

3.3: Bayes's Formula

Alex L.

October 16, 2024

Proposition: (Bayes's Formula) $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$

Proof: Event E is a disjoint union of $(E \cap F) \cup (E \cap F^C)$, in other words, event E is equal to E and F or E and not F . Since these are disjoint, $P(E) = P(E \cap F) + P(E \cap F^C)$. We know that $P(E|F) = \frac{P(E \cap F)}{P(F)}$, and multiplying by $P(F)$, then substituting, we get $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$.

This result makes intuitive sense: weighting the average of the individual conditions with the chance of conditions happening should give us our original probability.

This finding is useful when $P(E|F)$ and $P(F)$ are easy to calculate, but $P(E)$ is not easy to calculate.

Example: (3a) An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. The company's statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability .4, whereas this probability decreases to .2 for a person who is not accident prone. If we assume that 30 percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?

Solution: Let's let event E be the event someone gets into an accident, and event F be the chance someone is accident-prone. Then, $P(E|F) = .4$, and $P(E|F^C) = .2$, $P(F) = .3$, and $P(F^C) = .7$. We plug these numbers into our equation to get $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C) = .4 * .3 + .7 * .2 = .12 + .14 = .26$

Example: (3a Part 2) Suppose that a new policyholder has an accident within a year of purchasing a policy. What is the probability that he or she is accident prone?

Solution: We are now given event E and want to find event F . $P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{.12}{.26} = \frac{6}{13}$

Example: (3b) Consider the following game played with an ordinary deck of 52 playing cards: The cards are shuffled and then turned over one at a time. At any time, the player can guess that the next card to be turned over will be the ace of spades; if it is, then the player wins. In addition, the player is said to win if the ace of spades has not yet appeared when only one card remains and no guess has yet been made. What is a good strategy? What is a bad strategy?

Solution: Every strategy has a $\frac{1}{n}$ chance of winning in an n card deck. Let's use induction to figure it out. By the rules of the game, you automatically win with a 1 card deck, so the win probability is 1. Now, let's assume that for an $n - 1$ card deck, the win chance for any strategy is $\frac{1}{n-1}$, and calculate the win probability for an n card deck. If your strategy is to guess that the first card is an ace of spades, you have a $\frac{1}{n}$ chance of winning. If your strategy is not guessing on the first card, then after the first card is removed, you have a deck with $n - 1$ cards, and we know that by assumption, the win chance is $\frac{1}{n-1}$. But, for the second strategy, the card flipped can't be the ace of spades, or else you instantly lose, so the chance the second strategy wins is $\frac{n-1}{n} \frac{1}{n-1}$ which is equal to $\frac{1}{n}$. As such, no matter what strategy you pursue, your win chance is always $\frac{1}{n}$.

Example: (3c) In answering a question on a multiple-choice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer and $1 - p$ be the probability that the student guesses. Assume

that a student who guesses at the answer will be correct with probability $\frac{1}{m}$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that he or she answered it correctly?

Solution: Let E be the event that the student gets the question right, and F be the event the student knows the answer. We want to know $P(F|E) = \frac{P(E \cap F)}{P(E)}$ and since $P(E \cap F) = P(E|F)P(F)$, and $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$, by substituting, we get $P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^C)P(F^C)} = \frac{p}{p + \frac{1}{m}(1-p)}$

Example: A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a “false positive” result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability .01, the test result will imply that he or she has the disease.) If .5 percent of the population actually has the disease, what is the probability that a person has the disease given that the test result is positive?

Solution: Let D be the event someone has the disease, and let T be the event that they test positive. We know that $P(T|D) = .95$ and $P(T|D^C) = .01$, and $P(D) = .005$. We want to know $P(D|T) = \frac{P(D \cap T)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^C)P(D^C)} = \frac{.00475}{.00475 + .00995} = .323129$, or about a third of those who test positive actually have the disease. What a shitty fucking test.

Example: Consider a medical practitioner pondering the following dilemma: “If I’m at least 80 percent certain that my patient has this disease, then I always recommend surgery, whereas if I’m not quite as certain, then I recommend additional tests that are expensive and sometimes painful. Now, initially I was only 60 percent certain that Jones had the disease, so I ordered the series A test, which always gives a positive result when the patient has the disease and almost never does when he is healthy. The test result was positive, and I was all set to recommend surgery when Jones informed me, for the first time, that he was diabetic. This information complicates matters because, although it doesn’t change my original 60 percent estimate of his chances of having the disease in question, it does affect the interpretation of the results of the A test. This is so because the A test, while never yielding a positive result when the patient is healthy, does unfortunately yield a positive result 30 percent of the time in the case of diabetic patients who are not suffering from the disease. Now what do I do? More tests or immediate surgery?”

Solution: Previously, we were almost certain Jones should go into surgery. Now, we need to compute the updated probability of Jones having the disease. Let D be the chance that Jones has the disease, and T be the chance he tests positive. $P(T|D^C) = .3$ and $P(T|D) = 1$. Also, $P(D)$ in a vacuum is .6. We want to know $P(D|T) = \frac{P(D \cap T)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^C)P(D^C)} = \frac{.6}{.6 + .12} = .833$, so the doctor should give Jones the surgery, just like House MD.