

LECTURE 23: BAYES' THEOREM II

- I. Conditional probability
 - 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' Theorem, 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.” Thus $P(A) + P(\sim A) = 1$.
 - iv. Employing the example from before, 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. Let's begin with an important equivalency.

$$P(B) = P(B|A)P(A) + P(B|\sim A)P(\sim A)$$

- i. Why does this work? If you're trying to figure out how likely is it for B to happen, recognize that there's a chance of B happening if A happens and there's a chance of B happening if A doesn't happen. Adjusting for the likelihood A happens, and you get the probability that B happens.
 - ii. For example, what is the chance of you having a good job in five years? You can have a good job if you get a bachelor's degree and you can have one if you don't get a bachelor's degree but the chance of the first conditional is much higher than the chance of the second conditional.
- II. Bayes' Theorem
 - a. Bayes' Theorem is as stated:

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

- b. 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 $P(U|+)$?

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

$$= \frac{0.8 * 0.01}{0.8 * 0.01 + 0.1 * 0.99} = \frac{0.008}{0.008 + 0.099} = \frac{0.008}{0.107} \cong 0.0748$$

- i. Note the result is as before: 7.48%. Only 7.48% of positive results are actually drug users!
- c. 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).

III. 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.
- i. Excel does not have a built-in function.
- b. Begin by turning a column into labels: Chance; Sensitivity, and Specificity. The adjacent cells will be the values (B1 is 0.01; B2 is 0.8, B3 is 0.9).

	A	B
1	Chance	0.01
2	Sensitivity	0.8
3	Specificity	0.9

- c. Below label three cells: $P(U|+)$, numerator, denominator and then fill in as so:

	A	B
1	Chance	0.01
2	Sensitivity	0.8
3	Specificity	0.9
4		
5	numerator	=B2*B1
6	denominator	=B5+(1-B3)*(1-B1)
7	P(U +)	=B5/B6

- d. Keep in mind that these numbers are what they are because the underlying values in Bayes' Theorem. For example, consider the second part of the denominator is: $P(+|\sim U)*P(\sim U)$.
- i. $\sim U$ means we're dealing with a negative state, thus the relevant accuracy measure is specificity.
 - ii. But we're asking for the chance of an incorrect result, or the chance of a positive test result given that the person isn't a drug user. Thus, we subtract specificity from 1.
 - iii. We also subtract the chance of being a drug user from 1 because we want the chance of someone not being a drug user.
- e. Now watch what happens as we turn up the specificity from 90% to 99% to 99.9%:

	A	B		A	B		A	B
1	Chance	0.01	1	Chance	0.01	1	Chance	0.01
2	Sensitivity	0.8	2	Sensitivity	0.8	2	Sensitivity	0.8
3	Specificity	0.9	3	Specificity	0.99	3	Specificity	0.999
4			4			4		
5	numerator	0.008	5	numerator	0.008	5	numerator	0.008
6	denominator	0.107	6	denominator	0.0179	6	denominator	0.00899
7	P(U +)	7.48%	7	P(U +)	44.69%	7	P(U +)	88.99%

- f. A positive test result becomes much more informative! Again, the earlier problem was that false positives overwhelmed true positives but by increasing specificity, we're able to get far fewer false positives.