1
Probability

1.1 Introduction

Probability is a mathematical language for quantifying uncertainty. In this Chapter we introduce the basic concepts underlying probability theory. We begin with the sample space, which is the set of possible outcomes.

1.2 Sample Spaces and Events

The **sample space** Ω is the set of possible outcomes of an experiment. Points ω in Ω are called **sample outcomes**, **realizations**, or **elements**. Subsets of Ω are called **Events**.

1.1 Example. If we toss a coin twice then $\Omega = \{HH, HT, TH, TT\}$. The event that the first toss is heads is $A = \{HH, HT\}$. ■

1.2 Example. Let ω be the outcome of a measurement of some physical quantity, for example, temperature. Then $\Omega = \mathbb{R} = (-\infty, \infty)$. One could argue that taking $\Omega = \mathbb{R}$ is not accurate since temperature has a lower bound. But there is usually no harm in taking the sample space to be larger than needed. The event that the measurement is larger than 10 but less than or equal to 23 is $A = (10, 23]$. ■

L. Wasserman, *All of Statistics*
© Springer Science+Business Media New York 2004

1.3 Example. If we toss a coin forever, then the sample space is the infinite set

$$\Omega = \Big\{\omega = (\omega_1, \omega_2, \omega_3, \ldots,) : \ \omega_i \in \{H, T\}\Big\}.$$

Let E be the event that the first head appears on the third toss. Then

$$E = \Big\{(\omega_1, \omega_2, \omega_3, \ldots,) : \ \omega_1 = T, \omega_2 = T, \omega_3 = H, \ \omega_i \in \{H, T\} \text{ for } i > 3\Big\}. \ \blacksquare$$

Given an event A, let $A^c = \{\omega \in \Omega : \ \omega \notin A\}$ denote the complement of A. Informally, A^c can be read as "not A." The complement of Ω is the empty set $\emptyset$. The union of events A and B is defined

$$A \bigcup B = \{\omega \in \Omega : \ \omega \in A \text{ or } \omega \in B \text{ or } \omega \in \text{both}\}$$

which can be thought of as "A or B." If $A_1, A_2, \ldots$ is a sequence of sets then

$$\bigcup_{i=1}^{\infty} A_i = \Big\{\omega \in \Omega : \ \omega \in A_i \text{ for at least one i}\Big\}.$$

The intersection of A and B is

$$A \bigcap B = \{\omega \in \Omega : \ \omega \in A \text{ and } \omega \in B\}$$

read "A and B." Sometimes we write $A \bigcap B$ as AB or (A, B). If $A_1, A_2, \ldots$ is a sequence of sets then

$$\bigcap_{i=1}^{\infty} A_i = \Big\{\omega \in \Omega : \ \omega \in A_i \text{ for all i}\Big\}.$$

The set difference is defined by $A - B = \{\omega : \ \omega \in A, \omega \notin B\}$. If every element of A is also contained in B we write $A \subset B$ or, equivalently, $B \supset A$. If A is a finite set, let $|A|$ denote the number of elements in A. See the following table for a summary.

Summary of Terminology

Ω	sample space
ω	outcome (point or element)
A	event (subset of Ω)
A^c	complement of A (not A)
$A \bigcup B$	union (A or B)
$A \bigcap B$ or AB	intersection (A and B)
$A - B$	set difference (ω in A but not in B)
$A \subset B$	set inclusion
$\emptyset$	null event (always false)
Ω	true event (always true)

We say that $A_1, A_2, \ldots$ are **disjoint** or are **mutually exclusive** if $A_i \bigcap A_j = \emptyset$ whenever $i \neq j$. For example, $A_1 = [0,1), A_2 = [1,2), A_3 = [2,3), \ldots$ are disjoint. A **partition** of Ω is a sequence of disjoint sets $A_1, A_2, \ldots$ such that $\bigcup_{i=1}^{\infty} A_i = \Omega$. Given an event A, define the **indicator function of** A by

$$I_A(\omega) = I(\omega \in A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

A sequence of sets $A_1, A_2, \ldots$ is **monotone increasing** if $A_1 \subset A_2 \subset \cdots$ and we define $\lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$. A sequence of sets $A_1, A_2, \ldots$ is **monotone decreasing** if $A_1 \supset A_2 \supset \cdots$ and then we define $\lim_{n\to\infty} A_n = \bigcap_{i=1}^{\infty} A_i$. In either case, we will write $A_n \to A$.

1.4 Example. Let $\Omega = \mathbb{R}$ and let $A_i = [0, 1/i)$ for $i = 1, 2, \ldots$. Then $\bigcup_{i=1}^{\infty} A_i = [0,1)$ and $\bigcap_{i=1}^{\infty} A_i = \{0\}$. If instead we define $A_i = (0, 1/i)$ then $\bigcup_{i=1}^{\infty} A_i = (0,1)$ and $\bigcap_{i=1}^{\infty} A_i = \emptyset$. ■

1.3 Probability

We will assign a real number $\mathbb{P}(A)$ to every event A, called the **probability** of A. [1] We also call $\mathbb{P}$ a **probability distribution** or a **probability measure.** To qualify as a probability, $\mathbb{P}$ must satisfy three axioms:

1.5 Definition. *A function* $\mathbb{P}$ *that assigns a real number* $\mathbb{P}(A)$ *to each event* A *is a* **probability distribution** *or a* **probability measure** *if it satisfies the following three axioms:*
Axiom 1: $\mathbb{P}(A) \geq 0$ *for every* A
Axiom 2: $\mathbb{P}(\Omega) = 1$
Axiom 3: *If* $A_1, A_2, \ldots$ *are disjoint then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

[1] It is not always possible to assign a probability to every event A if the sample space is large, such as the whole real line. Instead, we assign probabilities to a limited class of set called a σ-field. See the appendix for details.

There are many interpretations of $\mathbb{P}(A)$. The two common interpretations are frequencies and degrees of beliefs. In the frequency interpretation, $\mathbb{P}(A)$ is the long run proportion of times that A is true in repetitions. For example, if we say that the probability of heads is 1/2, we mean that if we flip the coin many times then the proportion of times we get heads tends to 1/2 as the number of tosses increases. An infinitely long, unpredictable sequence of tosses whose limiting proportion tends to a constant is an idealization, much like the idea of a straight line in geometry. The degree-of-belief interpretation is that $\mathbb{P}(A)$ measures an observer's strength of belief that A is true. In either interpretation, we require that Axioms 1 to 3 hold. The difference in interpretation will not matter much until we deal with statistical inference. There, the differing interpretations lead to two schools of inference: the frequentist and the Bayesian schools. We defer discussion until Chapter 11.

One can derive many properties of $\mathbb{P}$ from the axioms, such as:

$$
\begin{array}{rcl}
\mathbb{P}(\emptyset) & = & 0 \\
A \subset B & \Longrightarrow & \mathbb{P}(A) \le \mathbb{P}(B) \\
0 \le & \mathbb{P}(A) & \le 1 \\
\mathbb{P}(A^c) & = & 1 - \mathbb{P}(A) \\
A \bigcap B = \emptyset & \Longrightarrow & \mathbb{P}\left(A \bigcup B\right) = \mathbb{P}(A) + \mathbb{P}(B).
\end{array}
\tag{1.1}
$$

A less obvious property is given in the following Lemma.

1.6 Lemma. *For any events A and B,*

$$
\mathbb{P}\left(A \bigcup B\right) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB).
$$

PROOF. Write $A \bigcup B = (AB^c) \bigcup (AB) \bigcup (A^c B)$ and note that these events are disjoint. Hence, making repeated use of the fact that $\mathbb{P}$ is additive for disjoint events, we see that

$$
\begin{aligned}
\mathbb{P}\left(A \bigcup B\right) &= \mathbb{P}\left((AB^c) \bigcup (AB) \bigcup (A^c B)\right) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^c B) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^c B) + \mathbb{P}(AB) - \mathbb{P}(AB) \\
&= P\left((AB^c) \bigcup (AB)\right) + \mathbb{P}\left((A^c B) \bigcup (AB)\right) - \mathbb{P}(AB) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB). \quad \blacksquare
\end{aligned}
$$

1.7 Example. Two coin tosses. Let H_1 be the event that heads occurs on toss 1 and let H_2 be the event that heads occurs on toss 2. If all outcomes are

equally likely, then $\mathbb{P}(H_1 \bigcup H_2) = \mathbb{P}(H_1)+\mathbb{P}(H_2)-\mathbb{P}(H_1H_2) = \frac{1}{2}+\frac{1}{2}-\frac{1}{4} = 3/4$.
■

1.8 Theorem (Continuity of Probabilities). *If $A_n \to A$ then*

$$\mathbb{P}(A_n) \to \mathbb{P}(A)$$

as $n \to \infty$.

PROOF. Suppose that A_n is monotone increasing so that $A_1 \subset A_2 \subset \cdots$. Let $A = \lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$. Define $B_1 = A_1$, $B_2 = \{\omega \in \Omega : \omega \in A_2, \omega \notin A_1\}$, $B_3 = \{\omega \in \Omega : \omega \in A_3, \omega \notin A_2, \omega \notin A_1\}, \ldots$ It can be shown that $B_1, B_2, \ldots$ are disjoint, $A_n = \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$ for each n and $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$. (See exercise 1.) From Axiom 3,

$$\mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} \mathbb{P}(B_i)$$

and hence, using Axiom 3 again,

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}(A). \quad ■$$

1.4 Probability on Finite Sample Spaces

Suppose that the sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$ is finite. For example, if we toss a die twice, then Ω has 36 elements: $\Omega = \{(i,j);\ i,j \in \{1,\ldots 6\}\}$. If each outcome is equally likely, then $\mathbb{P}(A) = |A|/36$ where $|A|$ denotes the number of elements in A. The probability that the sum of the dice is 11 is 2/36 since there are two outcomes that correspond to this event.

If Ω is finite and if each outcome is equally likely, then

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

which is called the **uniform probability distribution.** To compute probabilities, we need to count the number of points in an event A. Methods for counting points are called combinatorial methods. We needn't delve into these in any great detail. We will, however, need a few facts from counting theory that will be useful later. Given n objects, the number of ways of ordering these objects is $n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$. For convenience, we define $0! = 1$. We also define

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \tag{1.2}$$

read "n choose k", which is the number of distinct ways of choosing k objects from n. For example, if we have a class of 20 people and we want to select a committee of 3 students, then there are

$$\binom{20}{3} = \frac{20!}{3!17!} = \frac{20\times 19\times 18}{3\times 2\times 1} = 1140$$

possible committees. We note the following properties:

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \text{and} \quad \binom{n}{k} = \binom{n}{n-k}.$$

1.5 Independent Events

If we flip a fair coin twice, then the probability of two heads is $\frac{1}{2}\times\frac{1}{2}$. We multiply the probabilities because we regard the two tosses as independent. The formal definition of independence is as follows:

1.9 Definition. *Two events A and B are* **independent** *if*

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B) \tag{1.3}$$

and we write $A \amalg B$. A set of events $\{A_i : i \in I\}$ is independent if

$$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J}\mathbb{P}(A_i)$$

for every finite subset J of I. If A and B are not independent, we write

$$A \not\amalg B$$

Independence can arise in two distinct ways. Sometimes, we explicitly **assume** that two events are independent. For example, in tossing a coin twice, we usually assume the tosses are independent which reflects the fact that the coin has no memory of the first toss. In other instances, we **derive** independence by verifying that $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$ holds. For example, in tossing a fair die, let $A = \{2,4,6\}$ and let $B = \{1,2,3,4\}$. Then, $A\bigcap B = \{2,4\}$,

$\mathbb{P}(AB) = 2/6 = \mathbb{P}(A)\mathbb{P}(B) = (1/2) \times (2/3)$ and so A and B are independent. In this case, we didn't assume that A and B are independent — it just turned out that they were.

Suppose that A and B are disjoint events, each with positive probability. Can they be independent? No. This follows since $\mathbb{P}(A)\mathbb{P}(B) > 0$ yet $\mathbb{P}(AB) = \mathbb{P}(\emptyset) = 0$. Except in this special case, there is no way to judge independence by looking at the sets in a Venn diagram.

1.10 Example. Toss a fair coin 10 times. Let $A =$"at least one head." Let T_j be the event that tails occurs on the j^{th} toss. Then

$$\begin{aligned} \mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\ &= 1 - \mathbb{P}(\text{all tails}) \\ &= 1 - \mathbb{P}(T_1 T_2 \cdots T_{10}) \\ &= 1 - \mathbb{P}(T_1)\mathbb{P}(T_2)\cdots\mathbb{P}(T_{10}) \quad \text{using independence} \\ &= 1 - \left(\frac{1}{2}\right)^{10} \approx .999. \quad \blacksquare \end{aligned}$$

1.11 Example. Two people take turns trying to sink a basketball into a net. Person 1 succeeds with probability 1/3 while person 2 succeeds with probability 1/4. What is the probability that person 1 succeeds before person 2? Let E denote the event of interest. Let A_j be the event that the first success is by person 1 and that it occurs on trial number j. Note that $A_1, A_2, \ldots$ are disjoint and that $E = \bigcup_{j=1}^{\infty} A_j$. Hence,

$$\mathbb{P}(E) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Now, $\mathbb{P}(A_1) = 1/3$. A_2 occurs if we have the sequence person 1 misses, person 2 misses, person 1 succeeds. This has probability $\mathbb{P}(A_2) = (2/3)(3/4)(1/3) = (1/2)(1/3)$. Following this logic we see that $\mathbb{P}(A_j) = (1/2)^{j-1}(1/3)$. Hence,

$$\mathbb{P}(E) = \sum_{j=1}^{\infty} \frac{1}{3}\left(\frac{1}{2}\right)^{j-1} = \frac{1}{3}\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^{j-1} = \frac{2}{3}.$$

Here we used that fact that, if $0 < r < 1$ then $\sum_{j=k}^{\infty} r^j = r^k/(1-r)$. $\blacksquare$

Summary of Independence

1. A and B are independent if and only if $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.
2. Independence is sometimes assumed and sometimes derived.
3. Disjoint events with positive probability are not independent.

1.6 Conditional Probability

Assuming that $\mathbb{P}(B) > 0$, we define the conditional probability of A given that B has occurred as follows:

1.12 Definition. *If* $\mathbb{P}(B) > 0$ *then the* **conditional probability** *of* A *given* B *is*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}. \tag{1.4}$$

Think of $\mathbb{P}(A|B)$ as the fraction of times A occurs among those in which B occurs. For any fixed B such that $\mathbb{P}(B) > 0$, $\mathbb{P}(\cdot|B)$ is a probability (i.e., it satisfies the three axioms of probability). In particular, $\mathbb{P}(A|B) \geq 0$, $\mathbb{P}(\Omega|B) = 1$ and if $A_1, A_2, \ldots$ are disjoint then $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$. But it is in general **not** true that $\mathbb{P}(A|B \bigcup C) = \mathbb{P}(A|B) + \mathbb{P}(A|C)$. The rules of probability apply to events on the left of the bar. In general it is **not** the case that $\mathbb{P}(A|B) = \mathbb{P}(B|A)$. People get this confused all the time. For example, the probability of spots given you have measles is 1 but the probability that you have measles given that you have spots is not 1. In this case, the difference between $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$ is obvious but there are cases where it is less obvious. This mistake is made often enough in legal cases that it is sometimes called the prosecutor's fallacy.

1.13 Example. A medical test for a disease D has outcomes $+$ and $-$. The probabilities are:

	D	D^c
$+$	.009	.099
$-$	.001	.891

From the definition of conditional probability,

$$\mathbb{P}(+|D) = \frac{\mathbb{P}(+\bigcap D)}{\mathbb{P}(D)} = \frac{.009}{.009 + .001} = .9$$

and

$$\mathbb{P}(-|D^c) = \frac{\mathbb{P}(-\bigcap D^c)}{\mathbb{P}(D^c)} = \frac{.891}{.891 + .099} \approx .9.$$

Apparently, the test is fairly accurate. Sick people yield a positive 90 percent of the time and healthy people yield a negative about 90 percent of the time. Suppose you go for a test and get a positive. What is the probability you have the disease? Most people answer .90. The correct answer is

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(+\bigcap D)}{\mathbb{P}(+)} = \frac{.009}{.009 + .099} \approx .08.$$

The lesson here is that you need to compute the answer numerically. Don't trust your intuition. ■

The results in the next lemma follow directly from the definition of conditional probability.

1.14 Lemma. *If A and B are independent events then $\mathbb{P}(A|B) = \mathbb{P}(A)$. Also, for any pair of events A and B,*

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

From the last lemma, we see that another interpretation of independence is that knowing B doesn't change the probability of A. The formula $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A)$ is sometimes helpful for calculating probabilities.

1.15 Example. Draw two cards from a deck, without replacement. Let A be the event that the first draw is the Ace of Clubs and let B be the event that the second draw is the Queen of Diamonds. Then $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A) = (1/52) \times (1/51)$. ■

Summary of Conditional Probability

1. If $\mathbb{P}(B) > 0$, then

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

2. $\mathbb{P}(\cdot|B)$ satisfies the axioms of probability, for fixed B. In general, $\mathbb{P}(A|\cdot)$ does not satisfy the axioms of probability, for fixed A.

3. In general, $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$.

4. A and B are independent if and only if $\mathbb{P}(A|B) = \mathbb{P}(A)$.

1.7 Bayes' Theorem

Bayes' theorem is the basis of "expert systems" and "Bayes' nets," which are discussed in Chapter 17. First, we need a preliminary result.

1.16 Theorem (The Law of Total Probability). *Let $A_1, \ldots, A_k$ be a partition of Ω. Then, for any event B,*

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

PROOF. Define $C_j = BA_j$ and note that $C_1, \ldots, C_k$ are disjoint and that $B = \bigcup_{j=1}^{k} C_j$. Hence,

$$\mathbb{P}(B) = \sum_j \mathbb{P}(C_j) = \sum_j \mathbb{P}(BA_j) = \sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)$$

since $\mathbb{P}(BA_j) = \mathbb{P}(B|A_j)\mathbb{P}(A_j)$ from the definition of conditional probability. ■

1.17 Theorem (Bayes' Theorem). *Let $A_1, \ldots, A_k$ be a partition of Ω such that $\mathbb{P}(A_i) > 0$ for each i. If $\mathbb{P}(B) > 0$ then, for each $i = 1, \ldots, k$,*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}. \tag{1.5}$$

1.18 Remark. We call $\mathbb{P}(A_i)$ the **prior probability of** A and $\mathbb{P}(A_i|B)$ the **posterior probability of** A.

PROOF. We apply the definition of conditional probability twice, followed by the law of total probability:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_iB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}. \blacksquare$$

1.19 Example. I divide my email into three categories: A_1 = "spam," A_2 = "low priority" and A_3 = "high priority." From previous experience I find that

$\mathbb{P}(A_1) = .7$, $\mathbb{P}(A_2) = .2$ and $\mathbb{P}(A_3) = .1$. Of course, $.7 + .2 + .1 = 1$. Let B be the event that the email contains the word "free." From previous experience, $\mathbb{P}(B|A_1) = .9$, $\mathbb{P}(B|A_2) = .01$, $\mathbb{P}(B|A_1) = .01$. (Note: $.9 + .01 + .01 \neq 1$.) I receive an email with the word "free." What is the probability that it is spam? Bayes' theorem yields,

$$\mathbb{P}(A_1|B) = \frac{.9 \times .7}{(.9 \times .7) + (.01 \times .2) + (.01 \times .1)} = .995. \quad \blacksquare$$

1.8 Bibliographic Remarks

The material in this chapter is standard. Details can be found in any number of books. At the introductory level, there is DeGroot and Schervish (2002); at the intermediate level, Grimmett and Stirzaker (1982) and Karr (1993); at the advanced level there are Billingsley (1979) and Breiman (1992). I adapted many examples and exercises from DeGroot and Schervish (2002) and Grimmett and Stirzaker (1982).

1.9 Appendix

Generally, it is not feasible to assign probabilities to all subsets of a sample space Ω. Instead, one restricts attention to a set of events called a **σ-algebra** or a **σ-field** which is a class $\mathcal{A}$ that satisfies:

(i) $\emptyset \in \mathcal{A}$,

(ii) if $A_1, A_2, \ldots, \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ and

(iii) $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$.

The sets in $\mathcal{A}$ are said to be **measurable**. We call $(\Omega, \mathcal{A})$ a **measurable space.** If $\mathbb{P}$ is a probability measure defined on $\mathcal{A}$, then $(\Omega, \mathcal{A}, \mathbb{P})$ is called a **probability space.** When Ω is the real line, we take $\mathcal{A}$ to be the smallest σ-field that contains all the open subsets, which is called the **Borel σ-field.**

1.10 Exercises

1. Fill in the details of the proof of Theorem 1.8. Also, prove the monotone decreasing case.

2. Prove the statements in equation (1.1).

3. Let Ω be a sample space and let $A_1, A_2, \ldots,$ be events. Define $B_n = \bigcup_{i=n}^{\infty} A_i$ and $C_n = \bigcap_{i=n}^{\infty} A_i$.

   (a) Show that $B_1 \supset B_2 \supset \cdots$ and that $C_1 \subset C_2 \subset \cdots$.

   (b) Show that $\omega \in \bigcap_{n=1}^{\infty} B_n$ if and only if ω belongs to an infinite number of the events $A_1, A_2, \ldots$.

   (c) Show that $\omega \in \bigcup_{n=1}^{\infty} C_n$ if and only if ω belongs to all the events $A_1, A_2, \ldots$ except possibly a finite number of those events.

4. Let $\{A_i : i \in I\}$ be a collection of events where I is an arbitrary index set. Show that

$$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c \quad \text{and} \quad \left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c$$

   Hint: First prove this for $I = \{1, \ldots, n\}$.

5. Suppose we toss a fair coin until we get exactly two heads. Describe the sample space S. What is the probability that exactly k tosses are required?

6. Let $\Omega = \{0, 1, \ldots,\}$. Prove that there does not exist a uniform distribution on Ω (i.e., if $\mathbb{P}(A) = \mathbb{P}(B)$ whenever $|A| = |B|$, then $\mathbb{P}$ cannot satisfy the axioms of probability).

7. Let $A_1, A_2, \ldots$ be events. Show that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

   Hint: Define $B_n = A_n - \bigcup_{i=1}^{n-1} A_i$. Then show that the B_n are disjoint and that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$.

8. Suppose that $\mathbb{P}(A_i) = 1$ for each i. Prove that

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 1.$$

9. For fixed B such that $\mathbb{P}(B) > 0$, show that $\mathbb{P}(\cdot|B)$ satisfies the axioms of probability.

10. You have probably heard it before. Now you can solve it rigorously. It is called the "Monty Hall Problem." A prize is placed at random

behind one of three doors. You pick a door. To be concrete, let's suppose you always pick door 1. Now Monty Hall chooses one of the other two doors, opens it and shows you that it is empty. He then gives you the opportunity to keep your door or switch to the other unopened door. Should you stay or switch? Intuition suggests it doesn't matter. The correct answer is that you should switch. Prove it. It will help to specify the sample space and the relevant events carefully. Thus write $\Omega = \{(\omega_1, \omega_2) : \ \omega_i \in \{1, 2, 3\}\}$ where ω_1 is where the prize is and ω_2 is the door Monty opens.

11. Suppose that A and B are independent events. Show that A^c and B^c are independent events.

12. There are three cards. The first is green on both sides, the second is red on both sides and the third is green on one side and red on the other. We choose a card at random and we see one side (also chosen at random). If the side we see is green, what is the probability that the other side is also green? Many people intuitively answer 1/2. Show that the correct answer is 2/3.

13. Suppose that a fair coin is tossed repeatedly until both a head and tail have appeared at least once.

    (a) Describe the sample space Ω.

    (b) What is the probability that three tosses will be required?

14. Show that if $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$ then A is independent of every other event. Show that if A is independent of itself then $\mathbb{P}(A)$ is either 0 or 1.

15. The probability that a child has blue eyes is 1/4. Assume independence between children. Consider a family with 3 children.

    (a) If it is known that at least one child has blue eyes, what is the probability that at least two children have blue eyes?

    (b) If it is known that the youngest child has blue eyes, what is the probability that at least two children have blue eyes?

16. Prove Lemma 1.14.

17. Show that

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

18. Suppose k events form a partition of the sample space Ω, i.e., they are disjoint and $\bigcup_{i=1}^k A_i = \Omega$. Assume that $\mathbb{P}(B) > 0$. Prove that if $\mathbb{P}(A_1|B) < \mathbb{P}(A_1)$ then $\mathbb{P}(A_i|B) > \mathbb{P}(A_i)$ for some $i = 2, \ldots, k$.

19. Suppose that 30 percent of computer owners use a Macintosh, 50 percent use Windows, and 20 percent use Linux. Suppose that 65 percent of the Mac users have succumbed to a computer virus, 82 percent of the Windows users get the virus, and 50 percent of the Linux users get the virus. We select a person at random and learn that her system was infected with the virus. What is the probability that she is a Windows user?

20. A box contains 5 coins and each has a different probability of showing heads. Let $p_1, \ldots, p_5$ denote the probability of heads on each coin. Suppose that

$$p_1 = 0, \ p_2 = 1/4, \ p_3 = 1/2, \ p_4 = 3/4 \ \text{ and } \ p_5 = 1.$$

Let H denote "heads is obtained" and let C_i denote the event that coin i is selected.

(a) Select a coin at random and toss it. Suppose a head is obtained. What is the posterior probability that coin i was selected ($i = 1, \ldots, 5$)? In other words, find $\mathbb{P}(C_i|H)$ for $i = 1, \ldots, 5$.

(b) Toss the coin again. What is the probability of another head? In other words find $\mathbb{P}(H_2|H_1)$ where $H_j =$ "heads on toss j."

Now suppose that the experiment was carried out as follows: We select a coin at random and toss it until a head is obtained.

(c) Find $\mathbb{P}(C_i|B_4)$ where $B_4 =$ "first head is obtained on toss 4."

21. (Computer Experiment.) Suppose a coin has probability p of falling heads up. If we flip the coin many times, we would expect the proportion of heads to be near p. We will make this formal later. Take $p = .3$ and $n = 1,000$ and simulate n coin flips. Plot the proportion of heads as a function of n. Repeat for $p = .03$.

22. (Computer Experiment.) Suppose we flip a coin n times and let p denote the probability of heads. Let X be the number of heads. We call X a binomial random variable, which is discussed in the next chapter. Intuition suggests that X will be close to $n\,p$. To see if this is true, we can repeat this experiment many times and average the X values. Carry

out a simulation and compare the average of the X's to np. Try this for $p = .3$ and $n = 10$, $n = 100$, and $n = 1,000$.

23. (Computer Experiment.) Here we will get some experience simulating conditional probabilities. Consider tossing a fair die. Let $A = \{2, 4, 6\}$ and $B = \{1, 2, 3, 4\}$. Then, $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 2/3$ and $\mathbb{P}(AB) = 1/3$. Since $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$, the events A and B are independent. Simulate draws from the sample space and verify that $\widehat{\mathbb{P}}(AB) = \widehat{\mathbb{P}}(A)\widehat{\mathbb{P}}(B)$ where $\widehat{\mathbb{P}}(A)$ is the proportion of times A occurred in the simulation and similarly for $\widehat{\mathbb{P}}(AB)$ and $\widehat{\mathbb{P}}(B)$. Now find two events A and B that are not independent. Compute $\widehat{\mathbb{P}}(A), \widehat{\mathbb{P}}(B)$ and $\widehat{\mathbb{P}}(AB)$. Compare the calculated values to their theoretical values. Report your results and interpret.