

LECTURE 12: BAYES' THEOREM II

I. Bayes' Theorem

- a. This can all be summarized with *Bayes' Theorem*—a way to describe how one should adjust their beliefs to account for evidence.
- b. To better understand Bayes, let's first understand some terminology.
 - i. $P(A)$ —the probability event A will occur.
 - ii. $P(A/B)$ —the probability event A will occur assuming event B happens; this is the conditional probability.
 - iii. \sim —this symbol means “not.” $P(\sim A)$ means “the probability event A will not occur.”
 - iv. Employing the example from Part II, let's let **U** be “employee is a user” (so $\sim U$ is “employee is not a user”) and **+** be “the test came back positive for drugs” (so $\sim +$ is “the test is not positive, or came back negative for drugs”).
- c. Bayes' Theorem is as stated:

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

- i. Note why the denominators are equivalent. If A happens, how likely is B? If A doesn't happen, how likely is B? Adjusting for the likelihood A happens, and you get the probability that B happens. This last step is similar to how we solved the Simpson Paradox; it's a weighted average.
- d. We want to know “what is the probability that someone who tested positive for drugs is actually a user?” Or, given that the test is positive, how likely is it that they actually use drugs? Or, what is $P(U|+)$?

$$\begin{aligned} P(U|+) &= \frac{P(+|U) P(U)}{P(+)} = \frac{P(+|U) P(U)}{P(+|U) P(U) + P(+|\sim U) P(\sim U)} \\ &= \frac{0.9 * 0.01}{0.9 * 0.01 + 0.1 * 0.99} = \frac{0.009}{0.009 + 0.099} = \frac{0.009}{0.108} \cong 0.083 \end{aligned}$$

- i. Note the result is as before: 8.3%. Only 8.3% of positive results are actually drug users!
 - e. Bayes' Theorem thus tells us two interrelated things:
 - i. When you receive new information, it tells you how much to adjust your estimation of the truth.
 - ii. It reminds you that because no test is 100% accurate, its results should not be weighed too heavily, especially if it is testing for something rare (since the false positives will overwhelm the true positives).
- II. Bayes in Excel
 - a. We can turn Excel into a calculator for this Theorem, allowing us to change parameters and see how the results transform.
 - b. Begin by turning a column into labels: Population Chance; Sensitivity, and Specificity. The adjacent cells will be the values.
 - i. How should we represent those values?
 - c. Below label three cells: $P(Y|+)$, numerator, denominator.
 - d. In class, we'll build this in detail but I want us to work on it together!

