

MONT 107N – Understanding Randomness
Lecture Notes on Bayes' Rule
January 27, 2010

What Bayes' Rule says, and why it is true

Bayes' Rule is a probability statement about the relation between the two conditional probabilities $P(A|B)$ and $P(B|A)$. As an equation (the clearest form, I think!), it says: For any two events (with $P(A), P(B) \neq 0$),

$$(1) \quad P(B|A) = P(A|B) \cdot \frac{P(B)}{P(A)}.$$

The reason it works this way is because of the definition of the conditional probability:

$$(2) \quad P(A|B) = \frac{P(A \text{ and } B)}{P(B)}.$$

Applying this to $P(B|A)$, we can manipulate and rearrange algebraically as follows:

$$\begin{aligned} P(B|A) &= \frac{P(B \text{ and } A)}{P(A)} \\ &= \frac{P(A \text{ and } B)}{P(B)} \cdot \frac{P(B)}{P(A)} \\ &= P(A|B) \cdot \frac{P(B)}{P(A)}, \end{aligned}$$

which shows what we wanted to establish from (1).

An context where "Bayesian" reasoning is important

One important example of the sometimes unexpected consequences of Bayes' Rule is the interpretation of medical diagnostic tests (for example the kind of routine screening tests that are now commonly given as part of annual physical exams).

Let A = the event that a test gives a "positive" result. (That is, the test says that you actually have the condition the test is designed to detect – usually some illness or disease). Let B = be the event that you actually have that illness or disease.

In designing medical diagnostic tests, doctors are often most concerned with understanding the following two probabilities:

- $P(A|B)$ = the *sensitivity* of the test. This measures how likely the test is to give a positive result, *given that the person to whom the test is administered actually has that condition.*
- $P(\text{not } A|\text{not } B)$ = the *specificity* of the test. This measures how likely the test is to give a negative result *when the person to whom the test is administered does not have that condition.*

Values as close to 1 as possible in both cases are usually the goal.

So far we have been looking at the diagnostic test essentially from the point of view of a doctor or medical researcher analyzing how well the test operates. But of course, *most of us* are more likely to be in another situation – that of a patient (or a loved one of a patient) when the patient gets one of these tests and then must deal with the consequences of a “positive” result. Now, how sensitive or specific the test is are much less important than trying to assess *a different probability*: Given that the test gave a positive result, what is the probability that the condition or disease is actually present? In symbols, this asks for the “reverse” probability: $P(B|A)$. Note that this is exactly the situation dealt with by Bayes’ Rule from (1)(!) Knowing $P(A|B)$, we want to determine $P(B|A)$.

An Example

Suppose the test was relatively sensitive *and* relatively specific: $P(A|B) = .9$ and $P(\text{not } A|\text{not } B) = .9$. But now, let us also assume that the condition or disease is *relatively rare*. For instance, suppose that only about 1 in 1000 people actually have it, so that if we look at the results of giving the test to the whole population of patients, $P(B) = .001$. The question is: What is $P(B|A)$ in this case?

(Note: It is important to realize that the way we are analyzing this question only makes sense if we are considering the whole population of patients. The statement $P(B) = .001$ means that if we *randomly selected patients* a large number of times, then the selected patient would have the disease about 1 out of 1000 times. We are not claiming anything about any particular patient.)

If we look at (1) again, we can see that we know all of the information on the right side at this point, *except* $P(A)$, the probability of getting a positive test result (whether or not the patient has the condition or disease). We can figure this out, though, as follows. The positive test results are entirely made up of either “true positives” or “false positives” (“false positives” are cases where the result of the test is wrong – the patient does not have the condition, but the test says that they do). So, in symbols

$$P(A) = P(A \text{ and } B) + P(A \text{ and not } B).$$

By (2), this can be rewritten as

$$(3) \quad P(A) = P(A|B)P(B) + P(A|\text{not } B)P(\text{not } B).$$

Here, we know from the information we have that $P(A|B) = .9$, $P(B) = .001$, and $P(\text{not } B) = 1 - .001 = .999$. The only unknown is $P(A|\text{not } B)$. However, if we assume the condition $\text{not } B$, then either A happens or not. This shows that

$$P(A|\text{not } B) = 1 - P(\text{not } A|\text{not } B) = 1 - .9 = .1$$

(from the given specificity value). Hence, from (3), we now have

$$(4) \quad P(A) = (.9)(.001) + (.1)(.999) = .1008.$$

Then from (1)

$$P(B|A) = P(A|B) \cdot \frac{P(B)}{P(A)} = (.9) \cdot \frac{.001}{.1008} \doteq .0089.$$

In other words, there is a bit less than a 1% chance that you have the condition, given that the test results were positive.

Discussion of the example

- The value of $P(B|A)$ should be shockingly, even *distressingly* small (from the point of view of the doctors who designed the test, anyway!)
- On the other hand, it is also not hard to see that with the given numbers, the probability of not having the disease, given that the test results are *negative*, or $P(\text{not}B|\text{not}A)$, is *very close to 1*. So if your test result is negative, you can breathe a lot easier(!)
- The underlying *reason* for the way the example comes out is that when the condition is rare, a *large majority of the positive test results are actually “false positives”* – the “false positive” term $(.1)(.999)$ in (4) is much larger than the “true positive” term $(.9)(.001)$.
- A question you might ask is: What happens if the disease or condition is *more common*? For instance, say $P(B) = .05$ (about 1 in 20 people have the condition), instead of $P(B) = .001$. But still assume the sensitivity and specificity are .9. If you do the computations in that case, then you will see:

$$P(A) = (.9)(.05) + (.1)(.95) = .14$$

and then from (1),

$$P(B|A) = (.9) \cdot \frac{(.05)}{(.14)} \doteq .32$$

(about a 32% chance). This is not as small as in the other case, but still probably a lot smaller than we would expect intuitively. In fact, if you play around with some other cases, you can see that with the sensitivity and specificity both .9, $P(B|A) < .9$ for all values of $P(B) < .1$.

- Another question you might ask is: Well couldn't the problem be that the test is not sensitive or specific enough? Suppose those probabilities were .99, but $P(B) = .001$? It is not difficult to see that in those cases, $P(B|A)$ is still significantly smaller than we might expect intuitively. And if the condition were less common – $P(B) < .001$, we could get similar results as in the main example no matter how close the sensitivity and specificity were to 1(!)