David Youngberg
BSAD 210—Montgomery College

LECTURE 22: BAYES' THEOREM I

I. Testing

a. Suppose you're interested which of your employees are on drugs. This is very rare, of course, but if they work with heavy machinery or sensitive information it's particularly important that they're clean.

b. Suppose national surveys reveal that **one out of every 100 people** who work in your industry use drugs.

c. The problem is that no drug test is 100% accurate. Still, most can get close. Suppose you use a test which has a **sensitivity of 80% and a specificity of 90%**:

   i. *Sensitivity*—describes the ability to detect a positive state. For every 10 drug users, 8 will get a (correct) positive test result. Subtracting sensitivity from one tells you the chance of getting a false negative.

   ii. *Specificity*— describes the ability to detect a negative state. For every 10 non-users, 9 will get a (correct) negative test result. Subtracting specificity from one tells you the chance of getting a false positive.

   iii. In practice, there is often a trade-off between the two. A test that is made to be very sensitive often means its specificity decreases. For example, a sensitive metal detector will not only pick up more threats (true positives) but also detect more false positives such as cell phones and belt buckles.

   iv. Sensitivity and specificity are based on c*onditional probability*: the probability one event will occur given another event has occurred.

      1. The probability I'm carrying an umbrella on any random day is low. But the probability I'm carrying an umbrella given it rained today is much higher than the nonconditional. The probability I'm carrying an umbrella given that it snowed today is much lower than the nonconditional.

   v. **If** someone is on drugs then the chance of detecting it is the sensitivity.

   vi. **If** they aren't on drugs, then the chance of discovering that they aren't is specificity.

II. Tree Diagram

a. Let's return to the original problem: which employees are on drugs and which are not?

b. If we leave it only to the test, we must recognize we could be misled:

   i. Any positive result can either be true (the test was right and the subject is a drug user) or false (the test was wrong and the subject is not a drug user).

   ii. Any negative result can either be true (the test was right and the subject is not a drug user or false (the test was wrong and the subject is a drug user.

   iii. We call these results true positive, false positive, true negative, and false negative.

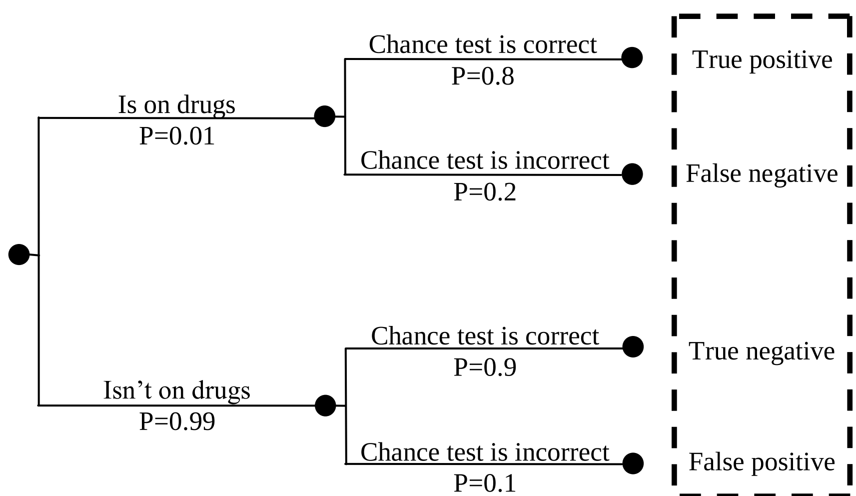
c. In other words, just because the test comes back positive doesn't mean the employee is on drugs. But how likely are they on drugs?

   i. Before the test, they were 1% likely. Without any other information, we can only assume they have the same chance as everyone else.

   ii. With a positive result, it should go up. But how much? Not to 100% (because the result could be wrong).

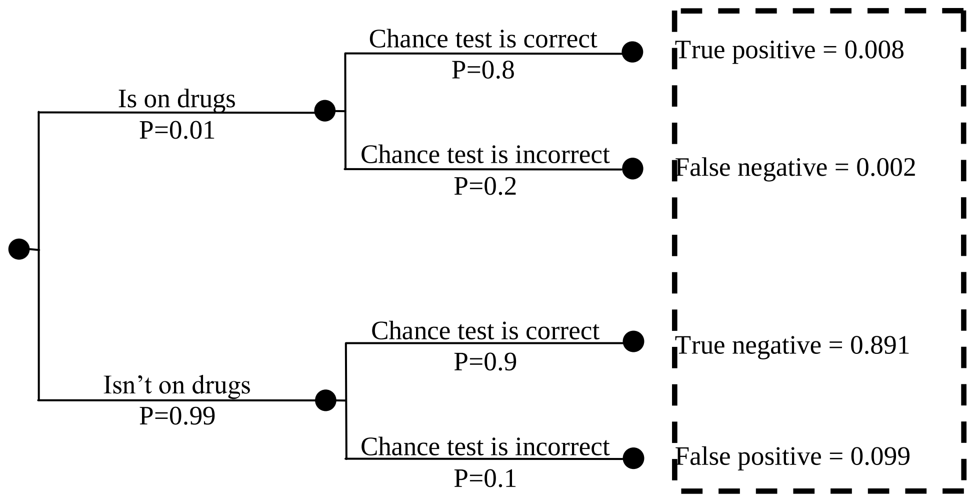
   iii. With a negative result, it should go down. But how much? No to 0% (because the result could be wrong).

d. This is the question Bayes' Theorem answers. Before I show you the equation, let's make a tree diagram of the various possible outcomes:



   i. Note that a false negative is "paired" with a true positive. The result of the test would be negative and that result would be wrong.

e. Now let's calculate the chance of each scenario. We do this by multiplying each conditional probability by the appropriate condition.
    i. Recall for a true positive to occur, the subject must be on drugs AND the test must come back positive.



f. Note there are two ways to get a positive result (true and false) and two ways to get a negative result (true and false).
    i. Chance of a positive result: 0.008 + 0.099 = 0.107 = 10.7%
    ii. Chance of a negative result: 0.002 + 0.891 = 0.893 = 89.3%

g. Now the really interesting part: if you get a positive result, what are the chances the person is actually a drug user?
    i. To answer, determine what portion of positive results are true positives:

Chance of a true positive

$$\frac{0.008}{0.107} \cong 0.0748, or\ 7.48\%$$

Chance of a positive result

    ii. Before testing, you had a 1% chance of being a user. After testing, a positive result increased your likelihood of being a user but because the condition (drug user) is so rare, the very many false positives overwhelm the true positives.
    iii. In other words, it's hard to detect rare things.