Conditional Probability: Diagnostic Test Performance

BIOS 6611

CU Anschutz

Week 5

Background and Example

Conditional Probability and the Confusion Matrix

One common use of conditional probability is to summarize the performance of screening and diagnostic tests.

Based on a gold standard or true disease status, a study to evaluate the performance of a new test can be set up as a **confusion matrix** (also known as a **2x2 table**):

Test	**Gold Standard/ Disease Status** **Positive** (D)	**Negative** ($\bar{D}$)	**Total**
Positive (T)	a (true positives)	b (false positives)	a+b
Negative ($\bar{T}$)	c (false negatives)	d (true negatives)	c+d
Total	a+c	b+d	N (a+b+c+d)

How to Summarize Performance of Diagnostic Tests

Diagnostic tests can be thought of as a prediction problem. Based on our test, we are trying to predict the *true* outcome. Oftentimes we want to develop a test because it might be cheaper, quicker, less invasive, or have some other benefit relative to the gold standard.

There are multiple types of summaries we will discuss in our slide sets on diagnostic testing:

- properties of the test (accuracy, sensitivity, specificity)
- predictive utility of the test (PPV, NPV, LR$+$, LR-, posterior odds)
- performance across different possible thresholds (ROC curves)

Some of these summaries depend on the **prevalence** of the disease, or the probability of having the disease. This is also called the *prior probability* of having the disease.

Motivating Example

Coronary heart disease (CHD) is a disease of the blood vessels that supply the heart and is the most common type of heart disease. For our example, we know that the prevalence of CHD in our population of interest is 20% (i.e., $P(\text{CHD}) = 0.2$).

Assume there are two ways of diagnosing CHD: [1] angiogram (gold standard) and [2] treadmill test (new test). How good is the treadmill test as an approximation (it's cheaper and easier to administer) to the gold standard angiogram?

To answer this question we designed a study that enrolled 50 with and 50 without CHD. In other words, we have fixed the numbers in our study based on the results of the angiogram (i.e., the gold standard), so the sample prevalence of 50% does not match the population prevalence of 20%.

In general, the prevalence estimated from a study with fixed cases will not reflect the population prevalence.

Study Results

Treadmill Test	Angiogram (GS) Positive (D)	Angiogram (GS) Negative ($\bar{D}$)	Total
Positive (T)	40	5	45
Negative ($\bar{T}$)	10	45	55
Total	50	50	100

Performance of the Test

Measuring Test Performance

Accuracy, sensitivity, and specificity are attributes of screening and diagnostic tests. These do *not* depend on the prevalence of the disease in the population and are usually estimated from studies with a large number of cases with and without disease.

Accuracy serves as an overall summary of performance, whereas sensitivity and specificity focus on performance of the test in those who do or don't have the disease, respectively.

Accuracy

One of the most straightforward measures of test performance is its **accuracy**, the proportion of correct classifications:

$$\text{Accuracy} = P(T \cap D) + P(\bar{T} \cap \bar{D}) = \frac{a+d}{a+b+c+d}$$

Interpretation: The treadmill test is ____% accurate at predicting if someone has CHD.

Sensitivity (and False Negative Rate)

Sensitivity (True Positive Rate, TPR): the probability that a test will indicate "disease" among those with the disease (i.e., how good is a test at detecting (ruling in) disease when disease is there?):

$$\text{Sensitivity} = P(T|D) = \frac{P(T \cap D)}{P(D)} = \frac{P(+\text{treadmill test} \cap \text{CHD})}{P(\text{CHD})} = \frac{a}{a+c}$$

1 - Sensitivity = 1 - TPR = **False Negative Rate (FNR)**

Interpretation: If someone has CHD, there is an ____% probability that the treadmill test will be positive.

Specificity (and False Positive Rate)

Specificity (True Negative Rate, TNR): the probability that a test will *not* indicate "disease" among those *without* the disease (i.e., how good is a test at ruling out disease when disease is not there?):

$$\text{Specificity} = P(\bar{T}|\bar{D}) = \frac{P(\bar{T} \cap \bar{D})}{P(\bar{D})} = \frac{P(\text{-treadmill test} \cap \text{no CHD})}{P(\text{no CHD})} = \frac{d}{b+d}$$

1 - Specificity = 1 - TNR = **False Positive Rate (FPR)**

Interpretation: If someone does not have CHD, there is a ____% probability that the treadmill test will be negative.

FNR and FPR Calculation Example

Based on our results from the previous two slides we can easily calculate our false negative and false positive rates:

FNR = 1 - Sensitivity = 1 - TPR =

FPR = 1 - Specificity = 1 - TNR =