Bayes' Theorem (Conditional Probability)

Example 1 - Prison and Plea

In a study of pleas and prison sentences in Arizona, it was found that 45% of the subjects studied were sent to prison. Among those sent to prison, 40% chose to plead guilty. Among those not sent to prison, 55% chose to plead not guilty. If it is known that a randomly selected subject entered a guilty plea, find the probability that this person was not sent to prison.

Prevalence: known percentage of subjects that were sent to prison = 45%

Sensitivity: percentage of true positives (pled guilty and were sent to prison) = 40%

Specificity: percentage of true negatives (pled not guilty and not sent to prison) = 55%

First, we organize the given information and desired information into a table. For the purposes of having "nice" numbers, let us assume that there are 10,000 subjects. This assumption does not affect any of the probabilities.

Status

Test Result	Sent to Prison	Not Sent to Prison	Total
Pled Guilty	B 1800	F 2475	G **4275**
Pled Not Guilty	C 2700	E 3025	H **5725**
Total	A **4500**	D **5500**	I **10000**

A: Number of subjects sent to prison = (prevalence)*(population) = 45% of 10000 = (0.45)*(10000) = 4500

B: Number of true positives = (sensitivity)*(population sent to prison) = 40% of 4500 = (0.40)*(4500) = 1800

C: Number of false positives = (total sent to prison) – (true positives) = 4500 – 1800 = 2700

D: Number not sent to prison = (total population) – (sent to prison) = 10000 – 4500 = 5500

E: Number of true negatives = (specificity)*(not sent to prison) = 55% of 5500 = (0.55)*(5500) = 3025

F: Number of false negatives = (total not sent to prison) – (true negatives) = 5500 – 3025 = 2475

G: Total who pled guilty = **B** + **F** = 1800 + 2475 = 4275

H: Total who pled not guilty = **C** + **E** = 2700 + 3025 = 5725

I: Total number of subjects = **G** + **H** = 4275 + 5725 = 10000 ***AND*** **A** + **D** = 4500 + 5500 = 10000

PVP = Predictive Value Positive = the probability that a person who pled guilty was actually sent to prison =

$$\frac{1800}{4275} \approx 0.421 = 42.1\%$$

PVN = Predictive Value Negative = the probability that a person who pled not guilty was not sent to prison =

$$\frac{3025}{5725} \approx 0.528 = 52.8\%$$

To solve the problem posed: P(not sent to prison given pled guilty) = P(not sent to prison|pled guilty) =

$$\frac{2475}{4275} \approx 0.579 = 57.9\%$$

These questions refer to the example on page 1 (the sent to prison example).

1. If a randomly selected subject pled not guilty, find the probability that this person was sent to prison.

2. If a randomly selected subject was sent to prison, find the probability that this person had pled not guilty.

3. If a randomly selected subject was not sent to prison, find the probability that this person had pled guilty.

A question involving Bayes' Theorem for you to try! – HIV

The New York State Health Department reports a 12% rate of the HIV virus for the "at-risk" population. A preliminary screening test for the HIV virus is accurate 95% of the time for those who do not have the HIV virus and has a 92% accuracy rate for those who do have the HIV virus. If someone is randomly selected from the at-risk population, what is the probability that they have the HIV virus if it is known that they have tested negative in the initial screening?

1. Construct a table summarizing the information given.

2. What is the prevalence?

3. What is the sensitivity?

4. What is the specificity?

5. What is the PVP?

6. What is the PVN?

7. What is the solution to the question posed?

Solutions to page 2 questions on the sent to prison scenario:

1. $\frac{2700}{5725} \approx .4716$

2. $\frac{2700}{4500} = .6$

3. $\frac{2475}{5500} = .45$

Solutions to page 2 questions on the HIV scenario:

1. Using just the probabilities, the table would look like this:

Status

Test Result		Has HIV	Does not have HIV	Total
	Positive Result	0.1104	0.044	**0.1544**
	Negative Result	0.0096	0.836	**0.8456**
	Total	**0.12**	**0.88**	**1.0**

Again, in order to have "nice" numbers to work with, we will assume an "at-risk" population of 10,000. Now the table looks like this:

Status

Test Result		Has HIV	Does not have HIV	Total
	Positive Result	1104	440	**1544**
	Negative Result	96	8360	**8456**
	Total	**1200**	**8800**	**10000**

2. 12%

3. 92%

4. 95%

5. With the "nice" numbers: $\frac{1104}{1544} \approx 0.7150 = 71.5\%$; using the probabilities: $\frac{0.1104}{0.1544} \approx 0.7150 = 71.5\%$

6. With the "nice" numbers: $\frac{8360}{8456} \approx 0.9886 = 98.86\%$; using the probabilities: $\frac{0.836}{0.8456} \approx 0.9886 = 98.86\%$

7. With the "nice" numbers: $\frac{96}{8456} \approx 0.0114 = 1.14\%$; using the probabilities: $\frac{0.0096}{0.8456} \approx 0.0114 = 1.14\%$