Richard Alan (Satterfield)
www.richard-alan.com .

# Conditional Probability

Example (PSU Lesson 4). A researcher is interested in evaluating how well a new diagnostic test works for detecting renal disease in patients with high blood pressure. The test is called the **BLUE** test, and she has used it on 137 patients. The test results for each of the patients was denoted in the standard way with a blue ***p*** (positive) indicating renal disease or a blue ***n*** (negative) indicating no renal disease. She wants to compare those results with known patient conditions to ascertain the effectiveness of the new test. The known conditions were established by highly complex procedures referred to as the **ORANGE** test, and whose results were denoted in the same way as her new test except they were shown in orange. Thus there are 4 possible outcomes for each patient in the study: ***pp, pn, np, nn.*** Ideally, the results from the new test should agree completely with conditions known to be true in which case the results would only have values of ***pp*** and ***nn***. But such ideals are never met in real life. The following table shows the results of her experiment. It shows a tally of each of the 4 possible outcomes, and it shows overall totals.

| | | Estimated Condition | | |
|---|---|---|---|---|
| | | ***p*** | ***n*** | ***p*** + ***n*** |
| Known Condition | ***p*** | 44 | 23 | 67 |
| | ***n*** | 10 | 60 | 70 |
| Totals | | 54 | 83 | 137 |

We can use the relative frequency approach to assigning probability by noting that of the 137 patients, 54 tested positive using the **BLUE** test. Thus,

1. P(*p*) = 54/137

Similarly we note that of the 137 patients, 67 tested positive using the **ORANGE** test. Thus,

2. P(***p***) = 67/137

Also, of the 137 patients, 44 tested positive using **BLUE** and **ORANGE** tests. Thus,

3. P(p∩***p***) = 44/137

The following summary is the preferred approach for dealing with conditional probability.

**Definition.** The **conditional probability** of an event *A* given that an event *B* has occurred is written:

P(A|B)

and is calculated using:

P(A|B) = P(A∩B) / P(B)

To apply this general formula for the renal disease example we define events A and B as follows:

4. A == ***p***

5. B == ***p*** thus,

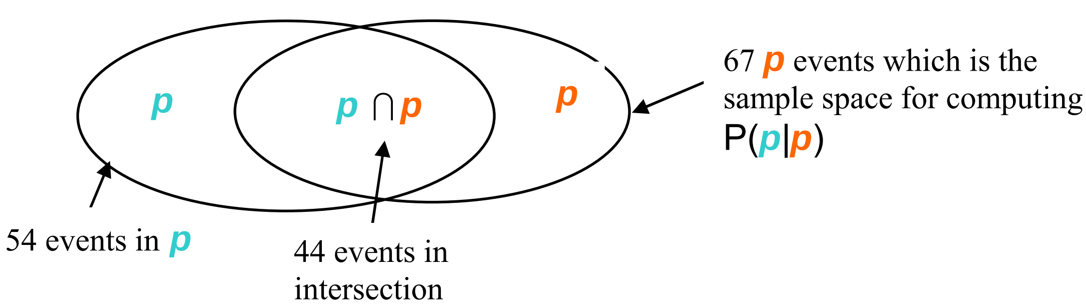
6. P(***p***|***p***) = P(***p*** ∩ ***p***) / P(***p***)

Substituting from eq. 2 and 3 gives us,

7. P(***p***|***p***) = $\frac{44/137}{67/137}$ = 44/67

<u>Observation</u>

Although the preferred approach shown in eq. 4-7 clearly gives the correct result you could express that same result directly using the relative frequency approach. We note from the table of outcomes that there were 44 positive results ***p*** of the 67 positive results ***p***. This is shown graphically in the following Venn diagram:




Applying the relative frequency approach we divide the number of events in the intersection with the number in the sample space to get,

8. P(***p***|***p***) = = 44/67

Example (Math Goodies Unit 6 Lesson 9). A math teacher gave her class two tests. 25% of the class passed both tests and 42% of the class passed the first test. What percent of those who passed the first test also passed the second test? This problem describes a conditional probability since it asks us to find the probability that the second test was passed given that the first test was passed. In the last example , the preferred notation for conditional probability seemed less apropos than the relative frequency approach. However, the opposite is true here. Using the preferred approach we have,

9. P(Second | First) = P(First and Second) / P(First)

= 0.25 / 0.42

= 60%

Observation

Clearly, the approach used to compute conditional probability depends on the problem. The above example fits the preferred approach perfectly, although it seems a bit contrived. It seems more likely the teacher would have the percentage for who passed the first test and the percentage who passed the second test. In that case you would not know the percentage who passed the first AND second test and therefore you could not deduce the answer regardless of what approach you used.