

6.4: Bayes Theorem

In this section we concentrate on the more complex conditional probability problems we began looking at in the last section.

✓ Example 6.4.1

Suppose a certain disease has an incidence rate of 0.1% (that is, it afflicts 0.1% of the population). A test has been devised to detect this disease. The test does not produce false negatives (that is, anyone who has the disease will test positive for it), but the false positive rate is 5% (that is, about 5% of people who take the test will test positive, even though they do not have the disease). Suppose a randomly selected person takes the test and tests positive. What is the probability that this person actually has the disease?

Solution

There are two ways to approach the solution to this problem. One involves an important result in probability theory called Bayes' theorem. We will discuss this theorem a bit later, but for now we will use an alternative and, we hope, much more intuitive approach.

Let's break down the information in the problem piece by piece.

Suppose a certain disease has an incidence rate of 0.1% (that is, it afflicts 0.1% of the population). The percentage 0.1% can be converted to a decimal number by moving the decimal place two places to the left, to get 0.001. In turn, 0.001 can be rewritten as a fraction: $\frac{1}{1000}$. This tells us that about 1 in every 1000 people has the disease. (If we wanted we could write $P(\text{disease})=0.001$.)

A test has been devised to detect this disease. The test does not produce false negatives (that is, anyone who has the disease will test positive for it). This part is fairly straightforward: everyone who has the disease will test positive, or alternatively everyone who tests negative does not have the disease. (We could also say $P(\text{positive} \mid \text{disease})=1$.)

The false positive rate is 5% (that is, about 5% of people who take the test will test positive, even though they do not have the disease). This is even more straightforward. Another way of looking at it is that of every 100 people who are tested and do not have the disease, 5 will test positive even though they do not have the disease. (We could also say that $P(\text{positive} \mid \text{no disease})=0.05$.)

Suppose a randomly selected person takes the test and tests positive. What is the probability that this person actually has the disease? Here we want to compute $P(\text{disease} \mid \text{positive})$. We already know that $P(\text{positive} \mid \text{disease})=1$, but remember that conditional probabilities are not equal if the conditions are switched.

Rather than thinking in terms of all these probabilities we have developed, let's create a hypothetical situation and apply the facts as set out above. First, suppose we randomly select 1000 people and administer the test. How many do we expect to have the disease? Since about $\frac{1}{1000}$ of all people are afflicted with the disease, $\frac{1}{1000}$ of 1000 people is 1. (Now you know why we chose 1000.) Only 1 of 1000 test subjects actually has the disease; the other 999 do not.

We also know that 5% of all people who do not have the disease will test positive. There are 999 disease-free people, so we would expect $(0.05)(999) = 49.95$ (so, about 50) people to test positive who do not have the disease.

Now back to the original question, computing $P(\text{disease} \mid \text{positive})$. There are 51 people who test positive in our example (the one unfortunate person who actually has the disease, plus the 50 people who tested positive but don't). Only one of these people has the disease, so

$$P(\text{disease} \mid \text{positive}) \approx \frac{1}{51} \approx 0.0196$$

or less than 2%. Does this surprise you? This means that of all people who test positive, over 98% *do not have the disease*.

The answer we got was slightly approximate, since we rounded 49.95 to 50. We could redo the problem with 100,000 test subjects, 100 of whom would have the disease and $(0.05)(99,900) = 4995$ test positive but do not have the disease, so the exact probability of having the disease if you test positive is

$$P(\text{disease} \mid \text{positive}) \approx \frac{100}{5095} \approx 0.0196$$

which is pretty much the same answer.

But back to the surprising result. *Of all people who test positive, over 98% do not have the disease.* If your guess for the probability a person who tests positive has the disease was wildly different from the right answer (2%), don't feel bad. The exact same problem was posed to doctors and medical students at the Harvard Medical School 25 years ago and the results revealed in a 1978 *New England Journal of Medicine* article. Only about 18% of the participants got the right answer. Most of the rest thought the answer was closer to 95% (perhaps they were misled by the false positive rate of 5%).

So at least you should feel a little better that a bunch of doctors didn't get the right answer either (assuming you thought the answer was much higher). But the significance of this finding and similar results from other studies in the intervening years lies not in making math students feel better but in the possibly catastrophic consequences it might have for patient care. If a doctor thinks the chances that a positive test result nearly guarantees that a patient has a disease, they might begin an unnecessary and possibly harmful treatment regimen on a healthy patient. Or worse, as in the early days of the AIDS crisis when being HIV-positive was often equated with a death sentence, the patient might take a drastic action and commit suicide.

As we have seen in this hypothetical example, the most responsible course of action for treating a patient who tests positive would be to counsel the patient that they most likely do *not* have the disease and to order further, more reliable, tests to verify the diagnosis.

One of the reasons that the doctors and medical students in the study did so poorly is that such problems, when presented in the types of statistics courses that medical students often take, are solved by use of Bayes' theorem, which is stated as follows:

Bayes' Theorem

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

In our earlier example, this translates to

$$P(\text{disease} | \text{positive}) = \frac{P(\text{disease}) \cdot P(\text{positive} | \text{disease})}{P(\text{disease}) \cdot P(\text{positive} | \text{disease}) + P(\text{no disease}) \cdot P(\text{positive} | \text{no disease})}$$

Plugging in the numbers gives

$$P(\text{disease} | \text{positive}) = \frac{(0.001)(1)}{(0.001)(1) + (0.999)(0.05)} \approx 0.0196$$

which is exactly the same answer as our original solution.

The problem is that you (or the typical medical student, or even the typical math professor) are much more likely to be able to remember the original solution than to remember Bayes' theorem. Psychologists, such as Gerd Gigerenzer, author of *Calculated Risks: How to Know When Numbers Deceive You*, have advocated that the method involved in the original solution (which Gigerenzer calls the method of "natural frequencies") be employed in place of Bayes' Theorem. Gigerenzer performed a study and found that those educated in the natural frequency method were able to recall it far longer than those who were taught Bayes' theorem. When one considers the possible life-and-death consequences associated with such calculations it seems wise to heed his advice.

Example 6.4.2

A certain disease has an incidence rate of 2%. If the false negative rate is 10% and the false positive rate is 1%, compute the probability that a person who tests positive actually has the disease.

Solution

Imagine 10,000 people who are tested. Of these 10,000, 200 will have the disease; 10% of them, or 20, will test negative and the remaining 180 will test positive. Of the 9800 who do not have the disease, 98 will test positive. So of the 278 total people who test positive, 180 will have the disease. Thus

$$P(\text{disease} | \text{positive}) = \frac{180}{278} \approx 0.647$$

so about 65% of the people who test positive will have the disease.

Using Bayes theorem directly would give the same result:

$$P(\text{disease} \mid \text{positive}) = \frac{(0.02)(0.90)}{(0.02)(0.90) + (0.98)(0.01)} = \frac{0.018}{0.0278} \approx 0.647$$

Try It 6.4.1

A certain disease has an incidence rate of 0.5%. If there are no false negatives and if the false positive rate is 3%, compute the probability that a person who tests positive actually has the disease.

Answer

Out of 100,000 people, 500 would have the disease. Of those, all 500 would test positive. Of the 99,500 without the disease, 2,985 would falsely test positive and the other 96,515 would test negative.

$$P(\text{disease} \mid \text{positive}) = \frac{500}{500 + 2985} = \frac{500}{3485} \approx 14.3\%$$

This page titled [6.4: Bayes Theorem](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [David Lippman](#) ([The OpenTextBookStore](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [12.4: Bayes Theorem](#) by [David Lippman](#) is licensed [CC BY-SA 3.0](#). Original source: <http://www.opentextbookstore.com/mathinsociety>.