

Conditional Probability

Reading

Sections 2.1.6, 2.2.1-2.2.5

Practice Problems

2.6.2 (Page 119) 2.15, 2.16, 2.17, 2.18, 2.19, 2.20

Notes

Marginal and Joint Probabilities

In many situations our probability models come from the interaction between 2 or more categorical variables. For instance, recall the CDC data, that had amongst other information two variables, the respondents' gender and their assessment of their health. The follow contingency table shows how many respondents answered each pair of the 10 possible combinations of values:

Health	poor	fair	good	very good	excellent	Total
Female	394	1135	2953	3590	2359	10431
Male	283	884	2722	3382	2298	9569
Total	677	2019	5675	6972	4657	20000

So for instance there were 884 respondents who were male and answered "fair" when asked about their health. We can turn these into probabilities by dividing by the total count:

Health	poor	fair	good	very good	excellent	Total
Female	0.01970	0.05675	0.14765	0.17950	0.11795	0.52155
Male	0.01415	0.04420	0.13610	0.16910	0.11490	0.47845
Total	0.03385	0.10095	0.28375	0.34860	0.23285	1.00000

For example those 884 respondents correspond to a probability of $\frac{884}{20000} = 0.0442 = 4.42\%$.

We now have a probability model, consisting of 10 possible outcomes (all those combinations). If we pick a student at random, these numbers tell us the chances that the student would be in each of those 10 outcomes. This is what we call the **joint probabilities**, as it involves more than one "variable": Our outcomes are formed out of pairs of outcomes from the two separate questions, Gender and Health.

These two separate questions also have their own probability distributions, reflected

in the “Total” row and column. For instance if we pick a respondent at random there is a 52.155% chance that they would be female, and a 28.375% chance that they would be in “good” health. We call these the **marginal probabilities**.

Using the table of joint probabilities, we can answer a number of questions, for instance:

If we pick the student at random, what are the chances that they would be Female and in at least good condition?

All we have to do for this is to add the 3 relevant entries in the table, so $0.14765 + 0.17950 + 0.11795 = 0.4451$ or 44.51%.

Conditional Probability

There is another important kind of question we want to ask, and it leads us to the notion of *conditional probability*. Suppose instead the question was the following (and there are two versions of it that are equivalent):

1. If we pick a respondent at random and they turn out to be female, what are the chances that they would be in at least good condition?
2. If we pick a female respondent at random, what are the chances that they would be in at least good condition?

In both cases the effect is the same: The total group of respondents is reduced to only those that are female. So instead of adding up all those respondents that are female and in at least good condition, and finding our answer, we now need to dig further, and figure out what part that is out of all the female respondents.

So in our case we could achieve this by doing either:

$$\frac{0.14765 + 0.17950 + 0.11795}{0.52155} = \frac{0.4451}{0.52155} = 0.8534$$

or

$$\frac{2953 + 3590 + 2359}{10431} = \frac{8902}{10431} = 0.8534$$

So the answer to this question is 85.34%.

We can also ask the question in the other direction, with a different answer:

If we pick a respondent at random and they turn out to have at least good health, what are the chances that they would be a female?

In both cases, this and the previous one, we are interested in the students that are female and in at least good health. what differs is what we compare them to: all the respondents, only female respondents, or only respondents that are in good health?

For this last case, we compare to all those that are in at least good health, so we would have:

$$\frac{0.14765 + 0.17950 + 0.11795}{0.28375 + 0.34860 + 0.23285} = \frac{0.4451}{0.8652} = 0.51445$$

Now we are ready for the general definition:

The probability of an event *given another event* is called **conditional probability**. It is one of the hardest concepts to master.

$P(A|B)$ is the probability of A occurring, given that B does occur.

Think of it as “restricting the sample space”: Only outcomes in B matter; how many of those are also in A ?

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Example: In the case of the 5-sided die, let us compute $P(A|B)$.

In words this says: Given the number is at least 4, what are the chances the number is even?

It will be: $\frac{P(4)}{P(4 \text{ or } 5)} = \frac{1/5}{2/5} = 1/2$.

Notice the difference between $P(A)$ and $P(A|B)$!!!