1
Introduction to probability

Why do electrical and computer engineers need to study probability?

Probability theory provides powerful tools to explain, model, analyze, and design technology developed by electrical and computer engineers. Here are a few applications.

Signal processing. My own interest in the subject arose when I was an undergraduate taking the required course in probability for electrical engineers. We considered the situation shown in Figure 1.1. To determine the presence of an aircraft, a known radar pulse $v(t)$

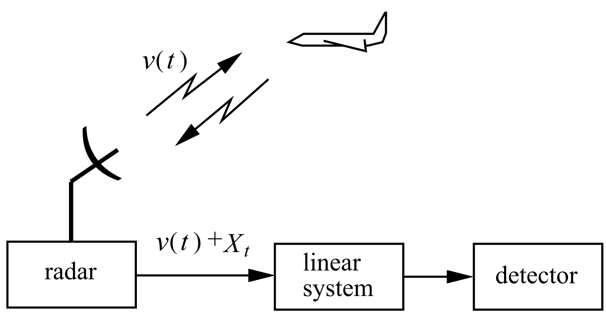

***Figure* 1.1.** Block diagram of radar detection system.

is sent out. If there are no objects in range of the radar, the radar's amplifiers produce only a noise waveform, denoted by X_t. If there is an object in range, the reflected radar pulse plus noise is produced. The overall goal is to decide whether the received waveform is noise only or signal plus noise. To get an idea of how difficult this can be, consider the signal plus noise waveform shown at the top in Figure 1.2. Our class addressed the subproblem of designing an optimal linear system to process the received waveform so as to make the presence of the signal more obvious. We learned that the optimal transfer function is given by the matched filter. If the signal at the top in Figure 1.2 is processed by the appropriate matched filter, we get the output shown at the bottom in Figure 1.2. You will study the matched filter in Chapter 10.

Computer memories. Suppose you are designing a computer memory to hold k-bit words. To increase system reliability, you employ an error-correcting-code system. With this system, instead of storing just the k data bits, you store an additional l bits (which are functions of the data bits). When reading back the $(k+l)$-bit word, if at least m bits are read out correctly, then all k data bits can be recovered (the value of m depends on the code). To characterize the quality of the computer memory, we compute the probability that at least m bits are correctly read back. You will be able to do this after you study the binomial random variable in Chapter 3.

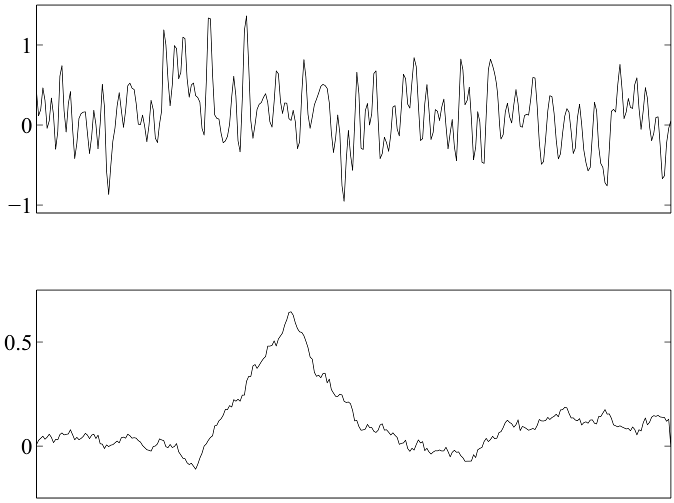

Figure **1.2.** Matched filter input (top) in which the signal is hidden by noise. Matched filter output (bottom) in which the signal presence is obvious.

Optical communication systems. Optical communication systems use photodetectors (see Figure 1.3) to interface between optical and electronic subsystems. When these sys-

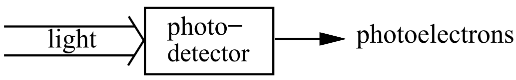

Figure **1.3.** Block diagram of a photodetector. The rate at which photoelectrons are produced is proportional to the intensity of the light.

tems are at the limits of their operating capabilities, the number of photoelectrons produced by the photodetector is well-modeled by the Poisson[a] random variable you will study in Chapter 2 (see also the Poisson process in Chapter 11). In deciding whether a transmitted bit is a zero or a one, the receiver counts the number of photoelectrons and compares it to a threshold. System performance is determined by computing the probability that the threshold is exceeded.

Wireless communication systems. In order to enhance weak signals and maximize the range of communication systems, it is necessary to use amplifiers. Unfortunately, amplifiers always generate thermal noise, which is added to the desired signal. As a consequence of the underlying physics, the noise is Gaussian. Hence, the Gaussian density function, which you will meet in Chapter 4, plays a prominent role in the analysis and design of communication systems. When noncoherent receivers are used, e.g., noncoherent frequency shift keying,

[a]Many important quantities in probability and statistics are named after famous mathematicians and statisticians. You can use an Internet search engine to find pictures and biographies of them on the web. At the time of this writing, numerous biographies of famous mathematicians and statisticians can be found at `http://turnbull.mcs.st-and.ac.uk/history/BiogIndex.html` and at `http://www.york.ac.uk/depts/maths/histstat/people/welcome.htm`. Pictures on stamps and currency can be found at `http://jeff560.tripod.com/`.

this naturally leads to the Rayleigh, chi-squared, noncentral chi-squared, and Rice density functions that you will meet in the problems in Chapters 4, 5, 7, and 9.

Variability in electronic circuits. Although circuit manufacturing processes attempt to ensure that all items have nominal parameter values, there is always some variation among items. How can we estimate the average values in a batch of items without testing all of them? How good is our estimate? You will learn how to do this in Chapter 6 when you study parameter estimation and confidence intervals. Incidentally, the same concepts apply to the prediction of presidential elections by surveying only a few voters.

Computer network traffic. Prior to the 1990s, network analysis and design was carried out using long-established Markovian models [41, p. 1]. You will study Markov chains in Chapter 12. As self similarity was observed in the traffic of local-area networks [35], wide-area networks [43], and in World Wide Web traffic [13], a great research effort began to examine the impact of self similarity on network analysis and design. This research has yielded some surprising insights into questions about buffer size vs. bandwidth, multiple-time-scale congestion control, connection duration prediction, and other issues [41, pp. 9–11]. In Chapter 15 you will be introduced to self similarity and related concepts.

In spite of the foregoing applications, probability was not originally developed to handle problems in electrical and computer engineering. The first applications of probability were to questions about gambling posed to Pascal in 1654 by the Chevalier de Mere. Later, probability theory was applied to the determination of life expectancies and life-insurance premiums, the theory of measurement errors, and to statistical mechanics. Today, the theory of probability and statistics is used in many other fields, such as economics, finance, medical treatment and drug studies, manufacturing quality control, public opinion surveys, etc.

Relative frequency

Consider an experiment that can result in M possible outcomes, $O_1,\ldots,O_M$. For example, in tossing a die, one of the six sides will land facing up. We could let O_i denote the outcome that the ith side faces up, $i = 1,\ldots,6$. Alternatively, we might have a computer with six processors, and O_i could denote the outcome that a program or thread is assigned to the ith processor. As another example, there are $M = 52$ possible outcomes if we draw one card from a deck of playing cards. Similarly, there are $M = 52$ outcomes if we ask which week during the next year the stock market will go up the most. The simplest example we consider is the flipping of a coin. In this case there are two possible outcomes, "heads" and "tails." Similarly, there are two outcomes when we ask whether or not a bit was correctly received over a digital communication system. No matter what the experiment, suppose we perform it n times and make a note of how many times each outcome occurred. Each performance of the experiment is called a **trial**.[b] Let $N_n(O_i)$ denote the number of times O_i occurred in n trials. The **relative frequency** of outcome O_i,

$$\frac{N_n(O_i)}{n},$$

is the fraction of times O_i occurred.

[b]When there are only two outcomes, the repeated experiments are called **Bernoulli trials**.

Here are some simple computations using relative frequency. First,

$$N_n(O_1) + \cdots + N_n(O_M) \;=\; n,$$

and so

$$\frac{N_n(O_1)}{n} + \cdots + \frac{N_n(O_M)}{n} \;=\; 1. \tag{1.1}$$

Second, we can group outcomes together. For example, if the experiment is tossing a die, let E denote the event that the outcome of a toss is a face with an even number of dots; i.e., E is the event that the outcome is O_2, O_4, or O_6. If we let $N_n(E)$ denote the number of times E occurred in n tosses, it is easy to see that

$$N_n(E) \;=\; N_n(O_2) + N_n(O_4) + N_n(O_6),$$

and so the relative frequency of E is

$$\frac{N_n(E)}{n} \;=\; \frac{N_n(O_2)}{n} + \frac{N_n(O_4)}{n} + \frac{N_n(O_6)}{n}. \tag{1.2}$$

Practical experience has shown us that as the number of trials n becomes large, the relative frequencies settle down and appear to converge to some limiting value. This behavior is known as **statistical regularity**.

***Example* 1.1.** Suppose we toss a fair coin 100 times and note the relative frequency of heads. Experience tells us that the relative frequency should be about $1/2$. When we did this,[c] we got 0.47 and were not disappointed.

The tossing of a coin 100 times and recording the relative frequency of heads out of 100 tosses can be considered an experiment in itself. Since the number of heads can range from 0 to 100, there are 101 possible outcomes, which we denote by $S_0, \ldots, S_{100}$. In the preceding example, this experiment yielded S_{47}.

***Example* 1.2.** We performed the experiment with outcomes $S_0, \ldots, S_{100}$ 1000 times and counted the number of occurrences of each outcome. All trials produced between 33 and 68 heads. Rather than list $N_{1000}(S_k)$ for the remaining values of k, we summarize as follows:

$$\begin{aligned}
N_{1000}(S_{33}) + N_{1000}(S_{34}) + N_{1000}(S_{35}) &= 4\\
N_{1000}(S_{36}) + N_{1000}(S_{37}) + N_{1000}(S_{38}) &= 6\\
N_{1000}(S_{39}) + N_{1000}(S_{40}) + N_{1000}(S_{41}) &= 32\\
N_{1000}(S_{42}) + N_{1000}(S_{43}) + N_{1000}(S_{44}) &= 98\\
N_{1000}(S_{45}) + N_{1000}(S_{46}) + N_{1000}(S_{47}) &= 165\\
N_{1000}(S_{48}) + N_{1000}(S_{49}) + N_{1000}(S_{50}) &= 230\\
N_{1000}(S_{51}) + N_{1000}(S_{52}) + N_{1000}(S_{53}) &= 214\\
N_{1000}(S_{54}) + N_{1000}(S_{55}) + N_{1000}(S_{56}) &= 144
\end{aligned}$$

[c]We did not actually toss a coin. We used a random number generator to simulate the toss of a fair coin. Simulation is discussed in Chapters 5 and 6.

$$\begin{aligned}
N_{1000}(S_{57}) + N_{1000}(S_{58}) + N_{1000}(S_{59}) &= 76 \\
N_{1000}(S_{60}) + N_{1000}(S_{61}) + N_{1000}(S_{62}) &= 21 \\
N_{1000}(S_{63}) + N_{1000}(S_{64}) + N_{1000}(S_{65}) &= 9 \\
N_{1000}(S_{66}) + N_{1000}(S_{67}) + N_{1000}(S_{68}) &= 1.
\end{aligned}$$

This summary is illustrated in the histogram shown in Figure 1.4. (The bars are centered over values of the form $k/100$; e.g., the bar of height 230 is centered over 0.49.)

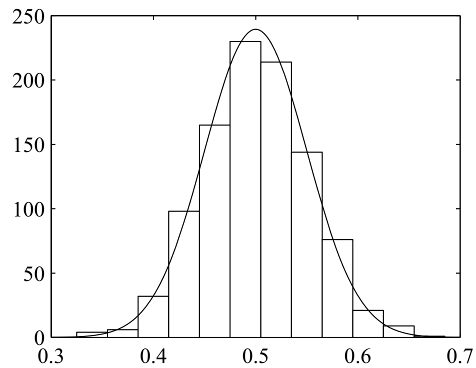

***Figure* 1.4.** Histogram of Example 1.2 with overlay of a Gaussian density.

Below we give an indication of why most of the time the relative frequency of heads is close to one half and why the bell-shaped curve fits so well over the histogram. For now we point out that the foregoing methods allow us to determine the bit-error rate of a digital communication system, whether it is a wireless phone or a cable modem connection. In principle, we simply send a large number of bits over the channel and find out what fraction were received incorrectly. This gives an estimate of the bit-error rate. To see how good an estimate it is, we repeat the procedure many times and make a histogram of our estimates.

What is probability theory?

Axiomatic probability theory, which is the subject of this book, was developed by **A. N. Kolmogorov**[d] in 1933. This theory specifies a set of axioms for a well-defined mathematical model of physical experiments whose outcomes exhibit random variability each time they are performed. The advantage of using a model rather than performing an experiment itself is that it is usually much more efficient in terms of time and money to analyze a mathematical model. This is a sensible approach only if the model correctly predicts the behavior of actual experiments. This is indeed the case for Kolmogorov's theory.

A simple prediction of Kolmogorov's theory arises in the mathematical model for the relative frequency of heads in n tosses of a fair coin that we considered in Example 1.1. In the model of this experiment, the relative frequency converges to $1/2$ as n tends to infinity;

[d]The website `http://kolmogorov.com/` is devoted to Kolmogorov.

this is a special case of the the **strong law of large numbers**, which is derived in Chapter 14. (A related result, known as the **weak law of large numbers**, is derived in Chapter 3.)

Another prediction of Kolmogorov's theory arises in modeling the situation in Example 1.2. The theory explains why the histogram in Figure 1.4 agrees with the bell-shaped curve overlaying it. In the model, the strong law tells us that for each k, the relative frequency of having exactly k heads in 100 tosses should be close to

$$\frac{100!}{k!(100-k)!}\frac{1}{2^{100}}.$$

Then, by the **central limit theorem**, which is derived in Chapter 5, the above expression is approximately equal to (see Example 5.19)

$$\frac{1}{5\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{k-50}{5}\right)^2\right].$$

(You should convince yourself that the graph of e^{-x^2} is indeed a bell-shaped curve.)

Because Kolmogorov's theory makes predictions that agree with physical experiments, it has enjoyed great success in the analysis and design of real-world systems.

1.1 Sample spaces, outcomes, and events

Sample spaces

To model systems that yield uncertain or random measurements, we let Ω denote the set of all possible distinct, indecomposable measurements that could be observed. The set Ω is called the **sample space**. Here are some examples corresponding to the applications discussed at the beginning of the chapter.

Signal processing. In a radar system, the voltage of a noise waveform at time t can be viewed as possibly being any real number. The first step in *modeling* such a noise voltage is to consider the sample space consisting of all real numbers, i.e., $\Omega = (-\infty, \infty)$.

Computer memories. Suppose we store an n-bit word consisting of all 0s at a particular location. When we read it back, we may not get all 0s. In fact, any n-bit word may be read out if the memory location is faulty. The set of all possible n-bit words can be modeled by the sample space

$$\Omega = \{(b_1, \dots, b_n) : b_i = 0 \text{ or } 1\}.$$

Optical communication systems. Since the output of a photodetector is a random number of photoelectrons. The logical sample space here is the nonnegative integers,

$$\Omega = \{0, 1, 2, \dots\}.$$

Notice that we include 0 to account for the possibility that no photoelectrons are observed.

Wireless communication systems. Noncoherent receivers measure the energy of the incoming waveform. Since energy is a nonnegative quantity, we model it with the sample space consisting of the nonnegative real numbers, $\Omega = [0, \infty)$.

Variability in electronic circuits. Consider the lowpass RC filter shown in Figure 1.5(a). Suppose that the exact values of R and C are not perfectly controlled by the manufacturing process, but are known to satisfy

$$95 \text{ ohms} \leq R \leq 105 \text{ ohms} \quad \text{and} \quad 300\ \mu\text{F} \leq C \leq 340\ \mu\text{F}.$$

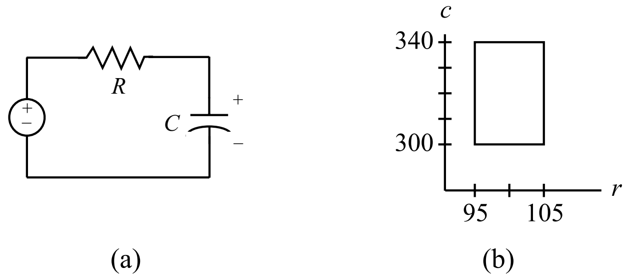

***Figure* 1.5.** (a) Lowpass *RC* filter. (b) Sample space for possible values of *R* and *C*.

This suggests that we use the sample space of ordered pairs of real numbers, (r,c), where $95 \leq r \leq 105$ and $300 \leq c \leq 340$. Symbolically, we write

$$\Omega = \{(r,c) : 95 \leq r \leq 105 \text{ and } 300 \leq c \leq 340\},$$

which is the rectangular region in Figure 1.5(b).

Computer network traffic. If a router has a buffer that can store up to 70 packets, and we want to model the actual number of packets waiting for transmission, we use the sample space

$$\Omega = \{0,1,2,\dots,70\}.$$

Notice that we include 0 to account for the possibility that there are no packets waiting to be sent.

Outcomes and events

Elements or points in the sample space Ω are called **outcomes**. Collections of outcomes are called **events**. In other words, an event is a subset of the sample space. Here are some examples.

If the sample space is the real line, as in modeling a noise voltage, the individual numbers such as 1.5, -8, and π are outcomes. Subsets such as the interval

$$[0,5] = \{v : 0 \leq v \leq 5\}$$

are events. Another event would be $\{2,4,7.13\}$. Notice that singleton sets, that is sets consisting of a single point, are also events; e.g., $\{1.5\}$, $\{-8\}$, $\{\pi\}$. Be sure you understand the difference between the outcome -8 and the event $\{-8\}$, which is the set consisting of the single outcome -8.

If the sample space is the set of all triples (b_1,b_2,b_3), where the b_i are 0 or 1, then any particular triple, say $(0,0,0)$ or $(1,0,1)$ would be an outcome. An event would be a subset such as the set of all triples with exactly one 1; i.e.,

$$\{(0,0,1),(0,1,0),(1,0,0)\}.$$

An example of a singleton event would be $\{(1,0,1)\}$.

In modeling the resistance and capacitance of the RC filter above, we suggested the sample space

$$\Omega = \{(r,c) : 95 \leq r \leq 105 \text{ and } 300 \leq c \leq 340\},$$

which was shown in Figure 1.5(b). If a particular circuit has $R = 101$ ohms and $C = 327$ μF, this would correspond to the outcome $(101,327)$, which is indicated by the dot in Figure 1.6. If we observed a particular circuit with $R \leq 97$ ohms and $C \geq 313$ μF, this would correspond to the event

$$\{(r,c) : 95 \leq r \leq 97 \text{ and } 313 \leq c \leq 340\},$$

which is the shaded region in Figure 1.6.

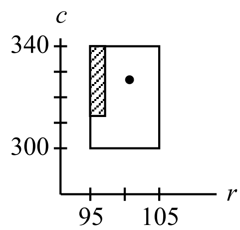

***Figure* 1.6.** The dot is the outcome $(101,327)$. The shaded region is the event $\{(r,c) : 95 \leq r \leq 97 \text{ and } 313 \leq c \leq 340\}$.

1.2 Review of set notation

Since sample spaces and events use the language of sets, we recall in this section some basic definitions, notation, and properties of sets.

Let Ω be a set of points. If ω is a point in Ω, we write $\omega \in \Omega$. Let A and B be two collections of points in Ω. If every point in A also belongs to B, we say that A is a **subset** of B, and we denote this by writing $A \subset B$. If $A \subset B$ and $B \subset A$, then we write $A = B$; i.e., two sets are equal if they contain exactly the same points. If $A \subset B$ but $A \neq B$, we say that A is a **proper subset** of B.

Set relationships can be represented graphically in **Venn diagrams**. In these pictures, the whole space Ω is represented by a rectangular region, and subsets of Ω are represented by disks or oval-shaped regions. For example, in Figure 1.7(a), the disk A is completely contained in the oval-shaped region B, thus depicting the relation $A \subset B$.

Set operations

If $A \subset \Omega$, and $\omega \in \Omega$ does not belong to A, we write $\omega \notin A$. The set of all such ω is called the **complement** of A in Ω; i.e.,

$$A^{\mathrm{c}} := \{\omega \in \Omega : \omega \notin A\}.$$

This is illustrated in Figure 1.7(b), in which the shaded region is the complement of the disk A.

The **empty set** or **null set** is denoted by $\varnothing$; it contains no points of Ω. Note that for any $A \subset \Omega$, $\varnothing \subset A$. Also, $\Omega^{\mathrm{c}} = \varnothing$.

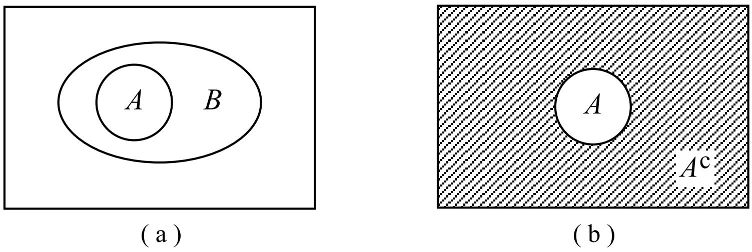

***Figure* 1.7.** (a) Venn diagram of $A \subset B$. (b) The complement of the disk A, denoted by A^c, is the shaded part of the diagram.

The **union** of two subsets A and B is

$$A \cup B := \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}.$$

Here "or" is inclusive; i.e., if $\omega \in A \cup B$, we permit ω to belong either to A or to B or to both. This is illustrated in Figure 1.8(a), in which the shaded region is the union of the disk A and the oval-shaped region B.

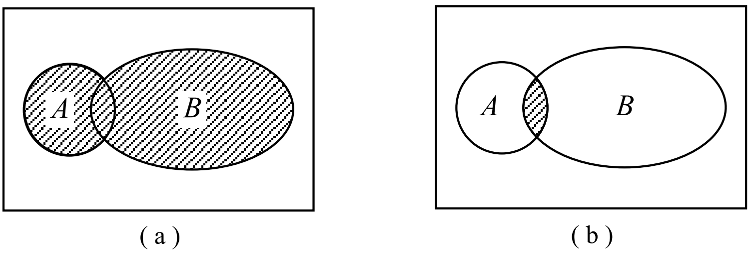

***Figure* 1.8.** (a) The shaded region is $A \cup B$. (b) The shaded region is $A \cap B$.

The **intersection** of two subsets A and B is

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

hence, $\omega \in A \cap B$ if and only if ω belongs to both A and B. This is illustrated in Figure 1.8(b), in which the shaded area is the intersection of the disk A and the oval-shaped region B. The reader should also note the following special case. If $A \subset B$ (recall Figure 1.7(a)), then $A \cap B = A$. In particular, we always have $A \cap \Omega = A$ and $\varnothing \cap B = \varnothing$.

The **set difference** operation is defined by

$$B \setminus A := B \cap A^c,$$

i.e., $B \setminus A$ is the set of $\omega \in B$ that do not belong to A. In Figure 1.9(a), $B \setminus A$ is the shaded part of the oval-shaped region B. Thus, $B \setminus A$ is found by starting with all the points in B and then removing those that belong to A.

Two subsets A and B are **disjoint** or **mutually exclusive** if $A \cap B = \varnothing$; i.e., there is no point in Ω that belongs to both A and B. This condition is depicted in Figure 1.9(b).

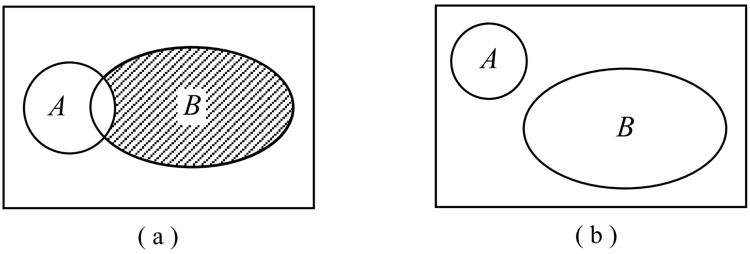

***Figure* 1.9.** (a) The shaded region is $B \setminus A$. (b) Venn diagram of disjoint sets A and B.

***Example* 1.3.** Let $\Omega := \{0,1,2,3,4,5,6,7\}$, and put

$$A := \{1,2,3,4\}, \quad B := \{3,4,5,6\}, \quad \text{and} \quad C := \{5,6\}.$$

Evaluate $A \cup B$, $A \cap B$, $A \cap C$, A^{c}, and $B \setminus A$.

Solution. It is easy to see that $A \cup B = \{1,2,3,4,5,6\}$, $A \cap B = \{3,4\}$, and $A \cap C = \varnothing$. Since $A^{\mathrm{c}} = \{0,5,6,7\}$,

$$B \setminus A = B \cap A^{\mathrm{c}} = \{5,6\} = C.$$

Set identities

Set operations are easily seen to obey the following relations. Some of these relations are analogous to the familiar ones that apply to ordinary numbers if we think of union as the set analog of addition and intersection as the set analog of multiplication. Let A, B, and C be subsets of Ω. The **commutative laws** are

$$A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A. \tag{1.3}$$

The **associative laws** are

$$A \cup (B \cup C) = (A \cup B) \cup C \quad \text{and} \quad A \cap (B \cap C) = (A \cap B) \cap C. \tag{1.4}$$

The **distributive laws** are

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{1.5}$$

and

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{1.6}$$

De Morgan's laws are

$$(A \cap B)^{\mathrm{c}} = A^{\mathrm{c}} \cup B^{\mathrm{c}} \quad \text{and} \quad (A \cup B)^{\mathrm{c}} = A^{\mathrm{c}} \cap B^{\mathrm{c}}. \tag{1.7}$$

Formulas (1.3)–(1.5) are exactly analogous to their numerical counterparts. Formulas (1.6) and (1.7) do not have numerical counterparts. We also recall that $A \cap \Omega = A$ and $\varnothing \cap B = \varnothing$; hence, we can think of Ω as the analog of the number one and $\varnothing$ as the analog of the number zero. Another analog is the formula $A \cup \varnothing = A$.