6 Conditional Probability

If we have partial information, this effectively shrinks the available sample space and changes the probabilities. We begin with an example.

Monty Hall show. The setting is a game show in which a prize is hidden behind one of three curtains. Call the curtains X, Y, and Z. You can win the prize by guessing the right curtain.

STEP 1. You choose a curtain.

This leaves two curtains you did not choose, and at least one of them does not hide the prize. Monty Hall opens this one curtain and this way demonstrates there is no prize hidden there. Then he asks whether you would like to reconsider. Would you?

STEP 2A. You stick to your initial choice.

STEP 2B. You change to the other available curtain.

Perhaps surprisingly, Step 2B is the better strategy. As shown in Figure 5, it doubles your chance to win the prize.

\[ P(AB) = P(A|B) \cdot P(B), \]
\[ P(ABC) = P(A|BC) \cdot P(B|C) \cdot P(C), \]

where in the second line B is conditioned on C, etc. Next, we rewrite the formula in a form that is often useful. Suppose \( \Omega = H_1 \sqcup H_2 \sqcup \ldots \sqcup H_n \) is a partition of the sample space. Then

\[ P(A) = P(A|H_1) + P(A|H_2) + \ldots + P(A|H_n) \]
\[ = \sum_{i=1}^{n} P(A|H_i) \cdot P(H_i). \]

Now fix an integer \( k \) and note that the probability of \( H_k \) given \( A \) is \( P(H_k|A) = P(H_kA)/P(A) \). Of course, \( P(H_kA) = P(AH_k) \), so we can write

\[ P(H_k|A) = \frac{P(A|H_k) \cdot P(H_k)}{\sum_{i=1}^{n} P(A|H_i) \cdot P(H_i)}, \]

which is usually referred to as Bayes’ Rule. It is often used in applications, for example for selecting candidates from different interpretations of data.

Law of succession. Consider a scenario in which we have \( N + 1 \) urns with \( N \) balls each. For \( k \) from 0 to \( n \), the \( k \)-th urn contains \( k \) red balls and \( n - k \) blue balls.

STEP 1. Select an urn randomly.

STEP 2. Draw \( m \) balls with replacement from the selected urn.

Let \( H \) be the event that all \( m \) balls are red. Now we draw another ball from the same urn, and we ask what is the probability that the \((m + 1)\)-st ball is also red. To analyze this question, we write \( A \) for the event that all \( m + 1 \) balls
are red, so we are interested in \( P(A|H) \). By definition, we have \( P(A|H) = P(AH)/A(H) = P(A)/P(H) \). Assuming we select the \( k \)-th urn, the probability of drawing \( m \) red balls in sequence is \((k/n)^m\). Hence, 

\[
P(H) = \frac{1^m + 2^m + \ldots + n^m}{n^m(n+1)}
\]

\[
\approx \frac{1}{n^m(n+1)} \int_0^n x^m \, dx 
\]

\[
= \frac{n}{n+1}.
\]

Assuming \( n \) is very large, we feel justified to take the limit, which is \( \lim_{n \to \infty} P(H) = \frac{1}{m+1} \). Similarly, \( \lim_{n \to \infty} P(A) = \frac{1}{m+1} \). Hence,

\[
\lim_{n \to \infty} P(A|H) = \frac{m+1}{m+2}.
\]

This is called the Law of Succession by Laplace (1812), in particular in the case in which the \( n+1 \) urns are replaced by one urn but we do not know the fraction of red versus blue balls so we assume this fraction is randomly distributed as explained.

**Independent events and trial processes.** We say \( A \) and \( B \) are independent if knowing \( B \) does not change the probability of \( A \), that is,

\[
P(A | B) = P(A).
\]

Since \( P(A | B) = \frac{P(A \cap B)}{P(B)} = P(A) \), we have

\[
P(B) = \frac{P(B \cap A)}{P(A)} = \frac{P(B | A)}{P(A)}.
\]

We thus see that independence is symmetric. Combining the definition of conditional probability with the condition of independence, we get a formula for the probability of two events occurring at the same time: if \( A \) and \( B \) are independent then

\[
P(A \cap B) = P(A) \cdot P(B).
\]

In many situations, a probabilistic experiment is repeated, possibly many times. We call this a trial process. It is independent if the \( i \)-th trial is not influenced by the outcomes of the preceding \( i-1 \) trials:

\[
P(A_i | A_1 \cap \ldots \cap A_{i-1}) = P(A_i),
\]

for each \( i \). A particular example is the Bernoulli trial process in which the probability of success is the same at each trial:

\[
P(\text{success}) = p; \quad P(\text{failure}) = 1 - p.
\]

If we do a sequence of \( n \) trials, we may define \( X \) equal to the number of successes. Hence, \( \Omega \) is the space of possible outcomes for a sequence of \( n \) trials or, equivalently, the set of binary strings of length \( n \). What is the probability of getting exactly \( k \) successes? The probability of having a sequence of \( k \) successes followed by \( n-k \) failures is \( p^k(1-p)^{n-k} \). Now we just have to multiply with the number of binary sequences that contain \( k \) successes.

**Binomial Probability Lemma.** The probability of having exactly \( k \) successes in a sequence of \( n \) trials is

\[
P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.
\]

As a sanity check, we make sure that the probabilities add up to one. Using the Binomial Theorem, we get

\[
\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k},
\]

which is equal to \((p + (1-p)) = 1\). Because of this connection, the probabilities in the Bernoulli trial process are called the binomial probabilities.

**Medical test example.** Trial processes that are not independent are generally more complicated and we need more elaborate tools to compute the probabilities. A useful such tool is the tree diagram as shown in Figure 7. We explain this using a realistic medical test problem. The outcome will show that probabilities can be counterintuitive, even in situations in which they are important. Consider a medical test for a disease, \( D \). The test mostly gives the right
answer, but not always. Say its false-negative rate is 1% and its false-positive rate is 2%:

\[
\begin{align*}
    P(y \mid D) &= 0.99; \\
    P(n \mid D) &= 0.01; \\
    P(y \mid \neg D) &= 0.02; \\
    P(n \mid \neg D) &= 0.98.
\end{align*}
\]

Assume that the chance you have disease \( D \) is only one in a thousand, that is, \( P(D) = 0.001 \). Now you take the test and the outcome is positive. What is the chance that you have the disease? In other words, what is \( P(D \mid y) \)? As illustrated in Figure 7,

\[
P(D \mid y) = \frac{P(D \cap y)}{P(y)} = \frac{0.00099}{0.02097} = 0.047 \ldots
\]

This is clearly a case in which you want to get a second opinion before starting a treatment.