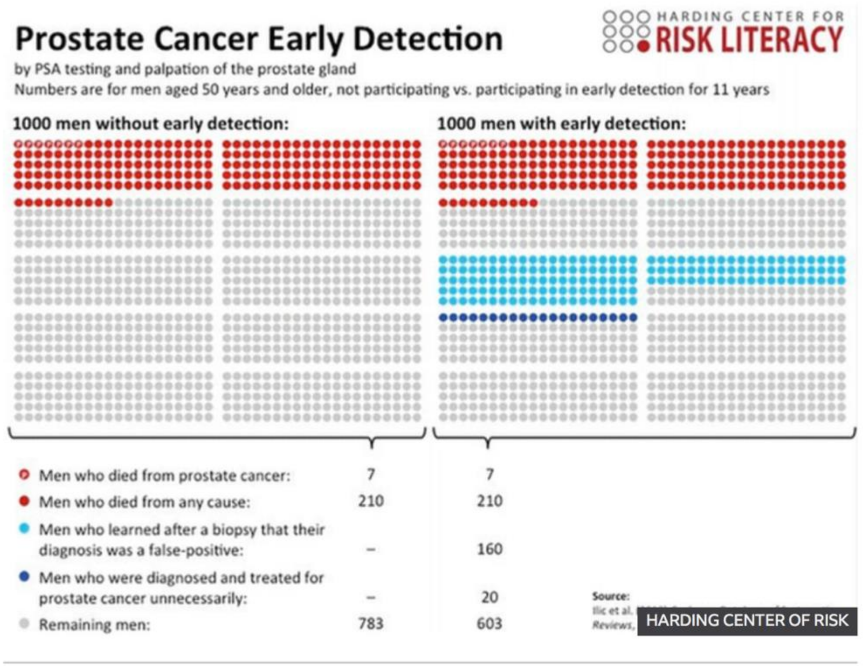
Don’t mix up P[A|B] with P[B|A] … that’s a different mistake called “The Prosecutor’s Fallacy”

Diagnostic Example

- We don’t know <u>the truth</u> about the disease state, we only have access to tests (data)
- There is no such thing as a perfect test: every test has a trade off between sensitivity and specificity
- **For a test to be useful,** it is not necessary for both sensitivity and specificity to be high, **but it IS** necessary for the user (health care provider) to interpret positives/negatives correctly.



https://www.bbc.com/news/magazine-28166019



Major Caveat

- Prior probabilities and prevalence are established with certain ancestries in mind

- For genetic information, >85% of our major databases are European ancestry ← this is a **massive problem**

- It is challenging to get prevalence rates in other ancestries.

What do I mean when I say, “Bayes allows us to easily update our information?”

Break out room: 4 scenarios

Each group will do the following three things:

(i) fill out a Bayes table using a cohort of 10,000 people

(ii) compute PPV and NPV, and

(iii) discuss how changing **prevalence** would alter the posterior.

Quick reference

- **PPV** = $P(Disease \mid +) = \frac{TP}{TP+FP}$
- **NPV** = $P(No\ disease \mid -) = \frac{TN}{TN+FN}$
- Use a **10,000** population for clean integers.
- Using counts instead of percentages can be helpful!

Break out room: 4 scenarios

A. Rare disease, decent test:
- **Prevalence:** 1%
- **Sensitivity:** 90%
- **Specificity:** 95%

B. Higher prevalence, weaker specificity
- **Prevalence:** 10%
- **Sensitivity:** 95%
- **Specificity:** 90%

C. High specificity, moderate sensitivity
- **Prevalence:** 20%
- **Sensitivity:** 80%
- **Specificity:** 99%

D. Screening: Catch almost everything
- **Prevalence:** 2%
- **Sensitivity:** 99%
- **Specificity: 85**%

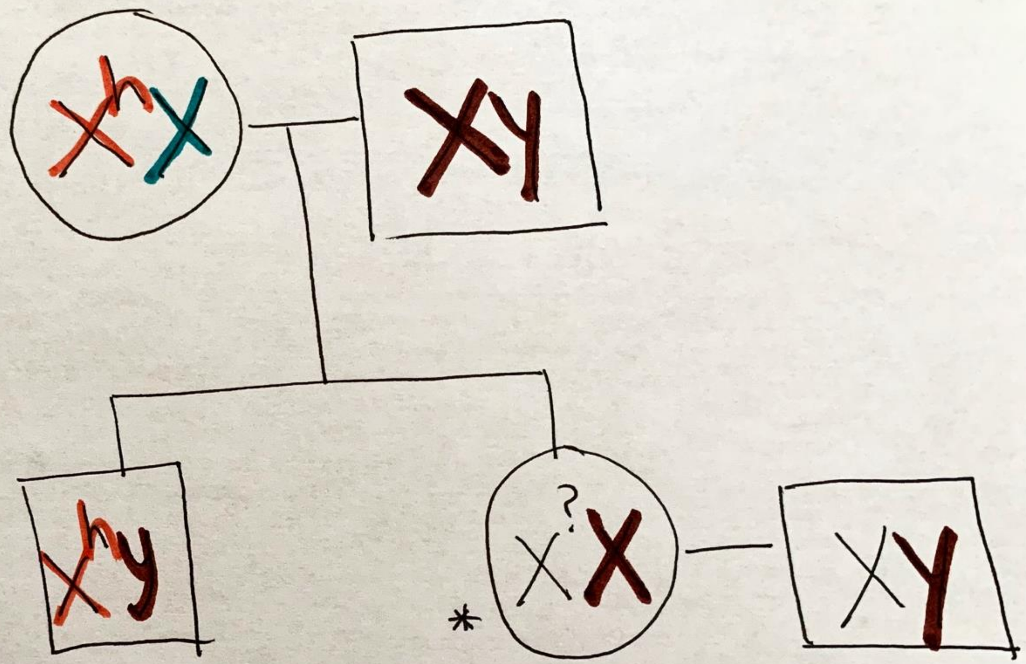
Before offspring:

Probability of being carrier:

$P(\theta=1) = \frac{1}{2}$

Probability of <u>not</u> being carrier:

$P(\theta=0) = \frac{1}{2}$



X-linked condition

- Woman is unaffected by Hemophilia, but she has a brother who is affected by Hemophilia. She refuses to get genetically tested.
- Hemophilia allele is located on the X chromosome
- Hemophilia is a recessive trait
- Their father is unaffected, and their mother is phenotypically unaffected (but she must be a carrier)

Since the woman has a brother with the disease, she can be a carrier for the recessive allele, or she may have inherited a typical X (doesn’t carry the recessive allele) from her mother (we know that she inherited the typical X from her father because he is unaffected and therefore must have a typical X)

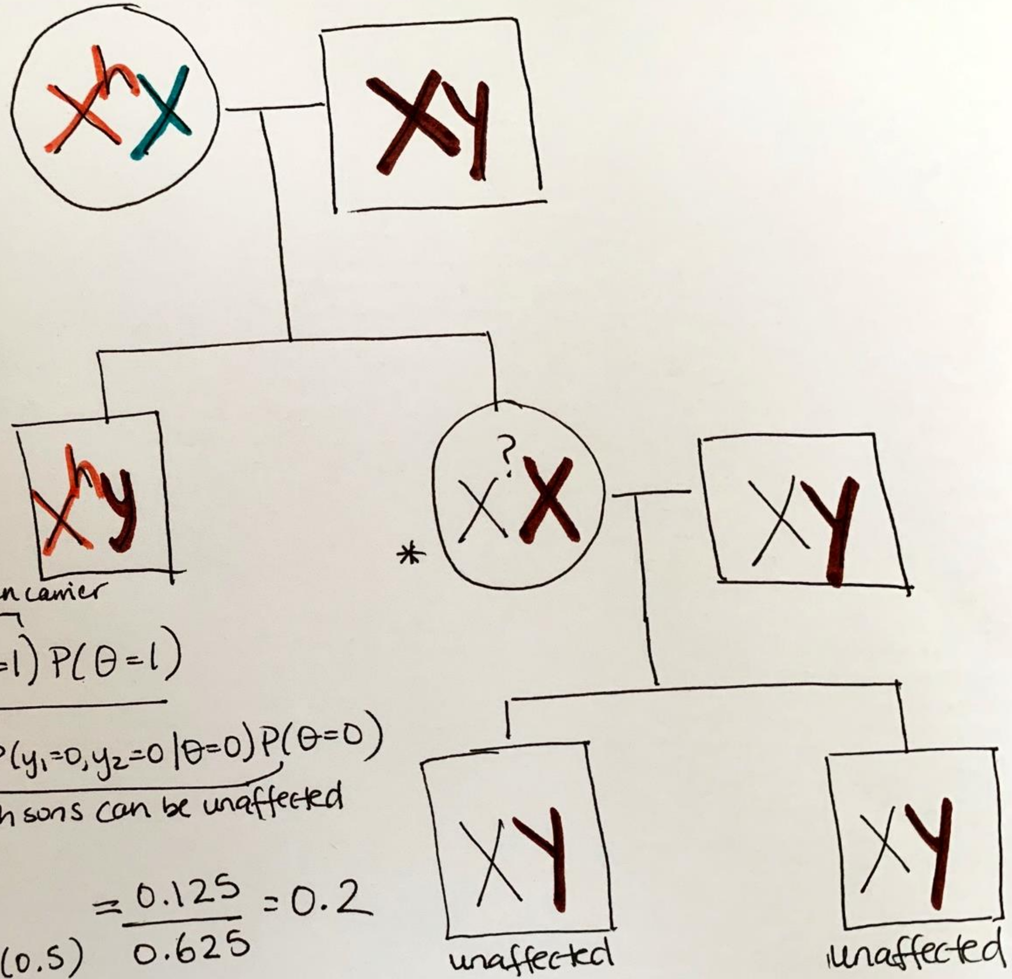
* Before offspring:

Probability of being carrier:

$$P(\theta=1) = \frac{1}{2}$$

Probability of not being carrier:

$$P(\theta=0) = \frac{1}{2}$$



After having two unaffected sons:

(Bayes)

$$P(\theta=1 \mid y_1=0, y_2=0) = \frac{\overbrace{P(y_1=0, y_2=0 \mid \theta=1)}^{\text{Two unaffected when carrier}} P(\theta=1)}{\underbrace{P(y_1=0, y_2=0 \mid \theta=1)P(\theta=1) + P(y_1=0, y_2=0 \mid \theta=0)}_{\text{all ways both sons can be unaffected}} P(\theta=0)}$$

$$= \frac{(0.25)(0.5)}{(0.25)(0.5) + 1\cdot(0.5)} = \frac{0.125}{0.625} = 0.2$$

- WITH EXTRA INFORMATION (2 unaffected sons), the probability that mom is a carrier has gone from $0.5 \rightarrow 0.2$
- Now what happens if mom has one more unaffected son?

Based on the pedigree and the known inheritance mechanism
P(woman being carrier) =P(Θ=1)= 0.5
P(woman not being carrier) =P(Θ=0)= 0.5

P(woman being carrier | two sons are unaffected)
= P(two unaffected sons and carrier)/P(all ways unaffected sons)

$$P[\Theta=1 \mid y_1=0, y_2=0] = \frac{P(y_1=0, y_2=0 \mid \Theta=1) * P(\Theta=1)}{P(y_1=0, y_2=0 \mid \Theta=1) * P(\Theta=1) + P(y_1=0, y_2=0 \mid \Theta=0) * P(\Theta=0)}$$

$$P[\Theta=1 \mid y_1=0, y_2=0] = \frac{0.25 * 0.5}{0.25 * 0.5 + 1 * 0.5} = \frac{0.125}{0.625} = 0.2$$

$$P[\Theta=1 \mid y_1=0, y_2=0] = \frac{0.25 * 0.5}{0.25 * 0.5 + 1 * 0.5} = \frac{0.125}{0.625} = 0.2$$

Based on the pedigree and the known inheritance mechanism
P(woman being carrier) =P(Θ=1)= 0.5
P(woman not being carrier) =P(Θ=0)= 0.5

$$P[\Theta=1 \mid y_1=0, y_2=0] = \frac{0.25*0.5}{0.25*0.5+1*0.5} = \frac{0.125}{0.625} = 0.2$$

We have now <u>updated our prior probability</u> of the woman being a carrier from a starting prior of <u>0.5</u> to <u>0.2</u>!

If she went on to have a **<u>third unaffected son</u>**, this would provide additional evidence and would continue to change the woman's probability of being a carrier:

We have now <u>updated our prior probability</u> of the woman being a carrier from a starting prior of <u>0.5</u> to <u>0.2</u>!

If she went on to have a **<u>third unaffected son</u>**, this would provide additional evidence and would continue to change the woman's probability of being a carrier. Note: we would use our new updated prior probability of 0.2 (instead of the original prior of 0.5):

$$P[\Theta=1 \mid y_1=0, y_2=0, y_3=0] = \frac{P(y_1=0, y_2=0, y_3=0 \mid \Theta=1) * P(\Theta=1)}{P(y_1=0, y_2=0, y_3=0 \mid \Theta=1) * P(\Theta=1) + P(y_1=0, y_2=0, y_3=0 \mid \Theta=0) * P(\Theta=0)}$$

$$P[\Theta=1 \mid y_1=0, y_2=0, y_3=0] = \frac{0.5*0.2}{0.5*0.2+1*0.8} = 0.111$$

Extra resources for understanding Bayes’

- One of the very best simple visual explanations of Bayes’ theorem is found here (it uses LEGO):

https://www.countbayesie.com/blog/2015/2/18/bayes-theorem-with-lego

- Count Bayesie has also put together a guide to Bayesian statistics

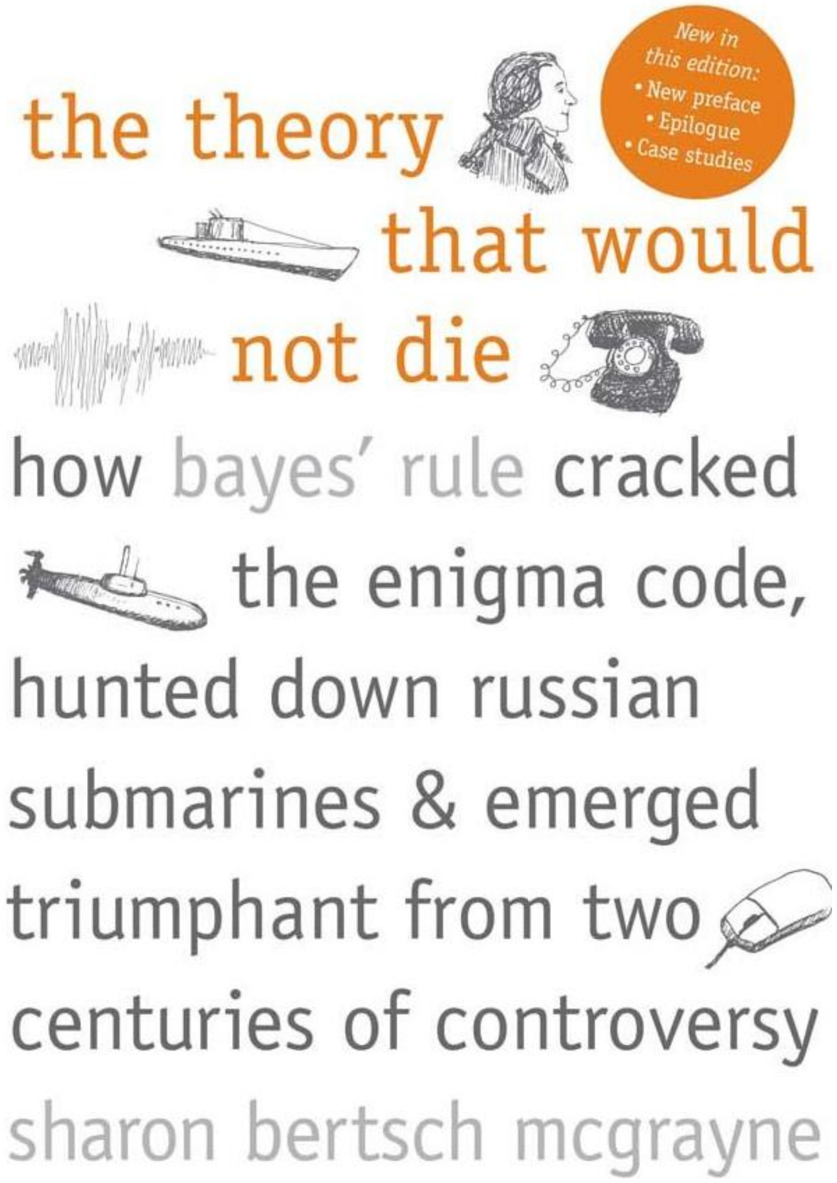
which includes some useful resources if you are struggling with some of the concepts that we have covered so far in lecture:

https://www.countbayesie.com/blog/2016/5/1/a-guide-to-bayesian-statistics

- There is an interesting video about using Bayesian inference to search for sunken treasure:

http://fivethirtyeight.com/features/how-data-nerds-found-a-131-year-old-sunken-treasure/

(it has also been used to find “black boxes” of airplanes that have crashed)



More Extra resources for Bayes’

Another one of the *best* websites to help you develop intuition about Bayes’ and it gives you multiple options for how sophisticated you want your examples to be:

https://arbital.com/p/bayes_rule/?l=1zq

Also see this article from 2014 on the particular importance of understanding statistics/test results as a physician (or scientist who must justify and discuss results):

http://www.bbc.com/news/magazine-28166019

Upload to Canvas as part of your Certification of Completion

Module 2B Questions:

1. *Suppose you work in a psychiatric institution and a patient is referred to you by the GP because of an elevated score on a depression questionnaire.*

--

- *In the practice of this GP, **10%** of patients have depression (the **'prevalence'** of the disorder).*
- *If a patient has depression, the likelihood that they have a positive score on this depression questionnaire is **90%** (the **'sensitivity'** of the test).*
- *A patient who does not have depression has an **80%** chance of a negative score (the **'specificity'** of the test)*

--

WHAT IS YOUR ESTIMATE THAT THE PATIENT HAS DEPRESSION?

2. (TBD) Please also finish up any questions we might not have finished during lecture.