L25: Monday, March 27. 1

Housekeeping:

- A14 due in class today
- Writing assignments $\left\{\begin{array}{l}\text{summary}\\ \text{discussion}\end{array}\right.$ due 11:59 p.m. today on Canvas
- Extra credit (Wage Gap rebuttal) due 11:59 p.m. on Canvas
- A15 due in class Wednesday
- Exam 2 on Friday
- Wednesday's class will be review

Last time: Cond'l probability
Independent events
Multiplication rules
Questions?

This time: Bayes' Theorem

Suppose that the dep't of public health tells us that 1 in 1,000 people has a certain disease;

Suppose, further, that researchers and clinicians tell us that a certain test for the disease yields a positive result 90% of the time when the disease is present, and 1% of the time when the disease is not present.

GIVEN THAT SOMEONE TESTS POSITIVE FOR THE DISEASE, WHAT IS THE PROBABILITY THAT THIS PERSON ACTUALLY HAS THE DISEASE?

We currently don't (yet) have a formula to tell us, so let's assume a population of 1,000,000 people (or any other number?), and fill in the table:

	T test pos.	T^c test neg.
disease D		
no disease D^c		

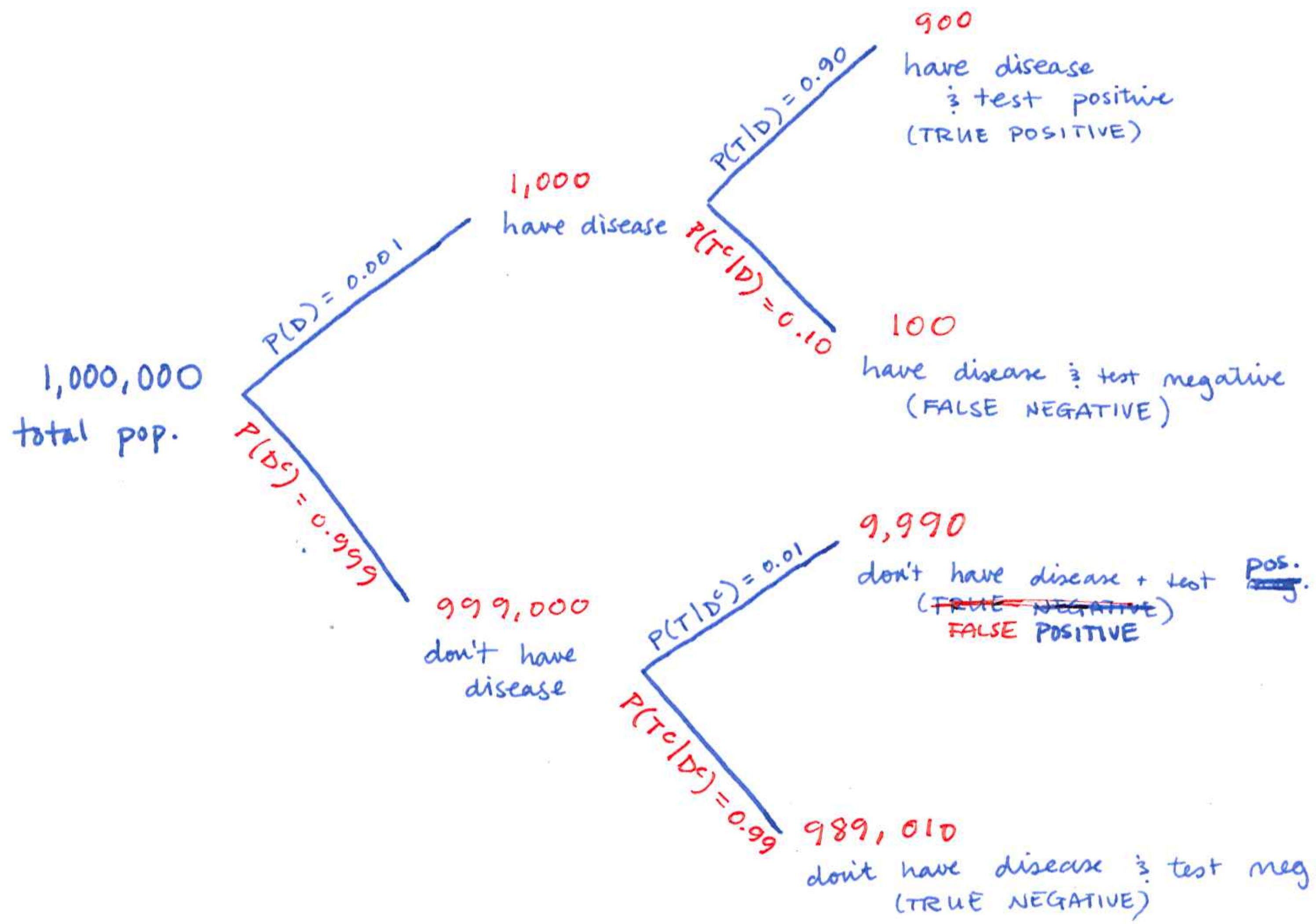
Once we fill in the table, we can compute $P(D|T)$.

~~[illegible]~~ **KNOWN:**

$$P(D) = \frac{1}{1{,}000} = 0.001$$

$$P(T \mid D) = 0.90$$

$$P(T \mid D^c) = 0.01$$



So, the table:

	T	T^c	total
D	900	100	1,000
D^c	9,990	989,010	999,000
total	10,890	989,110	1,000,000

So,

$$P(D \mid T) = \frac{900}{10{,}890}$$

$$\approx 0.083,$$

or 8.3%.

Doctors get this wrong ALL THE TIME.

The rule that systematizes computations like these (so you don't need to fill in a table every time) is called Bayes' Theorem:

For any two events A and B,

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|A^c)\,P(A^c)}.$$

$$\frac{a}{b+c} \neq \frac{a}{b} + \frac{a}{c}$$

Applying to our disease test:

$$P(D|T) = \frac{P(T|D)\,P(D)}{P(T|D)P(D) + P(T|D^c)\,P(D^c)}$$

$$= \frac{0.90(0.001)}{0.90(0.001) + 0.01(0.999)}$$

$$\approx 0.083 \quad \text{or} \quad 8.3\%$$

Known:

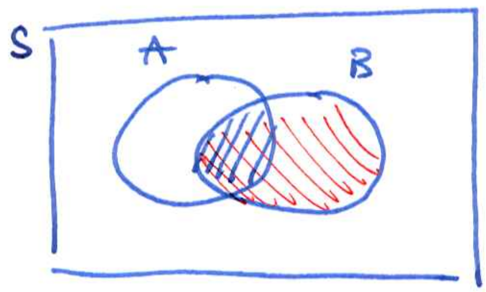
$P(D) = \frac{1}{1{,}000} = 0.001$

$P(D^c) = 1 - 0.001 = 0.999$

$P(T|D) = 0.90$

$P(T|D^c) = 0.01$

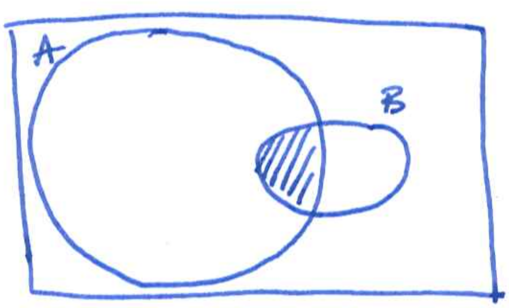
Recall: $P(A|B) = \dfrac{P(A \text{ and } B)}{P(B)} = \dfrac{\text{Area}(A \text{ and } B) / \cancel{\text{Area}(S)}}{\text{Area}(B) / \cancel{\text{Area}(S)}}$

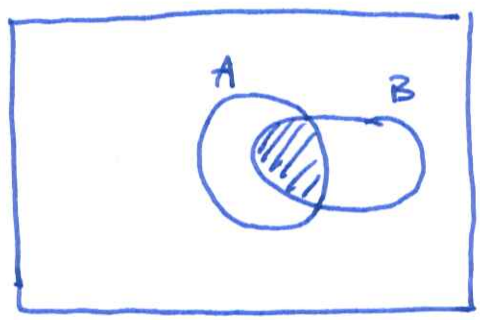


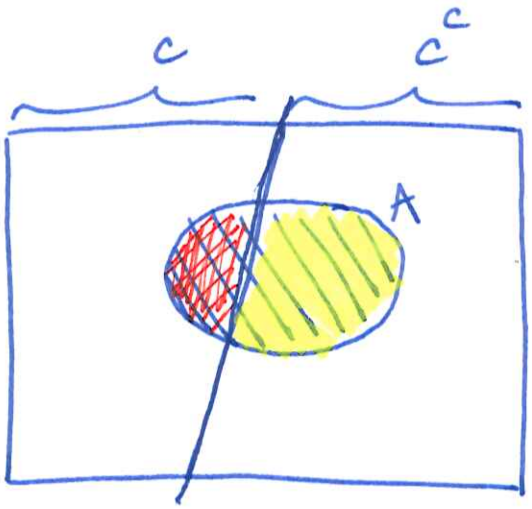
$= \dfrac{\text{Area}(A \text{ and } B)}{\text{Area}(B)}$

So, $P(A|B)$ is the proportion of outcomes from B that are also in A.

Notice: The size of A itself has very little to do with $P(A|B)$.







$$P(A) = \underline{P(A|C)P(C)} + \underline{P(A|C^c)P(C^c)} \quad \Big\} \text{ "LAW OF TOTAL PROBABILITY"}$$

<u>Recall</u>: Multiplication Rules:

$$P(A \text{ and } B) = \boxed{P(B)\,P(A|B)}$$

$$\boxed{P(A)\,P(B|A)} = P(A \text{ and } B)$$

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

$$P(A|B) = \frac{P(A)\,P(B|A)}{\boxed{P(B)}} \quad \Big\} \text{ new rule.}$$

By the law of total prob.:

$$P(B) = \boxed{P(B|A)\,P(A) + P(B|A^c)\,P(A^c)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)\,P(A^c)} \quad \Big\} \text{ Bayes' Theorem.}$$

Know: $P(D) = 0.05$

$P(T|D) = 0.95$

$P(T|D^c) = 0.05$

Seek: $P(D|T) = ?$

$P(D^c) = 1 - P(D) = 1 - 0.05 = 0.95$

$$P(D|T) = \frac{P(T|D)\,P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

$$= \frac{0.95(0.05)}{0.95(0.05) + 0.05(0.95)}$$

$$= \frac{1}{2} = 50\%.$$

if one thing were changed:

$P(D) = \cancel{0.95}$ 0.40, $P(D^c) = 0.60$

$P(T|D) = 0.95$

$P(T|D^c) = 0.05$

$$P(D|T) = \frac{0.95(0.40)}{0.95(\cancel{0.05}\,0.40) + 0.05(0.60)} \approx 0.926$$

$= 92.6\%$