On explaining $P(A|B)$ as fraction $\frac{P(A,B)}{P(B)}$

September 18, 2018

I mentioned in the lectures that $P(A|B)$ is defined as a ratio of two probabilities, $\frac{P(A,B)}{P(B)}$, though when you calculate you invariably work with a ratio of two counts $\frac{\text{count}(A,B)}{\text{count}(B)}$. Here’s an attempt to recapitulate the intuition I gave in the lecture for why this ratio of probabilities idea makes sense.

First suppose the following table represents a literal ‘table’ in a market on which second-hand mobile phones are placed. Suppose there are two brands Apple and HTC and that they fall into two prices Expensive and Cheap.

<table>
<thead>
<tr>
<th></th>
<th>Exp</th>
<th>Cheap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>50 (.25)</td>
<td>20 (.1)</td>
</tr>
<tr>
<td>HTC</td>
<td>30 (.15)</td>
<td>100 (.5)</td>
</tr>
</tbody>
</table>

The cells of the table show primarily the counts of the different phones available. There’s 200 in all and in brackets then a probability is shown, which divides the count by 200. You could think of this as the probability of grabbing a phone of that type (ie. brand and price-band) if you ran past this market stall and tried to steal one.

You should be able to readily work out things like $P(\text{cheap}|\text{Apple})$, by just looking at the Apple row, which has 70 phones in all, noting that 20 of those are cheap so $P(\text{cheap}|\text{Apple}) = 20/70 = 2/7$. This is following the straightforward intuition that for this conditional probability you zero in on the Apple outcomes, counting how many of these there are, and then within these, you see how many are also Cheap outcomes, and take the ratio.

The picture below illustrates how things can be thought of in a hierarchical way

Moving from the top-most down to the left represents the possibility the phone is an Apple phone, and the probability of that is 0.35. The two nodes beneath that then represent the price aspect, with the node to the left below saying the Apple is also cheap and the node to the right that the Apple is ¬cheap. These joint probabilities can be read off the table and are 0.1 and 0.25.

Now clearly there must be numbers (1) and (2) which scale the Apple prob of 0.35, down into these two alternatives:
0.35 \times (1) = 0.1 \quad \Rightarrow \quad (1) = \frac{0.1}{0.35} = \frac{2}{7}

0.35 \times (2) = 0.25 \quad \Rightarrow \quad (2) = \frac{0.25}{0.35} = \frac{5}{7}

Also, not at all accidentally, these two numbers (1) and (2) sum to 1, because 0.35 = 0.1 + 0.25.

Thus for cheap you have a number (1) which turns $P(Apple)$ into $P(Apple, cheap)$, and for ¬cheap you have a number (2) which turns $P(Apple)$ into $P(Apple, ¬cheap)$, and these sum to 1.

This motivates regarding (1) and (2) as a special kind of probability, for which people hit on the name conditional probability, and also hit on the notations $P(cheap|Apple)$ and $P(¬cheap|Apple)$ for (1) and (2).

If you bear in mind this kind of motivation for conditional probability it also drives home the most conspicuous ‘purpose’ of $P(A|B)$, which is to be multiplied by $P(B)$ and give $P(A, B)$, that is to generate a joint probability.