Conditional Conditionals

It is an interesting aspect of human language that exactly the same wording can mean exactly different things, with each variation hiding subtle nuances of meaning and context. This well-established principle makes human language expressive and evocative but also tends to obscure the channels of communication where precise communication is required. Case in point, the use of conditional probability.

Consider the following two ‘cause-and-effect’ scenarios. In the first, we survey two groups, each with 140 individuals, who are afflicted with similar diseases: a mild case of the flu and the common cold. For our research purposes, we focus on only two of the set of symptoms common to both ailments and ask whether an individual complains more about having a sore throat or body aches.

Given the current statistics, medical community estimates that for a case of mild flu, patients rank the sore throat as more bothersome than body aches 57.14% of the time and body aches more bothersome 42.86% of the time. The corresponding rankings for the common cold are 63.33% and 36.67% of the time for sore throat and body aches, respectively.

These rankings represent the conditional probabilities stating that if we know that an individual has a given disease then we should, within some statistical fluctuation, be able to say the probability that that individual will rank a sore throat ahead of body aches. Alternatively, given a sample of patient, each known to have either a mild flu or the common cold, we should be able to say what proportion of them will rank which symptom as being worse.

To be concrete, consider that we have $N\_f = 140$ people suffering with the flu and $N\_c = 150$ with the common cold. The proportions of these individuals that are predicted to rank the sore throat more would be given by the product of the conditional probability by the total size of the group. Mathematically, the number of flu patients $N\_{sf}$ ranking the sore throat as the primary problem is

\[ N\_{sf} = P(s|f) N\_f = 0.5714 \cdot N\_f = 80 \; , \]

while the number of cold patients $N\_{sc}$ agreeing that the sore throat is worse is

\[ N\_{sc} = P(s|c) N\_c = 0.4286 \cdot 140 = 60 \; . \]

In the same way, the number of flu and cold patients ranking body aches first can be obtained. The results summarize nicely in the usual joint probability table we know from statistics

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sore Throat | Body Aches |  |
| Flu | 80 | 60 | 140 |
| Cold | 95 | 55 | 150 |
|  | 175 | 115 |  |

This is one way, and probably the most familiar way, of defining what is meant by conditional probability. Other textbook examples involve surveying a population for attributes like the [Christmas tree light example](http://aristotle2digital.blogwyrm.com/?p=239) with the joint probability table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Red | Blue | Green | Yellow |  |
| Short Life | 0.12 | 0.12 | 0.10 | 0.06 | 0.40 |
| Medium Life | 0.105 | 0.105 | 0.0875 | 0.0525 | 0.35 |
| Long Life | 0.075 | 0.075 | 0.0625 | 0.0375 | 0.25 |
|  | 0.30 | 0.30 | 0.25 | 0.15 |  |

In order to follow what comes next, it is important to note that having the conditional probabilities is equivalent to having the table or vice versa. Having the complete information represented in one form allows the unambiguous construction of the other.

Interestingly, there seems to be a case in which the term ‘conditional probability’ is used such that, as far as I can tell, the values can’t be meaningfully cast into table form like the above. This case occurs when the use of the phrase ‘conditional probability’ is meant as a transition probability. A textbook example of this is given in terms of the weather in the Wikipedia article on [*Examples of Markov Chains*](https://en.wikipedia.org/wiki/Examples_of_Markov_chains#A_simple_weather_model).

In this example, the conditional probabilities represent the chance that the weather tomorrow will be like the weather today. For the sake of simplicity, the only two possible weather outcomes are ‘Sunny’ and ‘Rainy’. The chance that it will be Sunny tomorrow given that it is Sunny today is 0.9. Likewise, the chance that it will be Rainy tomorrow given that it is Sunny today is 0.1. These facts are summarized in the conditional probabilities

\[ P(S|S) = 0.9 \; \]

and

\[ P(R|S) = 0.1 \; . \]

In a similar fashion, the conditional probabilities for a Sunny or Rainy tomorrow given that it is Rainy today are fifty-fifty

\[ P(S|R) = 0.5 = P(R|R) \; .\]

These conditional probabilities look, at first glance, to be on the same footing as the earlier ones. The wording is the same; the notation is the same. However, the interpretation is quite different. The difference seems to lie in the fact that in the first case, the notion of having the flu or a cold and ranking which was worse, a sore throat or body aches, are underlying attributes of the sick individual. In the second case, sunny or rainy weather is an attribute of the environment plain enough but the fact that there are an unlimited number of possibilities (e.g. SSRS for a four day sequence of weather with S = Sunny and R = Rainy) precludes making the usual table.

Mathematically, the heart of the difficult lies in the usual definition of a conditional probability in terms of sets. For two possible outcomes $A$ and $B$ in a universe of possibilities the conditional probability that $B$ will occur given $A$ is

\[ P(B|A) = \frac{P\left( B \bigcap A \right)}{P(A)} \; , \]

where $P\left(B \bigcap A\right)$ is the probability of $A$ and $B$ occurring together (hence appearing in the joint probability distribution). For example, in the Christmas tree light example above, the probability that a bulb chosen at random is both yellow and has a long life is

\[ P\left(long \bigcap yellow \right) = 0.0375 \; ,\]

while the probability of choosing a yellow bulb to begin with is

\[ P(yellow) = 0.15 \; . \]

The conditional probability given that the bulb is yellow that it also has a long life is

\[ P(long|yellow) = \frac{ P\left( long \bigcap yellow \right)}{P(yellow)} = \frac{0.0375}{0.15} = 0.25 \; . \]

So, we can move easily between conditional probabilities and joint probability distributions in this case.

In the case of the transition probabilities between weather, it isn’t clear how to encode the event $S$ as sunny today versus $S$ as sunny tomorrow. The concepts of today and tomorrow are relative and constantly shifting during the passage of time.

One’s first reaction to that might be to not care. But then how does one answer the simple question as to the proportion of days that are sunny? This will be the topic of next month’s column.