

Base Rates

Prerequisites

[Basic Concepts](#)

Suppose that at your regular physical exam you test positive for Disease X. Although Disease X has only mild symptoms, you are concerned and ask your doctor about the accuracy of the test. It turns out that the test is 95% accurate. It would appear that the probability that you have Disease X is therefore 0.95. However, the situation is not that simple.

For one thing, more information about the accuracy of the test is needed because there two kinds of errors the test can make: [misses](#) and [false positives](#). If you actually had Disease X and the test failed to detect it, that would be a miss. If you did not have Disease X and the test indicated you did, that would be a false positive. The miss and false positive rates are not necessarily the same. For example, suppose that the test accurately indicates the disease in 99% of the people who have it and accurately indicates no disease in 91% of the people who do not have it. In other words, the test has a miss rate of 0.01 and a false positive rate of 0.09. This might lead you to revise your judgment and conclude that your chance of having the disease is 0.91 rather than 0.95. This would be true if half the people in your situation (people who show up for a regular physical exam) had disease X. The analysis becomes complicated if more or less than half the people in your situation have Disease X. The proportion of the people having the disease is called the [base rate](#).

Assume that Disease X is a rare disease, and only 2% of people in your situation have it. How does that affect the probability that you have it? Or, more generally, what is the probability that someone who tests positive actually has the disease. Lets consider what would happen if one million people were tested. Out of these one million people, 2% or 20,000 people would have the disease. Of these 20,000 with the disease, the test would accurately detect it in 99% of them. This means that 19,800 cases would be accurately identified. Now lets consider the 98% of the one million people (980,000) who do not have the disease. Since the false positive rate is 0.09, 9% of these 980,000 people will test positive for the disease. This is a total of 88,200 people incorrectly diagnosed.

To sum up, 19,800 people who tested positive would actually have the

disease and 88,200 people who tested positive would not have the disease. This means that of all those who tested positive, only

$$19,800 / (19,800 + 88,200) = 0.1833$$

of them would actually have the disease. So the probability that you have the disease is not 0.95, or 0.91, but only 0.1833.

These results are summarized in Table 1. The numbers of people diagnosed with the disease are shown in red. Of the one million people tested, the test was correct for 891,000 of those without the disease and for 19,800 with the disease; the test was correct 91% of the time. However, if you look only at the people testing positive (shown in red), only 19,800 (0.1833) of the 911,600 testing positive actually have the disease.

Table 1. Diagnosing Disease X.

True Condition			
No Disease 980,000		Disease 20,000	
Test Result		Test Result	
Positive 88,200	Negative 891,800	Positive 19,800	Negative 200

Bayes' Theorem

This same result can be obtained using Bayes' theorem. Bayes' theorem considers both the [prior probability](#) of an event and the diagnostic value of a test to determine the [posterior probability](#) of the event. For the current example, the event is that you have Disease X. Let's call this Event D. Since only 2% of people in your situation have Disease X, the prior probability of Event D is 0.02. Or, more formally, $P(D) = 0.02$. If D' represents the probability that Event D is false, then $P(D') = 1 - P(D) = 0.98$.

To define the diagnostic value of the test, we need to define another event: that you test positive for Disease X. Let's call this event T. The diagnostic value of the test depends on the probability you will test positive given that you actually have the disease, written as $P(T|D)$, and the probability you test positive given that you do not have the disease, written as $P(T|D')$. Bayes' theorem shown below allows you to calculate $P(D|T)$, the probability that

you have the disease given that you test positive for it.

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D')P(D')}$$

The various terms are:

$$P(T|D) = 0.99$$

$$P(T|D') = 0.09$$

$$P(D) = 0.02$$

$$P(D') = 0.98$$

Therefore,

$$P(D|T) = \frac{(0.99)(0.02)}{(0.99)(0.02) + (0.09)(0.98)} = 0.1833$$

which is the same value computed previously.