Lecture 6 - Conditional Probability I

Example: Toss 2 fair dice. If the first die shows 3, what is the probability that the sum of the 2 values equals 8?

Initially, the sample space is \( S = \{(i, j) : 1 \leq i, j \leq 6\} \). But we are told that the first die shows 3, so we know that most of the outcomes in \( S \) are irrelevant. The six remaining possibilities are

\[
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
\]

and exactly one of these – \((3, 5)\) – gives a sum of 8. These six remaining possibilities were initially equally likely, so a sensible answer is \(\frac{1}{6}\).

The modeling choice here is that if all outcomes in \( S \) were equally likely, and we are given information that a certain event happens, e.g. \( F = \{1st \, die \, shows \, 3\} \) we assume that the individual outcomes within \( F \) are still equally likely.

This idea can be generalized. Consider any experiment with sample space \( S \) and probability distribution \( P \). Let \( F \subset S \) be an event, and imagine that we are told that ‘\( F \) happens’, but are not given any other information about the outcome. In the light of this information, we should update our choice of the probability values: \( P \) won’t be a sensible choice any more.

For instance, if \( E \subset S \) is mutually exclusive with \( F \), then the updated probability of \( E \) should be zero, while our initial model may give \( P(E) \neq 0 \). The natural choice is to replace them with the ratios of probability value computed within the event \( F \).
The conditional probability that $E$ occurs given that $F$ has occurred is defined by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

if $P(F) > 0$.

Note

- $P(F|F) = P(F)/P(F) = 1$ ("conditioned on the event that $F$ happens, it is certain that $F$ happens"); and
- If $E \cap F = \emptyset$, then $P(E|F) = P(\emptyset)/P(F) = 0$ ("if $E$ and $F$ are mutually exclusive, then conditioned on the event that $F$ happens, $E$ has zero probability of happening").

Moreover, if we want to update our model of the experiment using the event that $F$ happens, then we can regard this as a whole new experiment. We keep the original sample space $S$, but now the probability values for $E \subset S$ are the numbers $P(E|F)$. This works because:

**Theorem 1.** (Ross Prop 3.5.1) If we fix $F \subset S$ with $P(F) > 0$, then the values $P(E|F)$ for all possible events $E \subset S$ satisfy the axioms of probability:

(a) $0 \leq P(E|F) \leq 1$;
(b) $P(S|F) = 1$;
(c) (Countable additivity) For any sequence of mutually exclusive events $E_i$ (that is events for which $E_i \cap E_j = \emptyset$ when $i \neq j$),

$$P\left(\bigcup_{i=1}^{\infty} E_i|F\right) = \sum_{i=1}^{\infty} P(E_i|F)$$

See Ross for a proof of this result.

**Choosing the sample space** Since $P(F|F) = 1$, the new conditioned probabilities put all the ‘weight’ of probability on outcomes inside $F$: any events which lie outside $F$ (i.e., are mutually exclusive with $F$) get probability zero. So an alternative is to also replace $S$ with the smaller sample space $F$, and just consider the probability values $P(E|F)$ for $E \subset F$. Either way is correct, but one may be more convenient than the other in a given situation. Compare Examples 3.2b and 3.2c.
For many experiments, it is more natural to choose how to model the conditional probabilities first and then derive unconditional probabilities from them:

\[ P(E \cap F) = P(E|F)P(F) \]

We see that the probability of \( E \) and \( F \) occurring is the probability that \( F \) has occurred times the probability that \( E \) has occurred given than \( F \) has occurred. Compare (a) in Example 3.2e where for the second draw all remaining outcomes conditioned on the color of the first ball are equally likely with (b) where they are not.

This result can be generalized to apply iteratively.

**Theorem 2.** (The multiplication rule) If \( P(E_1) > 0, \ldots, P(E_n) > 0 \),

\[ P(E_1 \cap \ldots \cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \ldots P(E_n|E_1 \cap \ldots \cap E_{n-1}) \]

*Proof.* Apply the definition of conditional probability to its right-hand side.

\[
P(E_1) \frac{P(E_2 \cap E_1)}{P(E_1)} \frac{P(E_1 \cap E_2 \cap E_3)}{P(E_1 \cap E_2)} \ldots \frac{P(E_1 \cap \ldots \cap E_n)}{P(E_1 \cap \ldots \cap E_{n-1})}
\]

\[ \square \]

See Examples 3.2f, 3.2g.

**References**