17.7 Independence

Suppose that we flip two fair coins simultaneously on opposite sides of a room. Intuitively, the way one coin lands does not affect the way the other coin lands. The mathematical concept that captures this intuition is called independence.

**Definition 17.7.1.** An event with probability 0 is defined to be independent of every event (including itself). If $\Pr[B] \neq 0$, then event $A$ is independent of event $B$ iff

$$\Pr[A \mid B] = \Pr[A].$$  \hfill (17.4)

In other words, $A$ and $B$ are independent if knowing that $B$ happens does not alter the probability that $A$ happens, as is the case with flipping two coins on opposite sides of a room.
Potential Pitfall

Students sometimes get the idea that disjoint events are independent. The opposite is true: if \( A \cap B = \emptyset \), then knowing that \( A \) happens means you know that \( B \) does not happen. Disjoint events are never independent—unless one of them has probability zero.

17.7.1 Alternative Formulation

Sometimes it is useful to express independence in an alternate form which follows immediately from Definition 17.7.1:

**Theorem 17.7.2.** \( A \) is independent of \( B \) if and only if

\[
\Pr[A \cap B] = \Pr[A] \cdot \Pr[B].
\]

(17.5)

Notice that Theorem 17.7.2 makes apparent the symmetry between \( A \) being independent of \( B \) and \( B \) being independent of \( A \):

**Corollary 17.7.3.** \( A \) is independent of \( B \) iff \( B \) is independent of \( A \).

17.7.2 Independence Is an Assumption

Generally, independence is something that you assume in modeling a phenomenon. For example, consider the experiment of flipping two fair coins. Let \( A \) be the event that the first coin comes up heads, and let \( B \) be the event that the second coin is heads. If we assume that \( A \) and \( B \) are independent, then the probability that both coins come up heads is:

\[
\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.
\]

In this example, the assumption of independence is reasonable. The result of one coin toss should have negligible impact on the outcome of the other coin toss. And if we were to repeat the experiment many times, we would be likely to have \( A \cap B \) about 1/4 of the time.

On the other hand, there are many examples of events where assuming independence isn’t justified. For example, an hourly weather forecast for a clear day might list a 10% chance of rain every hour from noon to midnight, meaning each hour has a 90% chance of being dry. But that does not imply that the odds of a rainless day are a mere \( 0.9^{12} \approx 0.28 \). In reality, if it doesn’t rain as of 5pm, the odds are higher than 90% that it will stay dry at 6pm as well—and if it starts pouring at 5pm, the chances are much higher than 10% that it will still be rainy an hour later.
Deciding when to assume that events are independent is a tricky business. In practice, there are strong motivations to assume independence since many useful formulas (such as equation (17.5)) only hold if the events are independent. But you need to be careful: we’ll describe several famous examples where (false) assumptions of independence led to trouble. This problem gets even trickier when there are more than two events in play.

17.8 Mutual Independence

We have defined what it means for two events to be independent. What if there are more than two events? For example, how can we say that the flips of \( n \) coins are all independent of one another? A set of events is said to be mutually independent if the probability of each event in the set is the same no matter which of the other events has occurred. This is equivalent to saying that for any selection of two or more of the events, the probability that all the selected events occur equals the product of the probabilities of the selected events.

For example, four events \( E_1, E_2, E_3, E_4 \) are mutually independent if and only if all of the following equations hold:

\[
\begin{align*}
\Pr[E_1 \cap E_2] &= \Pr[E_1] \cdot \Pr[E_2] \\
\Pr[E_1 \cap E_3] &= \Pr[E_1] \cdot \Pr[E_3] \\
\Pr[E_1 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_4] \\
\Pr[E_2 \cap E_3] &= \Pr[E_2] \cdot \Pr[E_3] \\
\Pr[E_2 \cap E_4] &= \Pr[E_2] \cdot \Pr[E_4] \\
\Pr[E_3 \cap E_4] &= \Pr[E_3] \cdot \Pr[E_4] \\
\Pr[E_1 \cap E_2 \cap E_3] &= \Pr[E_1] \cdot \Pr[E_2] \cdot \Pr[E_3] \\
\Pr[E_1 \cap E_2 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_2] \cdot \Pr[E_4] \\
\Pr[E_1 \cap E_3 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_3] \cdot \Pr[E_4] \\
\Pr[E_2 \cap E_3 \cap E_4] &= \Pr[E_2] \cdot \Pr[E_3] \cdot \Pr[E_4] \\
\Pr[E_1 \cap E_2 \cap E_3 \cap E_4] &= \Pr[E_1] \cdot \Pr[E_2] \cdot \Pr[E_3] \cdot \Pr[E_4]
\end{align*}
\]

The generalization to mutual independence of \( n \) events should now be clear.

17.8.1 DNA Testing

Assumptions about independence are routinely made in practice. Frequently, such assumptions are quite reasonable. Sometimes, however, the reasonableness of an
17.8. Mutual Independence

Independence assumption is not so clear, and the consequences of a faulty assumption can be severe.

Let’s return to the O. J. Simpson murder trial. The following expert testimony was given on May 15, 1995:

**Mr. Clarke:** When you make these estimations of frequency—and I believe you touched a little bit on a concept called independence?

**Dr. Cotton:** Yes, I did.

**Mr. Clarke:** And what is that again?

**Dr. Cotton:** It means whether or not you inherit one allele that you have is not—does not affect the second allele that you might get. That is, if you inherit a band at 5,000 base pairs, that doesn’t mean you’ll automatically or with some probability inherit one at 6,000. What you inherit from one parent is what you inherit from the other.

**Mr. Clarke:** Why is that important?

**Dr. Cotton:** Mathematically that’s important because if that were not the case, it would be improper to multiply the frequencies between the different genetic locations.

**Mr. Clarke:** How do you—well, first of all, are these markers independent that you’ve described in your testing in this case?

Presumably, this dialogue was as confusing to you as it was for the jury. Essentially, the jury was told that genetic markers in blood found at the crime scene matched Simpson’s. Furthermore, they were told that the probability that the markers would be found in a randomly-selected person was at most 1 in 170 million. This astronomical figure was derived from statistics such as:

- 1 person in 100 has marker $A$.
- 1 person in 50 marker $B$.
- 1 person in 40 has marker $C$.
- 1 person in 5 has marker $D$.
- 1 person in 170 has marker $E$. 
Then these numbers were multiplied to give the probability that a randomly-selected person would have all five markers:

\[
\Pr[A \cap B \cap C \cap D \cap E] = \Pr[A] \cdot \Pr[B] \cdot \Pr[C] \cdot \Pr[D] \cdot \Pr[E]
\]

\[
= \frac{1}{100} \cdot \frac{1}{50} \cdot \frac{1}{40} \cdot \frac{1}{5} \cdot \frac{1}{170} = \frac{1}{170,000,000}.
\]

The defense pointed out that this assumes that the markers appear mutually independently. Furthermore, all the statistics were based on just a few hundred blood samples.

After the trial, the jury was widely mocked for failing to “understand” the DNA evidence. If you were a juror, would you accept the 1 in 170 million calculation?

### 17.8.2 Pairwise Independence

The definition of mutual independence seems awfully complicated—there are so many selections of events to consider! Here’s an example that illustrates the subtlety of independence when more than two events are involved. Suppose that we flip three fair, mutually-independent coins. Define the following events:

- \(A_1\) is the event that coin 1 matches coin 2.
- \(A_2\) is the event that coin 2 matches coin 3.
- \(A_3\) is the event that coin 3 matches coin 1.

Are \(A_1\), \(A_2\), \(A_3\) mutually independent?

The sample space for this experiment is:

\[
\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.
\]

Every outcome has probability \((1/2)^3 = 1/8\) by our assumption that the coins are mutually independent.

To see if events \(A_1\), \(A_2\), and \(A_3\) are mutually independent, we must check a sequence of equalities. It will be helpful first to compute the probability of each event \(A_1\):

\[
\Pr[A_1] = \Pr[HHH] + \Pr[HHT] + \Pr[THH] + \Pr[TTH] + \Pr[TTT]
\]

\[
= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}.
\]
By symmetry, $\Pr[A_2] = \Pr[A_3] = 1/2$ as well. Now we can begin checking all the
equalities required for mutual independence:

$$
\Pr[A_1 \cap A_2] = \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}
$$

By symmetry, $\Pr[A_1 \cap A_3] = \Pr[A_1] \cdot \Pr[A_3]$ and $\Pr[A_2 \cap A_3] = \Pr[A_2] \cdot \Pr[A_3]$ must hold also. Finally, we must check one last condition:

$$
\Pr[A_1 \cap A_2 \cap A_3] = \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \\
\neq \frac{1}{8} = \Pr[A_1] \Pr[A_2] \Pr[A_3].
$$

The three events $A_1$, $A_2$, and $A_3$ are not mutually independent even though any
two of them are independent! This not-quite mutual independence seems weird at
first, but it happens. It even generalizes:

**Definition 17.8.1.** A set $A_1, A_2, \ldots$, of events is $k$-way independent iff every set
of $k$ of these events is mutually independent. The set is pairwise independent iff it
is 2-way independent.

So the events $A_1$, $A_2$, $A_3$ above are pairwise independent, but not mutually inde-
pendent. Pairwise independence is a much weaker property than mutual indepen-
dence.

For example, suppose that the prosecutors in the O. J. Simpson trial were wrong
and markers $A$, $B$, $C$, $D$, and $E$ appear only pairwise independently. Then the
probability that a randomly-selected person has all five markers is no more than:

$$
\Pr[A \cap B \cap C \cap D \cap E] \leq \Pr[A \cap E] = \Pr[A] \cdot \Pr[E] \\
= \frac{1}{100} \cdot \frac{1}{170} = \frac{1}{17,000}.
$$

The first line uses the fact that $A \cap B \cap C \cap D \cap E$ is a subset of $A \cap E$. (We picked
out the $A$ and $E$ markers because they’re the rarest.) We use pairwise independence
on the second line. Now the probability of a random match is $1$ in $17,000$—a far cry
from $1$ in $170$ million! And this is the strongest conclusion we can reach assuming
only pairwise independence.

On the other hand, the $1$ in $17,000$ bound that we get by assuming pairwise
independence is a lot better than the bound that we would have if there were no
independence at all. For example, if the markers are dependent, then it is possible that
everyone with marker \( E \) has marker \( A \),
everyone with marker \( A \) has marker \( B \),
everyone with marker \( B \) has marker \( C \), and
everyone with marker \( C \) has marker \( D \).

In such a scenario, the probability of a match is

\[
\Pr[E] = \frac{1}{170}.
\]

So a stronger independence assumption leads to a smaller bound on the probability of a match. The trick is to figure out what independence assumption is reasonable. Assuming that the markers are \textit{mutually} independent may well \textit{not} be reasonable unless you have examined hundreds of millions of blood samples. Otherwise, how would you know that marker \( D \) does not show up more frequently whenever the other four markers are simultaneously present?