# 2 Probability on Events

In this chapter we introduce probability on events. We follow an axiomatic approach that uses elementary set theory.

## 2.1 Sample Space and Events

Probability is typically defined in terms of some **experiment**. The **sample space**, $\Omega$, of the experiment is the set of all possible outcomes of the experiment.

**Definition 2.1** *An* **event***, $E$, is any subset of the sample space, $\Omega$.*

For example, in an experiment where a die is rolled twice, each outcome (a.k.a. sample point) is denoted by the pair $(i, j)$, where $i$ is the first roll and $j$ is the second roll. There are 36 sample points. The event

$$E = \{\ (1,3) \text{ or } (2,2) \text{ or } (3,1)\ \}$$

denotes that the sum of the die rolls is 4.

In general, the sample space may be *discrete*, meaning that the number of outcomes is finite, or at least countably infinite, or *continuous*, meaning that the number of outcomes is uncountable.

One can talk of unions and intersections of events, because they are also sets. For example, we can talk of $E \cup F$, $E \cap F$, and $\overline{E}$. Here, $E$ and $F$ are events and $\overline{E}$, the complement of $E$, denotes the set of points in $\Omega$ but not in $E$, also written $\Omega \setminus E$.

**Question:** For the die-rolling experiment, consider events $E_1$ and $E_2$ defined on $\Omega$ in Figure 2.1. Do you think that $E_1$ and $E_2$ are independent?

**Answer:** No, they are not independent. We get to this later when we define independence. We say instead that $E_1$ and $E_2$ are mutually exclusive.

$$
\begin{array}{ccccccc}
 & & E_1 & & & E_2 & \\
\Omega = & \left\{ \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array}\right. \hspace{-2em}
\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array} & & & & &
\end{array}
$$

**Figure 2.1** *Illustration of two mutually exclusive events in sample space* $\Omega$.

**Definition 2.2** *If $E_1 \cap E_2 = \emptyset$, then $E_1$ and $E_2$ are* **mutually exclusive**.

**Definition 2.3** *If $E_1, E_2, \ldots, E_n$ are events such that $E_i \cap E_j = \emptyset$, $\forall i \neq j$, and such that $\bigcup_{i=1}^{n} E_i = F$, then we say that events $E_1, E_2, \ldots, E_n$* **partition** *set $F$.*

## 2.2 Probability Defined on Events

Given a sample space $\Omega$, we can talk about the probability of event $E$, written $\mathbf{P}\{E\}$. The probability of event $E$ is the probability that the outcome of the experiment lies in the set $E$.

Probability on events is defined via the Probability Axioms:

**Axiom 2.4 (The Three Probability Axioms)**
***Non-negativity:*** $\mathbf{P}\{E\} \geq 0$, *for any event $E$.*
***Additivity:*** *If $E_1, E_2, E_3, \ldots$ is a countable sequence of events, with $E_i \cap E_j = \emptyset$, $\forall i \neq j$, then*

$$\mathbf{P}\{E_1 \cup E_2 \cup E_3 \cup \cdots\} = \mathbf{P}\{E_1\} + \mathbf{P}\{E_2\} + \mathbf{P}\{E_3\} + \cdots .$$

***Normalization:*** $\mathbf{P}\{\Omega\} = 1$.

From the three Probability Axioms, it is easy to reason that if we roll a die, where each side is equally likely, then, by symmetry, $\mathbf{P}\{\text{roll is } 3\} = \frac{1}{6}$. Likewise, $\mathbf{P}\{\text{roll is } \leq 3\} = \mathbf{P}\{\text{roll is 1 or 2 or } 3\} = \frac{3}{6}$.

**Question:** Is something missing from these Axioms? What if $E \cap F \neq \emptyset$?

**Answer:** The case where events $E$ and $F$ overlap can be derived from the Additivity Axiom, as shown in Lemma 2.5.

**Lemma 2.5**

$$\mathbf{P}\{E \cup F\} = \mathbf{P}\{E\} + \mathbf{P}\{F\} - \mathbf{P}\{E \cap F\}.$$

Lemma 2.5 is illustrated in Figure 2.2, where events $E$ and $F$ are depicted as sets. The subtraction of $\mathbf{P}\{E \cap F\}$ term is necessary so that those sample points in the intersection are not counted twice.

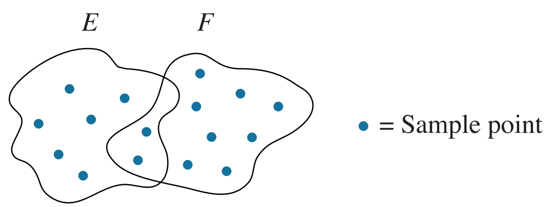


**Figure 2.2** *Venn diagram.*

**Proof**: We can express the set $E \cup F$ as a union of two mutually exclusive sets:

$$E \cup F = E \cup (F \setminus (E \cap F)),$$

where $F \setminus (E \cap F)$ denotes the points that are in $F$ but are *not* in $E \cap F$. Then, by the Additivity Axiom, we have:

$$\mathbf{P}\{E \cup F\} = \mathbf{P}\{E\} + \mathbf{P}\{F \setminus (E \cap F)\}. \tag{2.1}$$

Also by the Additivity Axiom we have:

$$\mathbf{P}\{F\} = \mathbf{P}\{F \setminus (E \cap F)\} + \mathbf{P}\{E \cap F\}. \tag{2.2}$$

We can rewrite (2.2) as:

$$\mathbf{P}\{F \setminus (E \cap F)\} = \mathbf{P}\{F\} - \mathbf{P}\{E \cap F\}. \tag{2.3}$$

Substituting (2.3) into (2.1), we get:

$$\mathbf{P}\{E \cup F\} = \mathbf{P}\{E\} + \mathbf{P}\{F\} - \mathbf{P}\{E \cap F\}. \qquad \blacksquare$$

**Lemma 2.6 (Union bound)** $\mathbf{P}\{E \cup F\} \leq \mathbf{P}\{E\} + \mathbf{P}\{F\}$.

**Proof**: This follows immediately from Lemma 2.5. ■

**Question:** When is Lemma 2.6 an equality?

**Answer:** When $E$ and $F$ are mutually exclusive.

**Question:** Suppose your experiment involves throwing a dart, which is *equally likely* to land anywhere in the interval $[0, 1]$. What is the probability that the dart lands at exactly 0.3?

**Answer:** The probability of landing at exactly 0.3 is defined to be 0. To see why, suppose that the probability were some $\epsilon > 0$. Then the probability of landing at 0.5 would also be $\epsilon$, as would the probability of landing at any rational point. But these different outcomes are mutually exclusive events, so their probabilities add. Thus, the probability of landing in $[0, 1]$ would be greater than 1, which contradicts $\mathbf{P}\{\Omega\} = 1$. While the probability of landing at exactly 0.3 is 0, the probability of landing in the interval $[0, 0.3]$ is defined to be 0.3.

## 2.3 Conditional Probabilities on Events

**Definition 2.7** *The* **conditional probability** *of event $E$ given event $F$ is written as $\mathbf{P}\{E \mid F\}$ and is given by the following, where we assume $\mathbf{P}\{F\} > 0$:*

$$\mathbf{P}\{E \mid F\} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}}. \tag{2.4}$$

$\mathbf{P}\{E \mid F\}$ should be thought of as the probability that event $E$ occurs, given that we have narrowed our sample space to points in $F$.

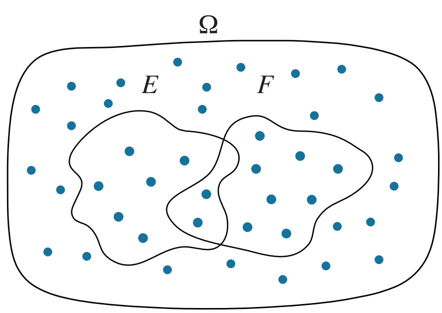


**Figure 2.3** *Sample space with 42 sample points, all equally likely.*

To visualize $\mathbf{P}\{E \mid F\}$, consider Figure 2.3, where $\mathbf{P}\{E\} = \frac{8}{42}$ and $\mathbf{P}\{F\} = \frac{10}{42}$. If we imagine that we narrow our space to the 10 points in $F$, then the probability that the outcome of the experiment is in set $E$, given that the outcome is in set

$F$, should be 2 out of 10. Indeed,

$$\mathbf{P}\{E \mid F\} = \frac{2}{10} = \frac{\frac{2}{42}}{\frac{10}{42}} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}}.$$

**Example 2.8 (Sandwich choices)**

Table 2.1 shows my sandwich choices each day. We define the "first half of the week" to be Monday through Wednesday (inclusive), and the "second half of the week" to be Thursday through Sunday (inclusive).

| Mon | Tue | Wed | Thu | Fri | Sat | Sun |
|---|---|---|---|---|---|---|
| Jelly | Cheese | Turkey | Cheese | Turkey | Cheese | None |

**Table 2.1** *My sandwich choices.*

**Question:** What is $\mathbf{P}\{\text{Cheese} \mid \text{Second half of week}\}$?

**Answer:** We want the fraction of days in the second half of the week when I eat a cheese sandwich. The answer is clearly 2 out of 4. Alternatively, via (2.4):

$$\mathbf{P}\{\text{Cheese} \mid \text{Second half of week}\} = \frac{\mathbf{P}\{\text{Cheese \& Second half}\}}{\mathbf{P}\{\text{Second half}\}} = \frac{\frac{2}{7}}{\frac{4}{7}} = \frac{2}{4}.$$

**Example 2.9 (Two offspring)**

The offspring of a horse is called a foal. A horse couple has at most one foal at a time. Each foal is equally likely to be a "colt" or a "filly." We are told that a horse couple has two foals, and at least one of these is a colt. Given this information, what's the probability that both foals are colts?

**Question:** What is $\mathbf{P}\{\text{both are colts} \mid \text{at least one is a colt}\}$?

**Answer:**

$$\begin{aligned} &\mathbf{P}\{\text{both are colts} \mid \text{at least one is a colt}\} \\ &= \frac{\mathbf{P}\{\text{both are colts } \textit{and} \text{ at least one is a colt}\}}{\mathbf{P}\{\text{at least one is a colt}\}} \\ &= \frac{\mathbf{P}\{\text{both are colts}\}}{\mathbf{P}\{\text{at least one is a colt}\}} \\ &= \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}. \end{aligned}$$

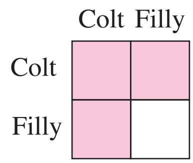


**Figure 2.4** *In Example 2.9, we've conditioned on being in the shaded region.*

**Question:** How might the question read if you wanted the answer to be $\frac{1}{2}$?

**Answer:** The question would ask what is $\mathbf{P}\{\text{both are colts} \mid \text{first born is a colt}\}$.

**Question:** Consider again the example of the couple with two colts, but where we're given the additional information that 10% of horse couples only produce colts, 10% of horse couples only produce fillies, and 80% are equally likely to produce either gender. Does this change your answer to $\mathbf{P}\{\text{both are colts} \mid \text{at least one is a colt}\}$?

**Answer:** Yes! See Exercise 2.11(b).

We now look at generalizing the notion of conditioning. By Definition 2.7, if $E_1$ and $E_2$ are events, where $\mathbf{P}\{E_1 \cap E_2\} > 0$, then

$$\mathbf{P}\{E_1 \cap E_2\} = \mathbf{P}\{E_1\} \cdot \mathbf{P}\{E_2 \mid E_1\} = \mathbf{P}\{E_2\} \cdot \mathbf{P}\{E_1 \mid E_2\}.$$

That is, the probability that the outcome is both in $E_1$ and in $E_2$ can be computed by multiplying two quantities: (1) first restrict the outcome to being in $E_1$ (probability $\mathbf{P}\{E_1\}$); (2) then further restrict the outcome to being in $E_2$, given that we've already restricted it to being in $E_1$ (probability $\mathbf{P}\{E_2 \mid E_1\}$). The next theorem presents a useful "chain rule" for conditioning. This chain rule will be proved in Exercise 2.9.

**Theorem 2.10 (Chain rule for conditioning)** *Let $E_1, E_2, \ldots, E_n$ be events, where $\mathbf{P}\left\{\bigcap_{i=1}^{n} E_i\right\} > 0$. Then*

$$\mathbf{P}\left\{\bigcap_{i=1}^{n} E_i\right\} = \mathbf{P}\{E_1\} \cdot \mathbf{P}\{E_2 \mid E_1\} \cdot \mathbf{P}\{E_3 \mid E_1 \cap E_2\} \cdots \mathbf{P}\left\{E_n \,\middle|\, \bigcap_{i=1}^{n-1} E_i\right\}.$$

## 2.4 Independent Events

**Definition 2.11** *Events $E$ and $F$ are* **independent**, *written $E \perp F$, if*

$$\mathbf{P}\{E \cap F\} = \mathbf{P}\{E\} \cdot \mathbf{P}\{F\}.$$

The above definition might seem less-than-intuitive to you. You might prefer to think of independence using the following definition:

**Definition 2.12 (Alternative)** *Assuming that $\mathbf{P}\{F\} \neq 0$, we say that events $E$ and $F$ are* **independent** *if*

$$\mathbf{P}\{E \mid F\} = \mathbf{P}\{E\}.$$

Definition 2.12 says that $\mathbf{P}\{E\}$ is not affected by whether $F$ is true or not.

**Lemma 2.13** *Definitions 2.11 and 2.12 are equivalent.*

**Proof**:
*Definition 2.11 $\Rightarrow$ Definition 2.12*: Assuming that $\mathbf{P}\{F\} > 0$, we have:

$$\mathbf{P}\{E \mid F\} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{F\}} \overset{\text{by } 2.11}{=} \frac{\mathbf{P}\{E\} \cdot \mathbf{P}\{F\}}{\mathbf{P}\{F\}} = \mathbf{P}\{E\}.$$

*Definition 2.12 $\Rightarrow$ Definition 2.11*:

$$\mathbf{P}\{E \cap F\} = \mathbf{P}\{F\} \cdot \mathbf{P}\{E \mid F\} \overset{\text{by } 2.12}{=} \mathbf{P}\{F\} \cdot \mathbf{P}\{E\}. \qquad \blacksquare$$

Generally people prefer Definition 2.11 because it doesn't require that $\mathbf{P}\{F\} > 0$ and because it shows clearly that a null event is independent of every event.

**Question:** Can two mutually exclusive (non-null) events ever be independent?

**Answer:** No. If $E$ and $F$ are mutually exclusive, then $\mathbf{P}\{E \mid F\} = 0 \neq \mathbf{P}\{E\}$.

**Question:** Suppose one is rolling a die twice. Which of these pairs of events are independent?

(a) $E_1$ = "First roll is 6" and $E_2$ = "Second roll is 6"
(b) $E_1$ = "Sum of the rolls is 7" and $E_2$ = "Second roll is 4"

**Answer:** They are both independent!

**Question:** Suppose we had defined: $E_1$ = "Sum of the rolls is 8" and $E_2$ = "Second roll is 4." Are they independent now?

**Answer:** No.

**Example 2.14 (The unreliable network)**

Suppose you are routing a packet from the source node to the destination node, as shown in Figure 2.5. On the plus side, there are 8 possible paths on which the packet can be routed. On the minus side, each of the 16 edges in the network independently only works with probability $p$. What is the probability that you are able to route the packet from the source to the destination?

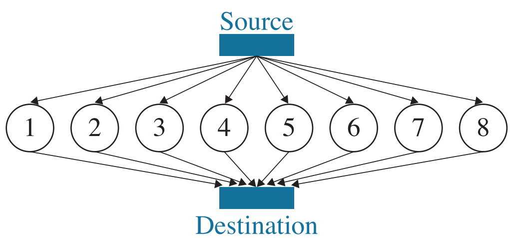


**Figure 2.5** *Unreliable network. Each edge only works with probability p.*

We want to figure out the probability that at least one path is working. We will first demonstrate an intuitive, but wrong, solution.

*Solution 1 (WRONG!)*:

There are eight possible two-hop paths to get from source to destination.

Let $E_1$ denote the event that the first two-hop path works, $E_2$ denote the event that the second two-hop path works, and $E_i$ denote the event that the $i$th two-hop path works:

$$\mathbf{P}\{E_i\} = p^2, \ \forall i.$$

Now the probability that at least one path works is the union of these eight events, namely:

$$\begin{aligned}\mathbf{P}\{\text{At least one path works}\} &= \mathbf{P}\{E_1 \cup E_2 \cup \cdots \cup E_8\}\\ &= \mathbf{P}\{E_1\} + \mathbf{P}\{E_2\} + \cdots + \mathbf{P}\{E_8\}\\ &= 8p^2.\end{aligned}$$

**Question:** What is wrong with Solution 1?

**Answer:** We cannot say that the probability of the union of the events equals the sum of their probabilities, unless the events are mutually exclusive. However, we know that the $E_i$'s are independent, and hence they *cannot* be mutually exclusive.

**Question:** How does the answer given in Solution 1 compare to the correct answer? Higher? Lower?

**Answer:** The answer given in Solution 1 is an upper bound on the correct answer, via the Union Bound in Lemma 2.6.

There's a lesson to be learned from Solution 1. When dealing with the probability of a *union* of independent events, it helps to turn the problem into an *intersection of independent events*. We will illustrate this idea in Solution 2.

*Solution 2 (CORRECT!)*:

$$\begin{aligned}\mathbf{P}\{\text{At least one path works}\} &= \mathbf{P}\{E_1 \cup E_2 \cup \cdots \cup E_8\} \\ &= 1 - \mathbf{P}\{\text{All paths are broken}\} \\ &= 1 - \mathbf{P}\left\{\overline{E_1} \cap \overline{E_2} \cap \cdots \cap \overline{E_8}\right\} \\ &= 1 - \mathbf{P}\left\{\overline{E_1}\right\} \cdot \mathbf{P}\left\{\overline{E_2}\right\} \cdots \mathbf{P}\left\{\overline{E_8}\right\}.\end{aligned}$$

$$\mathbf{P}\left\{\overline{E_1}\right\} = \mathbf{P}\{\text{path 1 is broken}\} = 1 - \mathbf{P}\{\text{path 1 works}\} = 1 - p^2.$$

Thus,

$$\mathbf{P}\{\text{At least one path works}\} = 1 - \left(1 - p^2\right)^8.$$

**Question:** Suppose we have three events: $A$, $B$, and $C$. Given that

$$\mathbf{P}\{A \cap B \cap C\} = \mathbf{P}\{A\} \cdot \mathbf{P}\{B\} \cdot \mathbf{P}\{C\}, \tag{2.5}$$

can we conclude that $A$, $B$, and $C$ are independent?

**Answer:** No. The problem is that (2.5) does not ensure that any *pair* of events are independent, as required by Definition 2.15.

**Definition 2.15** *Events* $A_1, A_2, \ldots, A_n$ *are* **independent** *if, for every subset* $S$ *of* $\{1, 2, \ldots, n\}$,

$$\mathbf{P}\left\{\bigcap_{i \in S} A_i\right\} = \prod_{i \in S} \mathbf{P}\{A_i\}.$$

A weaker version of independence is called *pairwise independence* and is defined in Definition 2.16.

**Definition 2.16** *Events $A_1, A_2, \ldots, A_n$ are* **pairwise independent** *if every pair of events is independent, i.e.,*

$$\forall i \neq j, \qquad \mathbf{P}\left\{A_i \cap A_j\right\} = \mathbf{P}\left\{A_i\right\} \cdot \mathbf{P}\left\{A_j\right\}.$$

Although pairwise independence is weaker than full independence, it still admits some nice properties, as we'll see in Exercise 5.38.

A different notion of independence that comes up frequently in problems (see for example, Exercise 2.14) is that of conditional independence.

**Definition 2.17** *Two events $E$ and $F$ are said to be* **conditionally independent** *given event $G$, where $\mathbf{P}\left\{G\right\} > 0$, if*

$$\mathbf{P}\left\{E \cap F \mid G\right\} = \mathbf{P}\left\{E \mid G\right\} \cdot \mathbf{P}\left\{F \mid G\right\}.$$

Independence does not imply conditional independence and vice-versa, see Exercise 2.19.

## 2.5 Law of Total Probability

Observe that the set $E$ can be expressed as

$$E = (E \cap F) \cup \left(E \cap \overline{F}\right).$$

That is, $E$ is the union of the set $E \cap F$ and the set $E \cap \overline{F}$, because any point in $E$ is also either in $F$ or not in $F$.

Now observe that $E \cap F$ and $E \cap \overline{F}$ are mutually exclusive. Thus,

$$\begin{aligned} \mathbf{P}\left\{E\right\} &= \mathbf{P}\left\{E \cap F\right\} + \mathbf{P}\left\{E \cap \overline{F}\right\} \\ &= \mathbf{P}\left\{E \mid F\right\} \mathbf{P}\left\{F\right\} + \mathbf{P}\left\{E \mid \overline{F}\right\} \mathbf{P}\left\{\overline{F}\right\}, \end{aligned}$$

where $\mathbf{P}\left\{\overline{F}\right\} = 1 - \mathbf{P}\left\{F\right\}$.

Theorem 2.18 is a generalization of this idea:

**Theorem 2.18 (Law of Total Probability)** *Let $F_1, F_2, \ldots, F_n$ partition the state space $\Omega$. Then,*

$$\begin{aligned}\mathbf{P}\{E\} &= \sum_{i=1}^{n} \mathbf{P}\{E \cap F_i\} \\ &= \sum_{i=1}^{n} \mathbf{P}\{E \mid F_i\} \cdot \mathbf{P}\{F_i\}.\end{aligned}$$

*Remark: This also holds if $F_1, F_2, \ldots, F_n$ partition $E$.*
*This also extends to the case where there are countably infinite partitions.*

**Proof**:

$$E = \bigcup_{i=1}^{n} (E \cap F_i).$$

Now, because the events $E \cap F_i$, $i = 1, \ldots, n$ are mutually exclusive, we have that

$$\mathbf{P}\{E\} = \sum_{i=1}^{n} \mathbf{P}\{E \cap F_i\}$$

$$= \sum_{i=1}^{n} \mathbf{P}\{E|F_i\} \cdot \mathbf{P}\{F_i\}. \qquad \blacksquare$$

**Question:** Suppose we are interested in the probability of a transaction failure. We know that if there is a caching failure, that will lead to transaction failures with probability $5/6$. We also know that if there is a network failure then that will lead to a transaction failure with probability $1/4$. Suppose that a caching failure occurs with probability $1/100$ and a network failure occurs with probability $1/100$. What is the probability of a transaction failure?

**Answer:** It is tempting to write (WRONGLY):

$$\begin{aligned}\mathbf{P}\{\text{transaction fails}\} = \mathbf{P}\{\text{transaction fails} \mid \text{caching failure}\} \cdot \frac{1}{100} \\ + \mathbf{P}\{\text{transaction fails} \mid \text{network failure}\} \cdot \frac{1}{100} \\ = \frac{5}{6} \cdot \frac{1}{100} + \frac{1}{4} \cdot \frac{1}{100}.\end{aligned}$$

**Question:** What is wrong with that solution?

**Answer:** The two events that we conditioned on – a network failure and a caching failure – do not partition the space. The sum of the probabilities of these events is clearly $< 1$. Furthermore, there may be a non-zero probability that *both* a network failure and a caching failure occur.

One needs to be very careful that the events that we condition on are (1) mutually exclusive and (2) sum to the whole space under consideration.

We can generalize the Law of Total Probability to apply to a conditional probability, as in Theorem 2.19.

**Theorem 2.19 (Law of Total Probability for conditional probability)** *Let $F_1, F_2, \ldots, F_n$ partition the sample space $\Omega$. Then:*

$$\mathbf{P}\{A \mid B\} = \sum_{i=1}^{n} \mathbf{P}\{A \mid B \cap F_i\} \cdot \mathbf{P}\{F_i \mid B\}.$$

**Proof**:

$$\begin{aligned}
\mathbf{P}\{A \mid B\} &= \frac{\mathbf{P}\{A \cap B\}}{\mathbf{P}\{B\}} \\
&= \frac{\sum_i \mathbf{P}\{A \cap B \cap F_i\}}{\mathbf{P}\{B\}} \\
&= \frac{\sum_i \mathbf{P}\{B\} \cdot \mathbf{P}\{F_i \mid B\} \cdot \mathbf{P}\{A \mid B \cap F_i\}}{\mathbf{P}\{B\}} \qquad \text{(chain rule)} \\
&= \sum_i \mathbf{P}\{F_i \mid B\} \cdot \mathbf{P}\{A \mid B \cap F_i\}.
\end{aligned}$$

■

## 2.6 Bayes' Law

Sometimes, one needs to know $\mathbf{P}\{F \mid E\}$, but all one knows is the reverse direction: $\mathbf{P}\{E \mid F\}$. Is it possible to get $\mathbf{P}\{F \mid E\}$ from $\mathbf{P}\{E \mid F\}$? It turns out that it is possible, assuming that we also know $\mathbf{P}\{E\}$ and $\mathbf{P}\{F\}$.

**Theorem 2.20 (Bayes' Law)** *Assuming $\mathbf{P}\{E\} > 0$,*

$$\mathbf{P}\{F \mid E\} = \frac{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\}}{\mathbf{P}\{E\}}.$$

**Proof**:

$$\mathbf{P}\{F \mid E\} = \frac{\mathbf{P}\{E \cap F\}}{\mathbf{P}\{E\}} = \frac{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\}}{\mathbf{P}\{E\}}. \qquad \blacksquare$$

The Law of Total Probability can be combined with Bayes' Law as follows: Let $F_1, F_2, \ldots, F_n$ partition $\Omega$. Then we can write: $\mathbf{P}\{E\} = \sum_{j=1}^{n} \mathbf{P}\{E \mid F_j\} \cdot \mathbf{P}\{F_j\}$. This yields:

**Theorem 2.21 (Extended Bayes' Law)** *Let $F_1, F_2, \ldots, F_n$ partition $\Omega$. Assuming $\mathbf{P}\{E\} > 0$,*

$$\mathbf{P}\{F \mid E\} = \frac{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\}}{\mathbf{P}\{E\}} = \frac{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\}}{\sum_{j=1}^{n} \mathbf{P}\{E \mid F_j\} \mathbf{P}\{F_j\}}.$$

**Example 2.22 (Cancer screening)**

Suppose that there is a rare child cancer that occurs in one out of one million kids. There's a test for this cancer, which is 99.9% effective (see Figure 2.6).

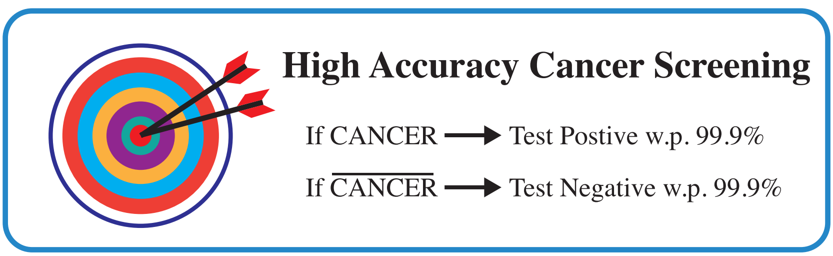


**Figure 2.6** *High accuracy cancer screening.*

**Question:** Suppose that my child's test result is positive. How worried should I be?

**Answer:**

$$\begin{aligned}
&\mathbf{P}\{\text{Cancer} \mid \text{Test pos.}\} \\
&= \frac{\mathbf{P}\{\text{Test pos.} \mid \text{Cancer}\} \cdot \mathbf{P}\{\text{Cancer}\}}{\mathbf{P}\{\text{Test pos.} \mid \text{Cancer}\} \cdot \mathbf{P}\{\text{Cancer}\} + \mathbf{P}\{\text{Test pos.} \mid \text{No Cancer}\} \cdot \mathbf{P}\{\text{No Cancer}\}} \\
&= \frac{0.999 \cdot 10^{-6}}{0.999 \cdot 10^{-6} + 10^{-3} \cdot (1 - 10^{-6})} \\
&\approx \frac{10^{-6}}{10^{-6} + 10^{-3}} \\
&= \frac{1}{1001}.
\end{aligned}$$

Thus, the probability that the child has the cancer is less than 1 in 1000.

**Question:** What was the key factor in obtaining the result?

**Answer:** There are two things going on here. First, the cancer is very rare: $10^{-6}$ likelihood. Second, there is a very low probability of error in the test: $10^{-3}$ chance of error. The *key* determining factor in the chance that the child has cancer is the *ratio* of these two. Consider the ratio of the rareness of the cancer, $10^{-6}$, to the low error probability of the test, $10^{-3}$. This ratio yields $10^{-3}$, which is (roughly) the probability that the child has the cancer. If the cancer were even rarer, say $10^{-7}$ likelihood, then the probability that the child has cancer would be approximately the ratio $\frac{10^{-7}}{10^{-3}} = 10^{-4}$.

## 2.7 Exercises

2.1 **Bombs and alarms**
A bomb detector alarm lights up with probability 0.99 if a bomb is present. If no bomb is present, the bomb alarm still (incorrectly) lights up with probability 0.05. Suppose that a bomb is present with probability 0.1. What is the probability that there is no bomb and the alarm lights up?

2.2 **More on independent events**
Suppose that we roll a die twice. Consider the following three events:

$$E_1 = \text{Second roll is 4}$$
$$E_2 = \text{Difference between the two rolls is 4}$$
$$E_3 = \text{Difference between the two rolls is 3}$$

(a) Are $E_1$ and $E_2$ independent?
(b) Are $E_1$ and $E_3$ independent?

2.3 **How much do vaccines help?**
The US Surgeon General recently declared that 99.5% of COVID deaths are among the unvaccinated [40]. Given that I'm vaccinated, what are my chances of dying from COVID? You may use any of the facts below:
- The fraction of people who are currently vaccinated in the United States is 50%.
- The fraction of people who die from COVID in the United States is 0.2%.

2.4 **Bayesian reasoning for weather prediction**
In the hope of having a dry outdoor wedding, John and Mary decide to get married in the desert, where the average number of rainy days per year is

10. Unfortunately, the weather forecaster is predicting rain for tomorrow, the day of John and Mary's wedding. Suppose that the weather forecaster is not perfectly accurate: If it rains the next day, 90% of the time the forecaster predicts rain. If it is dry the next day, 10% of the time the forecaster still (incorrectly) predicts rain. Given this information, what is the probability that it will rain during John and Mary's wedding?

2.5 **Assessing risk**
Airlines know that on average 5% of the people making flight reservations do not show up. They model this by assuming that each person independently does not show up with probability of 5%. Consequently, their policy is to sell 52 tickets for a flight that can only hold 50 passengers. What is the probability that there will be a seat available for every passenger who shows up?

2.6 **When one event implies another**
Suppose that we are told that event $A$ implies event $B$. Which of the following *must* be true:
(a) $\mathbf{P}\{A\} \leq \mathbf{P}\{B\}$
(b) $\mathbf{P}\{A\} > \mathbf{P}\{B\}$
(c) Neither

2.7 **Wearing masks and COVID**
This problem analyzes mask wearing and COVID via Figure 2.7.
(a) Does consistently wearing masks reduce your chance of catching COVID? Let $C$ (respectively, $M$) denote the event that a randomly chosen person catches COVID (respectively, consistently wears a mask). Compare $\mathbf{P}\{C \mid M\}$ with $\mathbf{P}\left\{C \mid \overline{M}\right\}$.
(b) Your friend comes down with COVID. Based on your answer to part (a), do you think it's more likely that they didn't consistently wear a mask, or that they did? Do the computation to see if you're right.

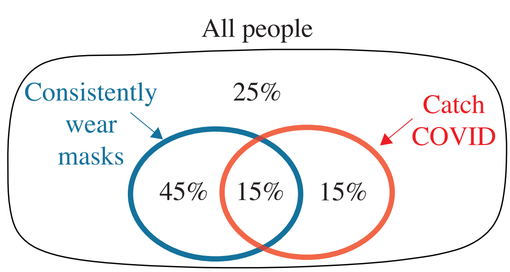


**Figure 2.7** *Venn diagram showing fraction of people who wear masks consistently and fraction of people who catch COVID, for Exercise 2.7.*

2.8 **Positive correlation**
We say that events $A$ and $B$ are *positively correlated* if

$$\mathbf{P}\{A \mid B\} > \mathbf{P}\{A\}. \tag{2.6}$$

Prove or disprove that (2.6) implies

$$\mathbf{P}\{B \mid A\} > \mathbf{P}\{B\}. \tag{2.7}$$

Assume that $\mathbf{P}\{A\} > 0$ and $\mathbf{P}\{B\} > 0$.

2.9 **Chain rule for conditioning**
Let $E_1, E_2, \ldots, E_n$ be $n$ events, where $\mathbf{P}\left\{\bigcap_{i=1}^{n} E_i\right\} > 0$. Prove via induction that

$$\mathbf{P}\left\{\bigcap_{i=1}^{n} E_i\right\} = \mathbf{P}\{E_1\} \cdot \mathbf{P}\{E_2 \mid E_1\} \cdot \mathbf{P}\{E_3 \mid E_1 \cap E_2\} \cdots \mathbf{P}\left\{E_n \,\middle|\, \bigcap_{i=1}^{n-1} E_i\right\}.$$

2.10 **Birthday paradox**
The famous birthday paradox considers the situation of a room of $m = 30$ people, where we ask what is the probability that no two have the same birthday. Let $A$ be the event that no two people have the same birthday. It would seem that $\mathbf{P}\{A\}$ is high, given that there are $n = 365$ possible birthdays, but it turns out that $\mathbf{P}\{A\} < e^{-1}$; hence with high likelihood at least two people have the same birthday.
Assume that all $n$ birthdays are equally likely. Prove the above claim via the following conditioning approach: Imagine that the people in the room are ordered, from 1 to $m$. Let $A_i$ be the event that person $i$ has a different birthday from each of the first $i-1$ people. Now observe that $A = \bigcap_{i=1}^{m} A_i$, and use the chain rule from Exercise 2.9. [Hint: Leave everything in terms of $n$ and $m$ until the final evaluation. You will need to use (1.14), which says that $1 - \frac{i}{n} \le e^{-\frac{i}{n}}$ for high $n$.]

2.11 **It's a colt!**
The offspring of a horse is called a foal. A horse couple has at most one foal at a time. Each foal is equally likely to be a "colt" or a "filly." We are told that a horse couple has two foals, and at least one of these is a colt. Given this information, what's the probability that both foals are colts?
(a) Compute the answer to the above question, assuming only that each foal is equally likely to be a colt or a filly.
(b) Now re-compute the answer given the latest discovery: Scientists have discovered that 10% of horse couples only produce colts, 10% of couples only produce fillies, and 80% are equally likely to produce either gender.
(c) Is your answer for (b) different from that for (a)? Why?

2.12 **It's a Sunday colt!**
As in Exercise 2.11, we are told that a horse couple has two foals. Additionally, we are told that at least one of these foals is a colt that was born on a Sunday. Given this information, what's the probability that both foals are colts? Assume that a foal is equally like to be born on any day of the week, and is equally likely to be a colt or a filly, and births are independent.

2.13 **Happy or sad**
Independently, on any given day, with probability 50% Mor is happy, and with probability 50% she is sad. While it's hard to know how Mor is feeling, her clothes offer a clue. On her happy days, Mor is 90% likely to wear red and 10% likely to wear black. On her sad days, Mor is 90% likely to wear black and 10% likely to wear red. For the last two days, Mor has worn black. What is the likelihood that Mor has been sad both of the last two days?

2.14 **Bayesian reasoning for healthcare testing**
A pharmaceutical company has developed a potential vaccine against the H1N1 flu virus. Before any testing of the vaccine, the developers assume that with probability 0.5 their vaccine will be effective and with probability 0.5 it will be ineffective. The developers do an initial laboratory test on the vaccine. This initial lab test is only partially indicative of the effectiveness of the vaccine, with an accuracy of 0.6. Specifically, if the vaccine is effective, then this laboratory test will return "success" with probability 0.6, whereas if the vaccine is ineffective, then this laboratory test will return "failure" with probability 0.6.

(a) What is the probability that the laboratory test returns "success"?
(b) What is the probability that the vaccine is effective, given that the laboratory test returned "success"?
(c) The developers decide to add a second experiment (this one on human beings) that is more indicative than the original lab test and has an accuracy of 0.8. Specifically, if the vaccine is effective, then the human being test will return "success" with probability 0.8. If the vaccine is ineffective, then the human being test will return "failure" with probability 0.8. What is the probability that the vaccine is effective, given that both the lab test and the human being test came up "success"? How useful was it to add this additional test? Assume that the two tests (human test and lab test) are conditionally independent on the vaccine being effective or ineffective.

2.15 **Independence of three events**
Natassa suggests the following definition for the independence of three

events: Events $A$, $B$, and $C$ are independent if

$$\mathbf{P}\{A \cap B \cap C\} = \mathbf{P}\{A\} \cdot \mathbf{P}\{B\} \cdot \mathbf{P}\{C\}.$$

Is Natassa correct? Specifically, does the above definition also ensure that any *pair* of events are independent? Either provide a proof or a counter-example. Assume that your events each have non-zero probability.

2.16 **Does independence imply independence of the complement?**
Haotian reasons that if event $E$ is independent of event $G$, then $E$ should also be independent of $\overline{G}$. He argues that if $E$ is not affected by whether $G$ is true, then it should also not be affected by whether $G$ is not true. Either provide a formal proof via the definition of independence, or find a counter-example where you define $E$, $G$, and $\overline{G}$.

2.17 **Corrupted packets**
CMU has two campuses: one in Pittsburgh and one in Qatar. Suppose all packets of a flow originate in either Pittsburgh or in Qatar. Packets originating in Pittsburgh are (independently) corrupted with probability $p$. Packets originating in Qatar are (independently) corrupted with probability $q$. We are watching a flow of packets (all from the same origin). At first, we don't know the origin, so we assume that each origin is equally likely. So far, we've seen two packets in the flow, both of which were corrupted. Given this information:
(a) What is the probability that the flow originated in Pittsburgh?
(b) What is the probability that the next packet will be corrupted?

2.18 **Pairwise independence and the mystery novel principle**
The mystery novel principle considers three events, $A$, $B$, and $C$, where:
- $A$ tells us nothing about $C$;
- $B$ tells us nothing about $C$;
- But $A$ and $B$ together tell us everything about $C$!

Another way of phrasing this is that $A$, $B$, and $C$ are "pairwise independent," meaning that any pair of these is independent. However, the three events together are not independent. Provide a simple example of three events with this property. [Hint: You shouldn't need more than two tosses of a coin.]

2.19 **Independence does not imply conditional independence**
Produce an example of two events, $A$ and $B$, that are independent, but are no longer independent once we condition on some event $C$. [Hint: Your example can be very simple. Consider, for instance, the simple experiment of flipping a coin two times, and define events based on that experiment.]

2.20 **Does conditional independence imply conditional independence on the complement?**

Jelena reasons that if events $E$ and $F$ are conditionally independent of event $G$, then they should also be conditionally independent of event $\overline{G}$. Either provide a formal proof, or find a counter-example.

2.21 **Another definition of conditional independence?**
Recall that events $E$ and $F$ are conditionally independent on event $G$ if

$$\mathbf{P}\{E \cap F \mid G\} = \mathbf{P}\{E \mid G\} \cdot \mathbf{P}\{F \mid G\}.$$

Taegyun proposes an alternative definition: events $E$ and $F$ are conditionally independent on event $G$ if

$$\mathbf{P}\{E \mid F \cap G\} = \mathbf{P}\{E \mid G\}.$$

Taegyun argues that "knowing $F$ gives no additional information about $E$, given that we already know $G$." Is Taegyun's definition equivalent to the original definition (i.e., each definition implies the other) or not? If so, prove it. If not, find a counter-example. Assume that $\mathbf{P}\{F \cap G\} > 0$.

2.22 **The famous Monty Hall problem**
A game show host brings the contestant into a room with three closed doors. Behind one of the doors is a car. Behind the other two doors is a goat. The contestant is asked to pick a door and state which door she has chosen (we'll assume she picks the door at random, because she has no insider knowledge).
Now the game show host, knowing what's behind each door, picks a door that was not chosen by the contestant and reveals that there is a goat behind that door (the game show host will always choose to open a door with a goat). The contestant is then asked, "Would you like to switch from your chosen door?"
One would think that it shouldn't matter whether the contestant switches to the other unopened door, since the car is equally likely to be behind the originally chosen door and the remaining unopened door. This intuition is wrong. Derive the probability that the contestant gets the car, both in the case that the contestant switches doors and the case that the contestant sticks with her original door.
[Hint: Assume WLOG that the contestant picks door 1. Let $D$ be a r.v. that represents the door hiding the car. Let $Y$ be a r.v. which represents the action of the game show host. Compare $\mathbf{P}\{D = 1 \mid Y = i\}$ with $\mathbf{P}\{D = 2 \mid Y = i\}$ for all logical values of $i$.]

2.23 **Weighty coins**
Imagine that there are two coins of weight $w_1$ and eight coins of weight $w_2$, where $w_1 \neq w_2$. All the coins look identical. We pick two pairs of coins, without replacement, from the pile of 10 coins. What is the probability that

all the chosen coins have weight $w_2$, given that the weights of the two pairs are equal?

2.24 **Winning streak**
Assume that the Pittsburgh Steelers win a game with probability $p$ irrespective of the opponent, and that the outcome of each game is independent of the others.

(a) Suppose you are told that the Steelers won four out of the eight games they played in a season. What is the probability that Steelers had a winning streak (i.e., continuous wins) of at least three matches in that season?

(b) Suppose $p = 0.5$. Suppose the Steelers play six games in a particular season. What is the probability that the Steelers will have a winning streak of at least three matches? (Is the hype that a winning streak receives by the media worth it?)

2.25 **Monty Hall with five doors**
A game show host brings the contestant into a room with five closed doors. Behind one of the doors is a car. Behind the other four doors is a goat. The contestant is asked to pick a door and state which door she has chosen (we'll assume she picks the door at random, because she has no insider knowledge).
Now the game show host, knowing what's behind each door, picks a door that was not chosen by the contestant and reveals that there is a goat behind that door (the game show host will always choose to open a door with a goat). The contestant is then asked, "Would you like to switch from your chosen door?"
Derive the probability that the contestant gets the car, both in the case that the contestant switches doors and the case that the contestant sticks with her original door.

2.26 **Another fun door problem**
Imagine there are two doors. Both doors have money behind them, but one contains twice as much money as the other. Suppose you choose one door randomly, and before you look behind the door you are given the chance to switch doors. Should you switch?

(a) Explain what is wrong with the following argument that favors switching:
Suppose $M$ is the money behind the door I chose. Then with probability $\frac{1}{2}$, I chose the door with less money and the other door contains $2M$. Also with probability $\frac{1}{2}$, I chose the door with more money and the other door contains $\frac{M}{2}$. Therefore the expected value of money in the

other door is

$$\frac{1}{2} \cdot 2M + \frac{1}{2} \cdot \frac{M}{2} = \frac{5}{4}M > M.$$

So we should switch.

(b) Prove that there is no point to switching.

2.27 **Prediction with an unknown source**

You have two dice. Die A is a fair die (each of the six numbers are equally likely) and die B is a biased die (the number six comes up with probability $\frac{2}{3}$ and the remaining $\frac{1}{3}$ probability is split evenly across all the other numbers). Kaige picks a die at random and rolls that die three times. Given that the first two rolls are both sixes, what is the probability that the third roll will also be a six?

2.28 **Modeling packet corruption**

Packet switched networks are the backbone of the Internet. Here, data is transferred from a source to a destination by encapsulating and transferring data as a series of packets. There are a number of reasons due to which packets get lost in the network and never reach the destination. Consider two models for packet losses in the network.

*Model 1*: Each packet is lost with probability $p$ independently.

*Model 2*: A packet is lost with probability $p_1$ if its previous packet was transmitted successfully, and is lost with probability $p_2$ if its previous packet was lost.

Suppose a source sends exactly three packets over the network to a destination. For this setup, under Model 2, assume that the probability of the first packet getting lost is $p_1$. Further assume $p = p_1 = 0.01$ and $p_2 = 0.5$.

(a) What is the probability that the second packet is lost under Model 1 and Model 2?

(b) Suppose you are told that the third packet is lost. Given this additional information, what is the probability that the second packet is lost under Model 1 and Model 2?

(c) Suppose we represent the loss pattern for the three packets using 0s and 1s, where 0 represents the packet being lost and 1 represents the packet being transferred successfully. For example, loss pattern 110 corresponds to the scenario when the first two packets are transferred successfully and the third packet is lost. What is the probability of loss patterns $\{010, 100, 001\}$ under Model 1 and Model 2?

(d) What do you observe from your answer to the above question? Specifically, what kind of loss patterns have higher probability in Model 2 as compared to Model 1?

*Aside:* Extensive measurements over the Internet have shown that packet losses in real-world networks are correlated. Models similar to Model 2 are used to model such correlated packet-loss scenarios. For example, one such model is called the Gilbert–Elliot model [23, 32].