Handout 05: Conditional Probability

Definitions

Suppose we have a box containing ten marbles, some large, some small, some red, some white; the exact contents of the box is represented below:

	Large	Small	Totals
Red	3	2	5
White	4	1	5
Totals	7	3	10

Assume that the marbles are well mixed, and that a single marble is drawn from the box at random. Since it is reasonable to think of each of the ten marbles as having the same chance of selection, you would undoubtedly compute the probability that a red marble is drawn as

$$\mathbb{P}(R) = \frac{5}{10} = \frac{1}{2}$$

Suppose, however, that I was the one to draw the marble, and without showing it to you, I proclaimed that the marble drawn was large. You would be forced to rethink your probability assignment, and you would undoubtedly be led to the alternative computation

$$\mathbb{P}(R|L) = \frac{3}{7}$$

The new symbol introduced here is pronounced "the probability of red, given large"; that is, the vertical slash is simply read as "given." We make this intuition formal by the following definition.

Definition 1 (Conditional Probability) *Consider a pair of events A and B associated with a random experiment, and suppose that $\mathbb{P}(B) > 0$. Then the conditional probability of A, given B, is defined as:*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Note that the conditional probability $P(A|B)$ remains undefined when the event B has probability zero. In discrete problems, one does not encounter the need to condition on an event that has no a priori chance of happening. This need does arise in the continuous case, however. Our treatment of conditional probability will be suitably generalized once we get there.

In our experiment with ten marbles, the formal definition of conditional probability yields:

$$\mathbb{P}(R|L) = \frac{\mathbb{P}(R \cap L)}{\mathbb{P}(L)} = \frac{3/10}{7/10} = \frac{3}{7}.$$

Fortunately, this computation agrees with our intuitive derivation.

Theorem 1 *For any $A \subset B$, define a new probability function $\mathbb{P}'(A) = P(A|B)$. All of the probability axioms hold for the pair $(B, \mathbb{P}'(\cdot))$.*

Because conditional probabilities "inherit" the three characteristics assumed to hold for the original probability function P, they also inherit all properties derived from these assumptions. We have already proven that this property holds for any set function satisfying the three fundamental axioms.

Theorem 2 (Multiplication Rule) *For any sets A and B where $P(B) > 0$, we have:*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A).$$

Proof. This is a simple re-arranging of the definition of conditional probabilities.

Bayes' Theorem

There is a famous result that we will frequently use that relates the conditional probability $\mathbb{P}(A|B)$ to the conditional probability $\mathbb{P}(B|A)$:

Theorem 3 (Bayes Theorem) *For any two events A and B such that $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, we have:*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A) \cdot \mathbb{P}(B|A)}{\mathbb{P}(B)}.$$

Proof. From the multiplication rule and commutivity of set intersection we have:

$$\begin{aligned}
\mathbb{P}(B) \cdot \mathbb{P}(A|B) &= \mathbb{P}(B \cap A) \\
&= \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) \cdot \mathbb{P}(B|A)
\end{aligned}$$

Dividing both sides by $\mathbb{P}(B)$ yields the result $\blacksquare$.

Independence

A natural definition of the independence of two events can be given by the relationship:

$$\mathbb{P}(A|B) = \mathbb{P}(A).$$

The difficulty with using this as a definition is that it does not hold when either A or B have a probability equal to zero. A more general definition is possible, from which the intuitive defintion follows whenever the respective events have non-zero probabilities.

Definition 2 (Independence) *We say that any events A and B are independent if and only if* $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Recognizing that two events are independent is not always a simple matter. The independence of two events can only be confirmed by calculating the relevant probabilities. The answer you derive will usually agree with your intuition, although we will encounter a situation or two where the answer is surprising.

Students often get independent events confused with mutually exclusive events. This appears to be due to the words involved, which seem somewhat similar, rather than to the closeness of the ideas themselves. Actually, mutually exclusive events are about as dependent as events can get. Recall the events A and B are mutually exclusive if $A \cap B = \emptyset$. This means that these events cannot happen simultaneously. If A occurs (that is, if the outcome of the experiment is a member of A), then B cannot possibly occur. It follows that for two mutually exclusive events A and B for which $P(A) > 0$ and $P(B) > 0$, we have $P(B|A) = 0 \neq P(B)$, that is, A and B are not independent. It should be intuitively clear that the occurrence of one of two mutually exclusive events carries a lot of information about the chance that the other event will occur, while the occurrence of one of two independent events carries none.