Probability and Random Processes for Electrical and Computer Engineers

John A. Gubner
University of Wisconsin–Madison

Discrete Random Variables

Bernoulli(p)

$\wp(X=1)=p, \quad \wp(X=0)=1-p.$
$\mathsf{E}[X]=p, \quad \mathsf{var}(X)=p(1-p), \quad G_X(z)=(1-p)+pz.$

---

binomial(n, p)

$$\wp(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,\ldots,n.$$

$\mathsf{E}[X]=np, \quad \mathsf{var}(X)=np(1-p), \quad G_X(z)=[(1-p)+pz]^n.$

---

geometric$_0$(p)

$\wp(X=k)=(1-p)p^k, \quad k=0,1,2,\ldots.$

$$\mathsf{E}[X]=\frac{p}{1-p}, \quad \mathsf{var}(X)=\frac{p}{(1-p)^2}, \quad G_X(z)=\frac{1-p}{1-pz}.$$

---

geometric$_1$(p)

$\wp(X=k)=(1-p)p^{k-1}, \quad k=1,2,3,\ldots.$

$$\mathsf{E}[X]=\frac{1}{1-p}, \quad \mathsf{var}(X)=\frac{p}{(1-p)^2}, \quad G_X(z)=\frac{(1-p)z}{1-pz}.$$

---

negative binomial or Pascal(m, p)

$$\wp(X=k)=\binom{k-1}{m-1}(1-p)^m p^{k-m}, \ k=m,m+1,\ldots.$$

$$\mathsf{E}[X]=\frac{m}{1-p}, \quad \mathsf{var}(X)=\frac{mp}{(1-p)^2}, \quad G_X(z)=\left[\frac{(1-p)z}{1-pz}\right]^m.$$

Note that Pascal$(1,p)$ is the same as geometric$_1(p)$.

Poisson(λ)

$$\wp(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0,1,\ldots.$$

$\mathsf{E}[X]=\lambda, \quad \mathsf{var}(X)=\lambda, \quad G_X(z)=e^{\lambda(z-1)}.$

Fourier Transforms*

Fourier Transform

$$H(f) \;=\; \int_{-\infty}^{\infty} h(t) e^{-j2\pi f t}\, dt$$

Inversion Formula

$$h(t) \;=\; \int_{-\infty}^{\infty} H(f) e^{j2\pi f t}\, df$$

$h(t)$	$H(f)$
$I_{[-T,T]}(t)$	$2T\dfrac{\sin(2\pi T f)}{2\pi T f}$
$2W\dfrac{\sin(2\pi W t)}{2\pi W t}$	$I_{[-W,W]}(f)$
$(1-\lvert t\rvert/T) I_{[-T,T]}(t)$	$T\left[\dfrac{\sin(\pi T f)}{\pi T f}\right]^2$
$W\left[\dfrac{\sin(\pi W t)}{\pi W t}\right]^2$	$(1-\lvert f\rvert/W) I_{[-W,W]}(f)$
$e^{-\lambda t} u(t)$	$\dfrac{1}{\lambda + j2\pi f}$
$e^{-\lambda \lvert t\rvert}$	$\dfrac{2\lambda}{\lambda^2 + (2\pi f)^2}$
$\dfrac{\lambda}{\lambda^2 + t^2}$	$\pi\, e^{-2\pi\lambda \lvert f\rvert}$
$e^{-(t/\sigma)^2/2}$	$\sqrt{2\pi}\,\sigma\, e^{-\sigma^2(2\pi f)^2/2}$

*The indicator function $I_{[a,b]}(t) := 1$ for $a \le t \le b$ and $I_{[a,b]}(t) := 0$ otherwise. In particular, $u(t) := I_{[0,\infty)}(t)$ is the unit step function.

Preface

Intended Audience

This book contains enough material to serve as a text for a two-course sequence in probability and random processes for electrical and computer engineers. It is also useful as a reference by practicing engineers.

The material for a first course can be offered at either the undergraduate or graduate level. The prerequisite is the usual undergraduate electrical and computer engineering course on signals and systems, e.g., Haykin and Van Veen [22] or Oppenheim and Willsky [36] (see the Bibliography at the end of the book).

The material for a second course can be offered at the graduate level. The additional prerequisite is greater mathematical maturity and some familiarity with linear algebra; e.g., determinants and matrix inverses.

Material for a First Course

In a first course, Chapters 1–4 and 6 would make up the core of any offering. These chapters cover the basics of probability and discrete and continuous random variables. The following additional topics are also appropriate for a first course. include

- Chapter 5, Statistics, can be studied any time after Chapter 4. This chapter covers parameter estimation and confidence intervals, histograms, and Monte–Carlo estimation. Numerous MATLAB scripts and problems can be found here. This is a stand-alone chapter, whose material is not used elsewhere in the text, with the exception of Problem 12 in Chapter 9 and Problem 6 in Chapter 13.
- Chapter 7, Introduction to Random Processes (Sections 7.1–7.7), can be studied any time after Chapter 6. (In a more advanced course that is covering Chapter 8, Random Vectors, the ordering of Chapters 7 and 8 can be reversed without any difficulty.)*
- Section 9.1, The Poisson Process, can be covered any time after Chapter 3.
- Sections 10.1–10.2 on discrete-time Markov chains can be covered any time after Chapter 2. (Section 10.3 on continuous-time Markov chains requires material from Chapters 4 and 9.)

As noted above, a first course can be offered at the undergraduate or the graduate level. In the aforementioned material, sections/remarks/problems marked with a star ($\star$) and Chapter Notes indicated by a numerical superscript in the text are directed at graduate students. These items are easily omitted in an undergraduate course.

*Logically, the material on random vectors in Chapter 8 could be incorporated into Chapter 6, and the material in Chapter 7 could be incorporated into Chapter 9. However, the present arrangement allows greater flexibility in course structure.

Material for a Second Course

The core material for a second course would consist of the following.

- Chapter 8, Random Vectors, emphasizes the finite-dimensional Karhunen–Loève expansion, nonlinear transformations of random vectors, Gaussian random vectors, and linear and nonlinear minimum mean squared error estimation. There is also an introduction to complex random variables and vectors, especially the circularly symmetric Gaussian due to its importance in wireless communications.
- Chapter 9, Advanced Concepts in Random Processes, begins with the Poisson, renewal, and Wiener processes. The general question of the existence of processes with specified finite-dimensional distributions is addressed using Kolmogorov's theorem.
- Chapter 11, Mean Convergence and Applications, covers convergence and continuity in mean of order p. Emphasis is on the vector-space structure of random variables with finite pth moment. Applications include the continuous-time Karhunen–Loève expansion, the Wiener process, and the spectral representation. The orthogonality principle and projection theorem are also derived and used to develop conditional expectation and probability in the general case; e.g., when two random variables are individually continuous but not jointly continuous.
- Chapter 12, Other Modes of Convergence, covers convergence in probability, convergence in distribution, and almost-sure convergence. Applications include showing that mean-square integrals of Gaussian processes are Gaussian, the strong and weak laws of large numbers, and the central limit theorem.

Additional material that may be included depending on student preparation and course objectives:

- Some instructors may wish to start a second course by covering some of the starred sections and some of the Notes from Chapters 1–4 and 6 and assigning some of the starred problems.
- If students are not familiar with the material in Chapter 7, it can be covered either immediately before or after Chapter 8.
- Since Chapter 10 on Markov chains is not needed for Chapters 11–13, it can be covered any time after Chapter 9.
- If Chapter 5, Statistics, is covered after Chapter 12, then attention can be focused on the derivations and their technical details, which require facts about the multivariate Gaussian, the strong law of large numbers, and Slutsky's Theorem.
- Chapter 13, Self Similarity and Long-Range Dependence, is provided to introduce students to these concepts because they arise in network traffic modeling. Fractional Brownian motion and fractional autoregressive moving average processes are developed as prominent examples exhibiting

self similarity and long-range dependence.

Chapter Features

- Each chapter begins with a Chapter Outline listing the main sections and subsections.
- Key equations are boxed:

$$\boxed{\wp(A|B) \ := \ \frac{\wp(A \cap B)}{\wp(B)}.}$$

- Important text passages are highlighted:

Two events A and B are said to be independent if $\wp(A \cap B) = \wp(A)\,\wp(B)$.

- Tables of discrete random variables and of Fourier transform pairs are found inside the front cover. A table of continuous random variables is found inside the back cover.
- The index was compiled as the book was being written. Hence, there are many cross-references to related information. For example, see "chi-squared random variable."
- When cdfs or other functions are encountered that do not have a closed form, MATLAB commands are given for computing them. See "Matlab commands" in the index.
- Each chapter contains a **Notes** section. Throughout each chapter, numerical superscripts refer to discussions in the Notes section. These notes are usually rather technical and address subtleties of the theory that are important for students in a graduate course. The notes can be skipped in an undergraduate course.
- Each chapter contains a **Problems** section. Problems are grouped according to the section they are based on, and this is clearly indicated. This enables the student to refer to the appropriate part of the text for background relating to particular problems, and it enables the instructor to make up assignments more quickly. In chapters intended for a first course, problems marked with a * are more challenging, and may be skipped in an undergraduate offering.
- Each chapter contains an **Exam Preparation** section. This serves as a chapter summary, drawing attention to key concepts and formulas.

Table of Contents

CHAPTER 1
Introduction to Probability

Chapter Outline

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 were introduced to the following problem of detecting a known signal in additive noise. Consider an air-traffic control system that sends out a known radar pulse. If there are no objects in range of the radar, the system returns only a noise waveform. If there is an object in range, the system returns the reflected radar pulse plus noise. The overall goal is to design a system that decides whether received waveform is noise only or signal plus noise. Our class addressed the subproblem of designing an optimal, linear, time-invariant system to process the received waveform so as amplify the signal and suppress the noise

*Sections marked with a * can be omitted in an introductory course.

by maximizing the signal-to-noise ratio. We learned that the optimal transfer function is given by the matched filter. You will study this in Chapter 7.

Computer Memories. Suppose you are designing a computer memory to hold k-bit words. To increase system reliability, you actually store n-bit words, where $n \Leftrightarrow k$ of the bits are devoted to error correction. The memory is reliable if when reading a location, any k or more bits (out of n) are correctly recovered. How should n be chosen to guarantee a specified reliability? You will be able to answer questions like these after you study the binomial random variable in Chapter 2.

Optical Communication Systems. Optical communication systems use photodetectors to interface between optical and electronic subsystems. When these systems are at the limits of the operating capabilities, the number of photoelectrons produced by the photodetector is well-modeled by the Poisson[*] random variable you will study in Chapter 2 (see also the Poisson process in Chapter 9). 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 of this event.

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 3, plays a very prominent role in the analysis and design of communication systems. When noncoherent receivers are used, e.g., noncoherent frequency shift keying, this naturally leads to the Rayleigh, chi-squared, noncentral chi-squared, and Rice density functions that you will meet in the problems in Chapters 3, 4, 6, and 8.

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 5 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 [38, p. 1]. You will study Markov chains in Chapter 10. As self similarity was observed in the traffic of local-area networks [32], wide-area networks [39], 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

[*]Many quantities in probability and statistics are named after famous mathematicians and statisticians. You can use an Internet search engine to find pictures and biographies. 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 `http://www.york.ac.uk/depts/maths/histstat/people/welcome.htm`

surprising insights into questions about buffer size vs. bandwidth, multiple-time-scale congestion control, connection duration prediction, and other issues [38, pp. 9–11]. In Chapter 13 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, $i = 1, \ldots, 6$. As another example, there are $M = 52$ possible outcomes if we draw one card from a deck of playing cards. The simplest example we consider is the flipping of a coin. In this case there are two possible outcomes, "heads" and "tails." 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**.† Let $N_n(O_i)$ denote the number of times O_i occurred in n trials. The **relative frequency** of outcome O_i is defined to be

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

Thus, the relative frequency $N_n(O_i)/n$ is the fraction of times O_i occurred.

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

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

† When there are only two outcomes, the repeated experiments are called **Bernoulli trials**.

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,‡ 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 the list $N_{1000}(S_k)$ for the remaining values of k, we summarize as follows:

$$\begin{aligned}
N_{1000}(S_{33}) + N_{1000}(S_{34}) + N_{1000}(S_{35}) &= 4 \\
N_{1000}(S_{36}) + N_{1000}(S_{37}) + N_{1000}(S_{38}) &= 6 \\
N_{1000}(S_{39}) + N_{1000}(S_{40}) + N_{1000}(S_{41}) &= 32 \\
N_{1000}(S_{42}) + N_{1000}(S_{43}) + N_{1000}(S_{44}) &= 98 \\
N_{1000}(S_{45}) + N_{1000}(S_{46}) + N_{1000}(S_{47}) &= 165 \\
N_{1000}(S_{48}) + N_{1000}(S_{49}) + N_{1000}(S_{50}) &= 230 \\
N_{1000}(S_{51}) + N_{1000}(S_{52}) + N_{1000}(S_{53}) &= 214 \\
N_{1000}(S_{54}) + N_{1000}(S_{55}) + N_{1000}(S_{56}) &= 144 \\
N_{1000}(S_{57}) + N_{1000}(S_{58}) + N_{1000}(S_{59}) &= 76 \\
N_{1000}(S_{60}) + N_{1000}(S_{61}) + N_{1000}(S_{62}) &= 21 \\
N_{1000}(S_{63}) + N_{1000}(S_{64}) + N_{1000}(S_{65}) &= 9 \\
N_{1000}(S_{66}) + N_{1000}(S_{67}) + N_{1000}(S_{68}) &= 1.
\end{aligned}$$

This data is illustrated in the histogram shown in Figure 1.1. (The bars are centered over values of the form $k/100$; e.g., the bar of height 230 is centered over 0.49.)

How can we explain this statistical regularity? Why does the bell-shaped curve fit so well over the histogram?

‡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 4 and 5.

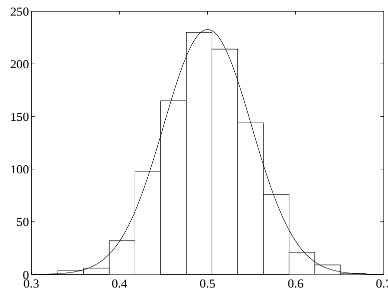


Figure 1.1. Histogram of Example 1.2 with overlay of a Gaussian density.

What Is Probability Theory?

Axiomatic probability theory, which is the subject of this book, was developed by **A. N. Kolmogorov** in 1933. This theory specifies a set of axioms that a well-defined mathematical model is required to satisfy such that if $\widehat{O}_1, \ldots, \widehat{O}_M$ are the mathematical objects that correspond to experimental outcomes $O_1, \ldots, O_M$, and if $\widehat{O}_i$ is assigned probability[§] p_i, then the probabilities of events constructed from the $\widehat{O}_i$ will satisfy the same additivity properties as relative frequency such as (1.1) and (1.2). Furthermore, using the axioms, it can be proved that

$$\lim_{n\to\infty} \frac{N_n(\widehat{O}_i)}{n} \;=\; p_i.$$

This result is a special case of the **strong law of large numbers**, which is derived in Chapter 12. (A related result, known as the **weak law of large numbers**, is derived in Chapter 2.) The law of large numbers says that the mathematical model has the statistical regularity that we observe experimentally. This is why probability theory has enjoyed great success in the analysis, design, and prediction of real-world systems.

Probability theory also explains why the histogram in Figure 1.1 agrees with the bell-shaped curve overlaying it. If probability p is assigned to $\widehat{O}_{\text{heads}}$, then the probability of $\widehat{S}_k$ (the mathematical object corresponding to S_k above) is given by the binomial probability formula (see Example 1.39 or Section 2.4)

$$\frac{n!}{k!(n \Leftrightarrow k)!}p^k(1 \Leftrightarrow p)^{n-k},$$

where $k = 0, \ldots, n$ (in Example 1.2, $n = 100$ and $p = 1/2$). By the **central limit theorem**, which is derived in Chapter 4, if n is large, the above expression is approximately equal to

$$\frac{1}{\sqrt{2\pi np(1 \Leftrightarrow p)}}\exp\left[\Leftrightarrow\frac{1}{2}\left(\frac{k \Leftrightarrow np}{\sqrt{np(1 \Leftrightarrow p)}}\right)^2\right].$$

[§]The concept of probability will be defined later.

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

Features of Each Chapter

The last three sections of every chapter are entitled Notes, Problems, and Exam Preparation, respectively. The Notes section contains additional information referenced in the text by numerical superscripts. These notes are usually rather technical and can be skipped by the beginning student. However, the notes provide a more in-depth discussion of certain topics that may be of interest to more advanced readers. The Problems section is an integral part of the book, and in some cases contains developments beyond those in the main text. The instructor may wish to solve some of these problems in the lectures. Remarks, problems, and sections marked by a * are intended for more advanced readers, and can be omitted in a first course. The Exam Preparation section provides a few study suggestions, including references to the more important concepts and formulas introduced in the chapter.

1.1. Review of Set Notation

Since Kolmogorov's axiomatic theory is expressed in 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.

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 1.2(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.,

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

This is illustrated in Figure 1.2(b), in which the shaded region is the complement of the disk A.

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

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

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

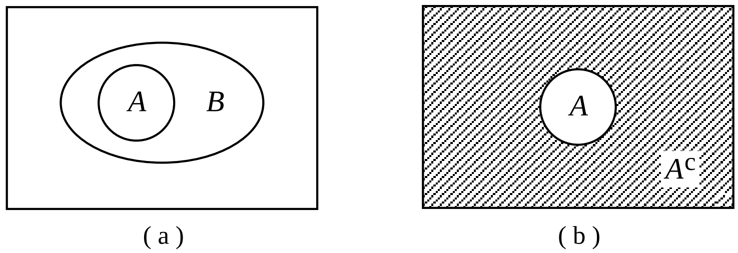


Figure 1.2. (a) Venn diagram of $A \subset B$. (b) The complement of the disk A, denoted by A^c, is the shaded part of the diagram.

Here "or" is inclusive; i.e., if $\omega \in A \cup B$, we permit ω to belong to either A or B or both. This is illustrated in Figure 1.3(a), in which the shaded region is the union of the disk A and the oval-shaped region 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 1.3(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 1.2(a)), then $A \cap B = A$. In particular, we always have $A \cap \Omega = A$ and $\varnothing \cap B = \varnothing$.

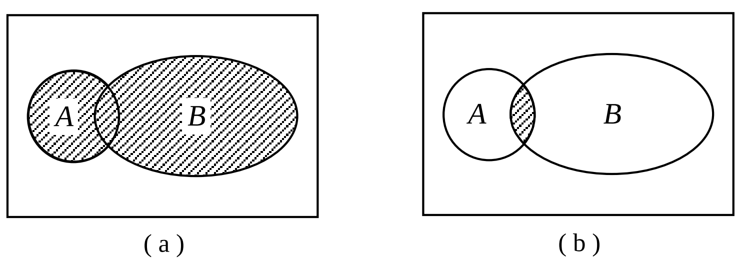


Figure 1.3. (a) The shaded region is $A \cup B$. (b) The shaded region is $A \cap B$.

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 1.4(a), $B \setminus A$ is the shaded part of the oval-shaped region B.

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 1.4(b).

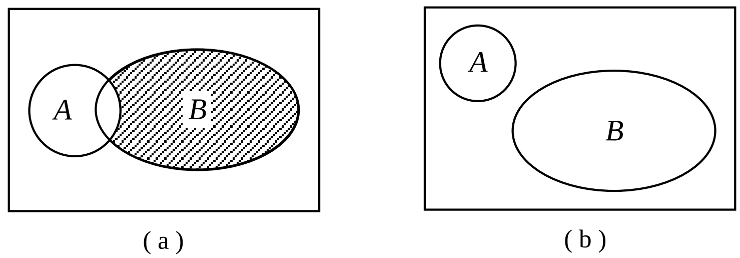


Figure 1.4. (a) The shaded region is $B \setminus A$. (b) Venn diagram of disjoint sets A and B.

***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 $A^c = \{0,5,6,7\}$,

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

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

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

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 $\emptyset \cap B = \emptyset$; hence, we can think of Ω as the analog of the number one and $\emptyset$ as the analog of the number zero. Another analog is the formula $A \cup \emptyset = A$.

We next consider infinite collections of subsets of Ω. Suppose $A_n \subset \Omega$, $n = 1, 2, \ldots$. Then

$$\bigcup_{n=1}^{\infty} A_n \;:=\; \{\omega \in \Omega : \omega \in A_n \text{ for some } 1 \leq n < \infty\}.$$

In other words, $\omega \in \bigcup_{n=1}^{\infty} A_n$ if and only if for at least one integer n satisfying $1 \leq n < \infty$, $\omega \in A_n$. This definition admits the possibility that $\omega \in A_n$ for more than one value of n. Next, we define

$$\bigcap_{n=1}^{\infty} A_n \;:=\; \{\omega \in \Omega : \omega \in A_n \text{ for all } 1 \leq n < \infty\}.$$

In other words, $\omega \in \bigcap_{n=1}^{\infty} A_n$ if and only if $\omega \in A_n$ for every positive integer n.

***Example* 1.4.** Let Ω denote the real numbers, $\Omega = \mathbb{R} := (-\infty, \infty)$. Then the following infinite intersections and unions can be simplified. Consider the intersection

$$\bigcap_{n=1}^{\infty} (-\infty, 1/n) \;=\; \{\omega : \omega < 1/n \text{ for all } 1 \leq n < \infty\}.$$

Now, if $\omega < 1/n$ for all $1 \leq n < \infty$, then ω cannot be positive; i.e., we must have $\omega \leq 0$. Conversely, if $\omega \leq 0$, then for all $1 \leq n < \infty$, $\omega \leq 0 < 1/n$. It follows that

$$\bigcap_{n=1}^{\infty} (-\infty, 1/n) \;=\; (-\infty, 0].$$

Consider the infinite union,

$$\bigcup_{n=1}^{\infty} (-\infty, -1/n] \;=\; \{\omega : \omega \leq -1/n \text{ for some } 1 \leq n < \infty\}.$$

Now, if $\omega \leq -1/n$ for some n with $1 \leq n < \infty$, then we must have $\omega < 0$. Conversely, if $\omega < 0$, then for large enough n, $\omega \leq -1/n$. Thus,

$$\bigcup_{n=1}^{\infty} (-\infty, -1/n] \;=\; (-\infty, 0).$$

In a similar way, one can show that

$$\bigcap_{n=1}^{\infty} [0, 1/n) \;=\; \{0\},$$

as well as

$$\bigcup_{n=1}^{\infty}(-\infty, n] \;=\; (-\infty, \infty) \quad \text{and} \quad \bigcap_{n=1}^{\infty}(-\infty, -n] \;=\; \varnothing.$$

The following **generalized distributive laws** also hold,

$$B \cap \left(\bigcup_{n=1}^{\infty} A_n \right) \;=\; \bigcup_{n=1}^{\infty} (B \cap A_n),$$

and

$$B \cup \left(\bigcap_{n=1}^{\infty} A_n \right) \;=\; \bigcap_{n=1}^{\infty} (B \cup A_n).$$

We also have the **generalized De Morgan's laws**,

$$\left(\bigcap_{n=1}^{\infty} A_n \right)^c \;=\; \bigcup_{n=1}^{\infty} A_n^c,$$

and

$$\left(\bigcup_{n=1}^{\infty} A_n \right)^c \;=\; \bigcap_{n=1}^{\infty} A_n^c.$$

Finally, we will need the following definition. We say that subsets $A_n, n = 1, 2, \ldots,$ are **pairwise disjoint** if $A_n \cap A_m = \varnothing$ for all $n \neq m$.

Partitions

A family of sets B_n is called a **partition** if the sets are pairwise disjoint and their union is the whole space Ω. A partition of three sets B_1, B_2, and B_3 is illustrated in Figure 1.5(a). Partitions are useful for chopping up sets into manageable, disjoint pieces. Given a set A, write

$$\begin{aligned} A \;&=\; A \cap \Omega \\ &=\; A \cap \left(\bigcup_n B_n \right) \\ &=\; \bigcup_n (A \cap B_n). \end{aligned}$$

Since the B_n are pairwise disjoint, so are the pieces $(A \cap B_n)$. This is illustrated in Figure 1.5(b), in which a disk is broken up into three disjoint pieces.

If a family of sets B_n is disjoint but their union is not equal to the whole space, we can always add the remainder set

$$R \;:=\; \left(\bigcup_n B_n \right)^c \tag{1.8}$$

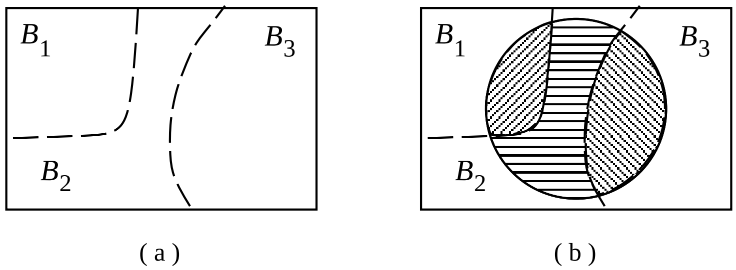


Figure 1.5. (a) The partition B_1, B_2, B_3. (b) Using the partition to break up a disk into three disjoint pieces (the shaded regions).

to the family to create a partition. Writing

$$\begin{aligned} \Omega &= R^c \cup R \\ &= \left(\bigcup_n B_n \right) \cup R, \end{aligned}$$

we see that the union of the augmented family is the whole space. It only remains to show that $B_k \cap R = \varnothing$. Write

$$\begin{aligned} B_k \cap R &= B_k \cap \left(\bigcup_n B_n \right)^c \\ &= B_k \cap \left(\bigcap_n B_n^c \right) \\ &= B_k \cap B_k^c \cap \left(\bigcap_{n \neq k} B_n^c \right) \\ &= \varnothing. \end{aligned}$$

*Functions

A **function** consists of a set X of admissible inputs called the **domain** and a **rule** or **mapping** f that associates to each $x \in X$ a value $f(x)$ that belongs to a set Y called the **co-domain**. We indicate this symbolically by writing $f: X \to Y$, and we say, "f maps X into Y." Two functions are the same if and only if they have the same domain, co-domain, and rule. If $f: X \to Y$ and $g: X \to Y$, then the mappings f and g are the same if and only if $f(x) = g(x)$ for all $x \in X$.

The set of all possible values of $f(x)$ is called the **range**. The range of a function is the set $\{f(x) : x \in X\}$. In general, the range is a proper subset of the co-domain.

A function is said to be **onto** if its range is equal to its co-domain. In other words, every value $y \in Y$ "comes from somewhere" in the sense that for every $y \in Y$, there is at least one $x \in X$ with $y = f(x)$.

*Sections marked with a * can be omitted in an introductory course.

A function is said to be **one-to-one** if the condition $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

Another way of thinking about the concepts of onto and one-to-one is the following. A function is onto if for every $y \in Y$, the equation $f(x) = y$ has a solution. This does not rule out the possibility that there may be more than one solution. A function is one-to-one if for every $y \in Y$, the equation $f(x) = y$ can have at most one solution. This does not rule out the possibility that for some values of $y \in Y$, there may be *no* solution.

A function is said to be **invertible** if for every $y \in Y$ there is a unique $x \in X$ with $f(x) = y$. Hence, a function is invertible if and only if it is both one-to-one and onto; i.e., for every $y \in Y$, the equation $f(x) = y$ has a unique solution.

***Example* 1.5.** For any real number x, put $f(x) := x^2$. Then

$$
\begin{aligned}
f&: (-\infty, \infty) \to (-\infty, \infty) \\
f&: (-\infty, \infty) \to [0, \infty) \\
f&: [0, \infty) \to (-\infty, \infty) \\
f&: [0, \infty) \to [0, \infty)
\end{aligned}
$$

specifies four different functions. In the first case, the function is not one-to-one because $f(2) = f(-2)$, but $2 \neq -2$; the function is not onto because there is no $x \in (-\infty, \infty)$ with $f(x) = -1$. In the second case, the function is onto since for every $y \in [0, \infty)$, $f(\sqrt{y}) = y$. However, since $f(-\sqrt{y}) = y$ also, the function is not one-to-one. In the third case, the function fails to be onto, but is one-to-one. In the fourth case, the function is onto and one-to-one and therefore invertible.

The last concept we introduce concerning functions is that of inverse image. If $f: X \to Y$, and if $B \subset Y$, then the **inverse image** of B is

$$f^{-1}(B) \;:=\; \{x \in X : f(x) \in B\},$$

which we emphasize is a subset of X. This concept applies to any function whether or not it is invertible. When the set X is understood, we sometimes write

$$f^{-1}(B) \;:=\; \{x : f(x) \in B\}$$

to simplify the notation.

***Example* 1.6.** If $f: (-\infty, \infty) \to (-\infty, \infty)$, where $f(x) = x^2$, find $f^{-1}([4, 9])$ and $f^{-1}([-9, -4])$.

Solution. In the first case, write

$$f^{-1}([4, 9]) \quad = \quad \{x : f(x) \in [4, 9]\}$$

$$\begin{aligned}
&= \{x : 4 \le f(x) \le 9\} \\
&= \{x : 4 \le x^2 \le 9\} \\
&= \{x : 2 \le x \le 3 \text{ or } -3 \le x \le -2\} \\
&= [2,3] \cup [-3,-2].
\end{aligned}$$

In the second case, we need to find

$$f^{-1}([-9,-4]) \;=\; \{x : -9 \le x^2 \le -4\}.$$

Since this is no $x \in (-\infty, \infty)$ with $x^2 < 0$, $f^{-1}([-9,-4]) = \varnothing$.

Remark. If we modify the function in the preceding example to be $f\colon [0,\infty) \to (-\infty,\infty)$, then $f^{-1}([4,9]) = [2,3]$ instead.

*Countable and Uncountable Sets

The number of points in a set A is denoted by $|A|$. We call $|A|$ the **cardinality** of A. The cardinality of a set may be finite or infinite. A little reflection should convince you that if A and B are two disjoint sets, then

$$|A \cup B| \;=\; |A| + |B|.$$

Use the convention that if x is a real number, then

$$x + \infty \;=\; \infty \quad \text{and} \quad \infty + \infty \;=\; \infty,$$

and be sure to consider the three cases: (i) A and B both have finite cardinality, (ii) one has finite cardinality and one has infinite cardinality, and (iii) both have infinite cardinality.

A set A is said to be **countable** if the elements of A can be enumerated or listed in a sequence: $a_1, a_2, \ldots$. In other words, a set A is countable if it can be written in the form

$$A \;=\; \bigcup_{k=1}^{\infty} \{a_k\},$$

where we emphasize that the union is over the positive integers, $k = 1, 2, \ldots$.

Remark. Since there is no requirement that the a_k be distinct, every finite set is countable by our definition. For example, you should verify that the set $A = \{1,2,3\}$ can be written in the above form by taking $a_1 = 1, a_2 = 2, a_3 = 3$, and $a_k = 3$ for $k = 4, 5, \ldots$. By a **countably infinite set**, we mean a countable set that is not finite.

***Example* 1.7.** Show that a set of the form

$$B \;=\; \bigcup_{i,j=1}^{\infty} \{b_{ij}\}$$

*Sections marked with a * can be omitted in an introductory course.

is countable.

Solution. The point here is that a sequence that is doubly indexed by positive integers forms a countable set. To see this, consider the array

$$\begin{array}{cccc} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & \\ b_{31} & b_{32} & & \\ b_{41} & & & \ddots \end{array}$$

Now list the array elements along antidiagonals from lower left to upper right defining

$$\begin{array}{llll} a_1 := b_{11} & & & \\ a_2 := b_{21}, & a_3 := b_{12} & & \\ a_4 := b_{31}, & a_5 := b_{22}, & a_6 := b_{13} & \\ a_7 := b_{41}, & a_8 := b_{32}, & a_9 := b_{23}, & a_{10} := b_{14} \\ \cdots & & & \end{array}$$

This shows that

$$B = \bigcup_{k=1}^{\infty} \{a_k\},$$

and so B is a countable set.

***Example* 1.8.** Show that the positive rational numbers form a countable subset.

Solution. Recall that a rational number is of the form i/j where i and j are integers with $j \neq 0$. Hence, the set of positive rational numbers is equal to

$$\bigcup_{i,j=1}^{\infty} \{i/j\}.$$

By the previous example, this is a countable set.

You will show in Problem 12 that the union of two countable sets is a countable set. It then easily follows that the set of all rational numbers is countable.

A set is **uncountable** or **uncountably infinite** if it is not countable.

***Example* 1.9.** Show that the set S of unending sequences of zeros and ones is uncountable.

Solution. Suppose, to obtain a contradiction, that S is countable. Then we can exhaustively list all the elements of S. Such a list would look like

$$\begin{array}{lcccccccccccc}
\mathbf{a}_1 & := & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & \cdots \\
\mathbf{a}_2 & := & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & \cdots \\
\mathbf{a}_3 & := & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & \cdots \\
\mathbf{a}_4 & := & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & \cdots \\
\mathbf{a}_5 & := & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
 & \vdots & & & & & & \ddots & & & &
\end{array}$$

But this list can never be complete. To construct a new binary sequence that is not in the list, use the following **diagonal argument**. Take $\mathbf{a} := 0\,1\,0\,0\,1\,\cdots$ to be such that kth bit of $\mathbf{a}$ is the complement of the kth bit of $\mathbf{a}_k$. In other words, viewing the above sequences as an infinite matrix, go along the diagonal and flip all the bits to construct $\mathbf{a}$. Then $\mathbf{a} \neq \mathbf{a}_1$ because they differ in the first bit. Similarly, $\mathbf{a} \neq \mathbf{a}_2$ because they differ in the second bit. And so on.

The same argument shows that the interval of real numbers $[0,1)$ is not countable. Write each fractional real number in its binary expansion, e.g., $0.11010101110\ldots$ and identify the expansion with the corresponding sequence of zeros and ones in the example.

1.2. Probability Models

In this section, we introduce a number of simple physical experiments and suggest mathematical probability models for them. These models are used to compute various probabilities of interest.

Consider the experiment of tossing a fair die and measuring, i.e., noting, the face turned up. Our intuition tells us that the "probability" of the ith face turning up is $1/6$, and that the "probability" of a face with an even number of dots turning up is $1/2$.

Here is a *mathematical model* for this experiment and measurement. Let Ω be any set containing six points. We call Ω the **sample space**. Each point in Ω corresponds to, or models, a possible outcome of the experiment. The individual points $\omega \in \Omega$ are called **sample points** or **outcomes**. For simplicity, let

$$\Omega := \{1,2,3,4,5,6\}.$$

Now put

$$F_i := \{i\}, \quad i = 1,2,3,4,5,6,$$

and

$$E := \{2,4,6\}.$$

We call the sets F_i and E events. An **event** is a collection of outcomes. The event F_i corresponds to, or models, the die's turning up showing the ith face.

Similarly, the event E models the die's showing a face with an even number of dots. Next, for every subset A of Ω, we denote the number of points in A by $|A|$. We call $|A|$ the **cardinality** of A. We *define* the **probability** of any event A by

$$\wp(A) \ := \ |A|/|\Omega|.$$

In other words, for the model we are constructing for this problem, the probability of an event A is defined to be the number of outcomes in A divided by the total number of possible outcomes. With this definition, it follows that $\wp(F_i) = 1/6$ and $\wp(E) = 3/6 = 1/2$, which agrees with our intuition.

We now make four observations about our model:

- (i) $\wp(\emptyset) = |\emptyset|/|\Omega| = 0/|\Omega| = 0$.
- (ii) $\wp(A) \geq 0$ for every event A.
- (iii) If A and B are mutually exclusive events, i.e., $A \cap B = \emptyset$, then $\wp(A \cup B) = \wp(A) + \wp(B)$; for example, $F_3 \cap E = \emptyset$, and it is easy to check that $\wp(F_3 \cup E) = \wp(\{2,3,4,6\}) = \wp(F_3) + \wp(E)$.
- (iv) When the die is tossed, *something* happens; this is modeled mathematically by the easily verified fact that $\wp(\Omega) = 1$.

As we shall see, these four properties hold for all the models discussed in this section.

We next modify our model to accommodate an unfair die as follows. Observe that for a fair die,¶

$$\wp(A) \ = \ \frac{|A|}{|\Omega|} \ = \ \sum_{\omega \in A} \frac{1}{|\Omega|} \ = \ \sum_{\omega \in A} p(\omega),$$

where $p(\omega) := 1/|\Omega|$. For an *unfair* die, we simply change the definition of the function $p(\omega)$ to reflect the likelihood of occurrence of the various faces. This new definition of $\wp$ still satisfies (i) and (iii); however, to guarantee that (ii) and (iv) still hold, we must require that p be nonnegative and sum to one, or, in symbols, $p(\omega) \geq 0$ and $\sum_{\omega \in \Omega} p(\omega) = 1$.

***Example* 1.10.** Construct a sample space Ω and probability $\wp$ to model an unfair die in which faces 1–5 are equally likely, but face 6 has probability 1/3. Using this model, compute the probability that a toss results in a face showing an even number of dots.

Solution. We again take $\Omega = \{1,2,3,4,5,6\}$. To make face 6 have probability 1/3, we take $p(6) = 1/3$. Since the other faces are equally likely, for $\omega = 1, \ldots, 5$, we take $p(\omega) = c$, where c is a constant to be determined. To find c we use the fact that

$$1 \ = \ \wp(\Omega) \ = \ \sum_{\omega \in \Omega} p(\omega) \ = \ \sum_{\omega=1}^{6} p(\omega) \ = \ 5c + \frac{1}{3}.$$

¶If $A = \emptyset$, the summation is taken to be zero.

It follows that $c = 2/15$. Now that $p(\omega)$ has been specified for all ω, we define the probability of any event A by

$$\wp(A) \;:=\; \sum_{\omega \in A} p(\omega).$$

Letting $E = \{2, 4, 6\}$ model the result of a toss showing a face with an even number of dots, we compute

$$\wp(E) \;=\; \sum_{\omega \in E} p(\omega) \;=\; p(2) + p(4) + p(6) \;=\; \frac{2}{15} + \frac{2}{15} + \frac{1}{3} \;=\; \frac{3}{5}.$$

This unfair die has a greater probability of showing an even numbered face that the fair die.

This problem is typical of the kinds of "word problems" to which probability theory is applied to analyze well-defined physical experiments. The application of probability theory requires the modeler to take the following steps:

1. Select a suitable sample space Ω.
2. Define $\wp(A)$ for all events A. For example, if Ω is a finite set and all outcomes ω are equally likely, we usually take $\wp(A) = |A|/|\Omega|$. If it is not the case that all outcomes are equally likely, e.g., as in the previous example, then $\wp(A)$ would be given by some other formula that you must determine based on the problem statement.
3. Translate the given "word problem" into a problem requiring the calculation of $\wp(E)$ for some specific event E.

The following example gives a family of constructions that can be used to model experiments having a finite number of possible outcomes.

***Example* 1.11.** Let M be a positive integer, and put $\Omega := \{1, 2, \ldots, M\}$. Next, let $p(1), \ldots, p(M)$ be nonnegative real numbers such that $\sum_{\omega=1}^{M} p(\omega) = 1$. For any subset $A \subset \Omega$, put

$$\wp(A) \;:=\; \sum_{\omega \in A} p(\omega).$$

In particular, to model equally likely outcomes, or equivalently, outcomes that occur "at random," we take $p(\omega) = 1/M$. In this case, $\wp(A)$ reduces to $|A|/|\Omega|$.

***Example* 1.12.** A single card is drawn at random from a well-shuffled deck of playing cards. Find the probability of drawing an ace. Also find the probability of drawing a face card.

Solution. The first step in the solution is to specify the sample space Ω and the probability $\wp$. Since there are 52 possible outcomes, we take $\Omega :=$

$\{1,\ldots,52\}$. Each integer corresponds to one of the cards in the deck. To specify $\wp$, we must define $\wp(E)$ for all events $E \subset \Omega$. Since all cards are equally likely to be drawn, we put $\wp(E) := |E|/|\Omega|$.

To find the desired probabilities, let $1,2,3,4$ correspond to the four aces, and let $41,\ldots,52$ correspond to the 12 face cards. We identify the drawing of an ace with the event $A := \{1,2,3,4\}$, and we identify the drawing of a face card with the event $F := \{41,\ldots,52\}$. It then follows that $\wp(A) = |A|/52 = 4/52 = 1/13$ and $\wp(F) = |F|/52 = 12/52 = 3/13$.

While the sample spaces Ω in Example 1.11 can model any experiment with a finite number of outcomes, it is often convenient to use alternative sample spaces.

***Example* 1.13.** Suppose that we have two well-shuffled decks of cards, and we draw one card at random from each deck. What is the probability of drawing the ace of spades followed by the jack of hearts? What is the probability of drawing an ace and a jack (in either order)?

Solution. The first step in the solution is to specify the sample space Ω and the probability $\wp$. Since there are 52 possibilities for each draw, there are $52^2 = 2{,}704$ possible outcomes when drawing two cards. Let $D := \{1,\ldots,52\}$, and put

$$\Omega \;:=\; \{(i,j) : i,j \in D\}.$$

Then $|\Omega| = |D|^2 = 52^2 = 2{,}704$ as required. Since all pairs are equally likely, we put $\wp(E) := |E|/|\Omega|$ for arbitrary events $E \subset \Omega$.

As in the preceding example, we denote the aces by $1,2,3,4$. We let 1 denote the ace of spades. We also denote the jacks by $41,42,43,44$, and the jack of hearts by 42. The drawing of the ace of spades followed by the jack of hearts is identified with the event

$$A \;:=\; \{(1,42)\},$$

and so $\wp(A) = 1/2{,}704 \approx 0.000370$. The drawing of an ace and a jack is identified with $B := B_{\mathrm{aj}} \cup B_{\mathrm{ja}}$, where

$$B_{\mathrm{aj}} \;:=\; \bigl\{(i,j) : i \in \{1,2,3,4\} \text{ and } j \in \{41,42,43,44\}\bigr\}$$

corresponds to the drawing of an ace followed by a jack, and

$$B_{\mathrm{ja}} \;:=\; \bigl\{(i,j) : i \in \{41,42,43,44\} \text{ and } j \in \{1,2,3,4\}\bigr\}$$

corresponds to the drawing of a jack followed by an ace. Since B_{aj} and B_{ja} are disjoint, $\wp(B) = \wp(B_{\mathrm{aj}}) + \wp(B_{\mathrm{ja}}) = (|B_{\mathrm{aj}}| + |B_{\mathrm{ja}}|)/|\Omega|$. Since $|B_{\mathrm{aj}}| = |B_{\mathrm{ja}}| = 16$, $\wp(B) = 2 \cdot 16/2{,}704 = 2/169 \approx 0.0118$.

***Example* 1.14.** Two cards are drawn at random from a *single* well-shuffled deck of playing cards. What is the probability of drawing the ace of spades followed by the jack of hearts? What is the probability of drawing an ace and a jack (in either order)?

Solution. The first step in the solution is to specify the sample space Ω and the probability $\wp$. There are 52 possibilities for the first draw and 51 possibilities for the second. Hence, the sample space should contain $52 \cdot 51 = 2{,}652$ elements. Using the notation of the preceding example, we take

$$\Omega \; := \; \{(i,j) : i, j \in D \text{ with } i \neq j\},$$

Note that $|\Omega| = 52^2 \Leftrightarrow 52 = 2{,}652$ as required. Again, all such pairs are equally likely, and so we take $\wp(E) := |E|/|\Omega|$ for arbitrary events $E \subset \Omega$. The events A and B are defined as before, and the calculation is the same except that $|\Omega| = 2{,}652$ instead of 2,704. Hence, $\wp(A) = 1/2{,}652 \approx 0.000377$, and $\wp(B) = 2 \cdot 16/2{,}652 = 8/663 \approx 0.012$.

In some experiments, the number of possible outcomes is countably infinite. For example, consider the tossing of a coin until the first heads appears. Here is a model for such situations. Let Ω denote the set of all positive integers, $\Omega := \{1, 2, \ldots\}$. For $\omega \in \Omega$, let $p(\omega)$ be nonnegative, and suppose that $\sum_{\omega=1}^{\infty} p(\omega) = 1$. For any subset $A \subset \Omega$, put

$$\wp(A) \; := \; \sum_{\omega \in A} p(\omega).$$

This construction can be used to model the coin tossing experiment by identifying $\omega = i$ with the outcome that the first heads appears on the ith toss. If the probability of tails on a single toss is α $(0 \leq \alpha < 1)$, it can be shown that we should take $p(\omega) = \alpha^{\omega-1}(1 \Leftrightarrow \alpha)$ (cf. Example 2.9). To find the probability that the first head occurs before the fourth toss, we compute $\wp(A)$, where $A = \{1, 2, 3\}$. Then

$$\wp(A) \; = \; p(1) + p(2) + p(3) \; = \; (1 + \alpha + \alpha^2)(1 \Leftrightarrow \alpha).$$

If $\alpha = 1/2$, $\wp(A) = (1 + 1/2 + 1/4)/2 = 7/8$.

For some experiments, the number of possible outcomes is more than countably infinite. Examples include the lifetime of a lightbulb or a transistor, a noise voltage in a radio receiver, and the arrival time of a city bus. In these cases, $\wp$ is usually defined as an integral,

$$\wp(A) \; := \; \int_A f(\omega)\, d\omega, \quad A \subset \Omega,$$

for some nonnegative function f. Note that f must also satisfy $\int_\Omega f(\omega)\, d\omega = 1$.

***Example* 1.15.** Consider the following model for the lifetime of a lightbulb. For the sample space we take the nonnegative half line, $\Omega := [0, \infty)$, and we put

$$\wp(A) \;:=\; \int_A f(\omega)\, d\omega,$$

where, for example, $f(\omega) := e^{-\omega}$. Then the probability that the lightbulb's lifetime is between 5 and 7 time units is

$$\wp([5,7]) \;=\; \int_5^7 e^{-\omega}\, d\omega \;=\; e^{-5} \Leftrightarrow e^{-7}.$$

***Example* 1.16.** A certain bus is scheduled to pick up riders at 9:15. However, it is known that the bus arrives randomly in the 20-minute interval between 9:05 and 9:25, and departs immediately after boarding waiting passengers. Find the probability that the bus arrives at or after its scheduled pick-up time.

Solution. Let $\Omega := [5, 25]$, and put

$$\wp(A) \;:=\; \int_A f(\omega)\, d\omega.$$

Now, the term "randomly" in the problem statement is usually taken to mean that $f(\omega) \equiv \text{constant}$. In order that $\wp(\Omega) = 1$, we must choose the constant to be $1/\text{length}(\Omega) = 1/20$. We represent the bus arriving at or after 9:15 with the event $L := [15, 25]$. Then

$$\wp(L) \;=\; \int_{[15,25]} \frac{1}{20}\, d\omega \;=\; \int_{15}^{25} \frac{1}{20}\, d\omega \;=\; \frac{25 \Leftrightarrow 15}{20} \;=\; \frac{1}{2}.$$

***Example* 1.17.** A dart is thrown at random toward a circular dartboard of radius 10 cm. Assume the thrower never misses the board. Find the probability that the dart lands within 2 cm of the center.

Solution. Let $\Omega := \{(x,y) : x^2 + y^2 \le 100\}$, and for any $A \subset \Omega$, put

$$\wp(A) \;:=\; \frac{\text{area}(A)}{\text{area}(\Omega)} \;=\; \frac{\text{area}(A)}{100\pi}.$$

We then identify the event $A := \{(x,y) : x^2 + y^2 \le 4\}$ with the dart's landing within 2 cm of the center. Hence,

$$\wp(A) \;=\; \frac{4\pi}{100\pi} \;=\; 0.04.$$

1.3. Axioms and Properties of Probability

In this section, we present Kolmogorov's axioms and derive some of their consequences.

The probability models of the preceding section suggest the following axioms that we now require of any probability model.

Given a nonempty set Ω, called the **sample space**, and a function $\wp$ defined on the subsets of Ω, we say $\wp$ is a **probability measure** if the following four axioms are satisfied:[1]

(i) The empty set $\varnothing$ is called the **impossible event**. The probability of the impossible event is zero; i.e., $\wp(\varnothing) = 0$.

(ii) Probabilities are nonnegative; i.e., for any event A, $\wp(A) \geq 0$.

(iii) If $A_1, A_2, \ldots$ are events that are mutually exclusive or pairwise disjoint, i.e., $A_n \cap A_m = \varnothing$ for $n \neq m$, then[‖]

$$\wp\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \wp(A_n). \tag{1.9}$$

This property is summarized by saying that the probability of the union of disjoint events is the sum of the probabilities of the individual events, or more briefly, "the probabilities of disjoint events add."

(iv) The entire sample space Ω is called the **sure event** or the **certain event**. Its probability is always one; i.e., $\wp(\Omega) = 1$.

A probability measure is a function whose argument is an event and whose value is a nonnegative real number. The foregoing axioms imply many other properties. In particular, we show later that the value of a probability measure must lie in the interval $[0, 1]$.

At this point, advanced readers, especially graduate students, should read Note 1 in the Notes section at the end of the chapter.

We now give an interpretation of how Ω and $\wp$ model randomness. We view the sample space Ω as being the set of all possible "states of nature." First, Mother Nature chooses a state $\omega_0 \in \Omega$. We do not know which state has been chosen. We then conduct an experiment, and based on some physical measurement, we are able to determine that $\omega_0 \in A$ for some event $A \subset \Omega$. In some cases, $A = \{\omega_0\}$, that is, our measurement reveals exactly which state ω_0 was chosen by Mother Nature. (This is the case for the events F_i defined at the beginning of Section 1.2). In other cases, the set A contains ω_0 as well as other points of the sample space. (This is the case for the event E defined at the beginning of Section 1.2). In either case, we do not know before making

[‖]See the paragraph **Finite Disjoint Unions** below and Problem 25 for further discussion regarding this axiom.

the measurement what measurement value we will get, and so we do not know what event A Mother Nature's ω_0 will belong to. Hence, in many applications, e.g., gambling, weather prediction, computer message traffic, etc., it is useful to compute $\wp(A)$ for various events to determine which ones are most probable.

Consequences of the Axioms

Axioms (i)–(iv) that characterize a probability measure have several important implications as discussed below.

Finite Disjoint Unions. Let N be a positive integer. By taking $A_n = \varnothing$ for $n > N$ in axiom (iii), we obtain the special case

$$\boxed{\wp\biggl(\bigcup_{n=1}^{N} A_n\biggr) \;=\; \sum_{n=1}^{N} \wp(A_n), \quad A_n \text{ pairwise disjoint.}}$$

Remark. It is not possible to go backwards and use this special case to derive axiom (iii).

***Example* 1.18.** If A is an event consisting of a finite number of sample points, say $A = \{\omega_1, \ldots, \omega_N\}$, then[2] $\wp(A) = \sum_{n=1}^{N} \wp(\{\omega_n\})$. Similarly, if A consists of a countably many sample points, say $A = \{\omega_1, \omega_2, \ldots\}$, then directly from axiom (iii), $\wp(A) = \sum_{n=1}^{\infty} \wp(\{\omega_n\})$.

Probability of a Complement. Given an event A, we can always write $\Omega = A \cup A^c$, which is a finite disjoint union. Hence, $\wp(\Omega) = \wp(A) + \wp(A^c)$. Since $\wp(\Omega) = 1$, we find that

$$\boxed{\wp(A^c) \;=\; 1 \Leftrightarrow \wp(A).} \tag{1.10}$$

Monotonicity. If A and B are events, then

$$\boxed{A \subset B \quad \text{implies} \quad \wp(A) \le \wp(B).} \tag{1.11}$$

To see this, first note that $A \subset B$ implies

$$B \;=\; A \cup (B \cap A^c).$$

This relation is depicted in Figure 1.6, in which the disk A is a subset of the oval-shaped region B; the shaded region is $B \cap A^c$. The figure shows that B is the disjoint union of the disk A together with the shaded region $B \cap A^c$. Since $B = A \cup (B \cap A^c)$ is a disjoint union, and since probabilities are nonnegative,

$$\begin{aligned} \wp(B) &= \wp(A) + \wp(B \cap A^c) \\ &\ge \wp(A). \end{aligned}$$

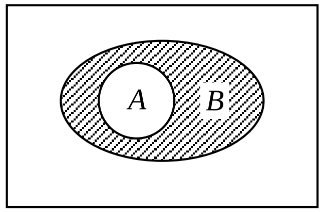


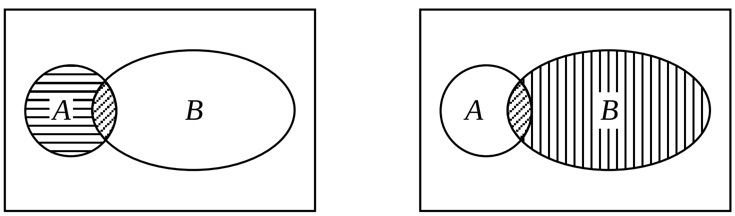
Figure 1.6. In this diagram, the disk A is a subset of the oval-shaped region B; the shaded region is $B \cap A^c$, and $B = A \cup (B \cap A^c)$.

Note that the special case $B = \Omega$ results in $\wp(A) \leq 1$ for every event A. In other words, *probabilities are always less than or equal to one.*

Inclusion-Exclusion. Given any two events A and B, we always have

$$\boxed{\wp(A \cup B) \;=\; \wp(A) + \wp(B) \Leftrightarrow \wp(A \cap B).} \tag{1.12}$$

To derive (1.12), first note that (see Figure 1.7)



(a) (b)

Figure 1.7. (a) Decomposition $A = (A \cap B^c) \cup (A \cap B)$. (b) Decomposition $B = (A \cap B) \cup (A^c \cap B)$.

$$A \;=\; (A \cap B^c) \cup (A \cap B)$$

and

$$B \;=\; (A \cap B) \cup (A^c \cap B).$$

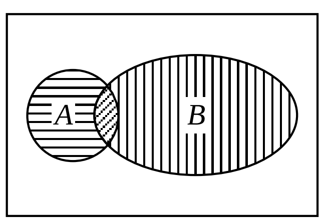


Figure 1.8. Decomposition $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$.

Hence,

$$A \cup B = \big[(A \cap B^c) \cup (A \cap B)\big] \cup \big[(A \cap B) \cup (A^c \cap B)\big].$$

The two copies of $A \cap B$ can be reduced to one using the identity $F \cup F = F$ for any set F. Thus,

$$A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B).$$

A Venn diagram depicting this last decomposition is shown in Figure 1.8. Taking probabilities of the preceding equations, which involve disjoint unions, we find that

$$\begin{aligned}
\wp(A) &= \wp(A \cap B^c) + \wp(A \cap B), \\
\wp(B) &= \wp(A \cap B) + \wp(A^c \cap B), \\
\wp(A \cup B) &= \wp(A \cap B^c) + \wp(A \cap B) + \wp(A^c \cap B).
\end{aligned}$$

Using the first two equations, solve for $\wp(A \cap B^c)$ and $\wp(A^c \cap B)$, respectively, and then substitute into the first and third terms on the right-hand side of the last equation. This results in

$$\begin{aligned}
\wp(A \cup B) &= \big[\wp(A) \Leftrightarrow \wp(A \cap B)\big] + \wp(A \cap B) \\
&\quad + \big[\wp(B) \Leftrightarrow \wp(A \cap B)\big] \\
&= \wp(A) + \wp(B) \Leftrightarrow \wp(A \cap B).
\end{aligned}$$

Limit Properties. Using axioms (i)–(iv), the following formulas can be derived (see Problems 26–28). For *any* sequence of events A_n,

$$\boxed{\wp\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{N \to \infty} \wp\left(\bigcup_{n=1}^{N} A_n\right),} \tag{1.13}$$

and

$$\boxed{\wp\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{N \to \infty} \wp\left(\bigcap_{n=1}^{N} A_n\right).} \tag{1.14}$$

In particular, notice that if the A_n are increasing in the sense that $A_n \subset A_{n+1}$ for all n, then the finite union in (1.13) reduces to A_N (see Figure 1.9(a)). Thus, (1.13) becomes

$$\boxed{\wp\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{N \to \infty} \wp(A_N), \quad \text{if } A_n \subset A_{n+1}.} \tag{1.15}$$

Similarly, if the A_n are decreasing in the sense that $A_{n+1} \subset A_n$ for all n, then the finite intersection in (1.14) reduces to A_N (see Figure 1.9(b)). Thus, (1.14) becomes

$$\boxed{\wp\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{N \to \infty} \wp(A_N), \quad \text{if } A_{n+1} \subset A_n.} \tag{1.16}$$

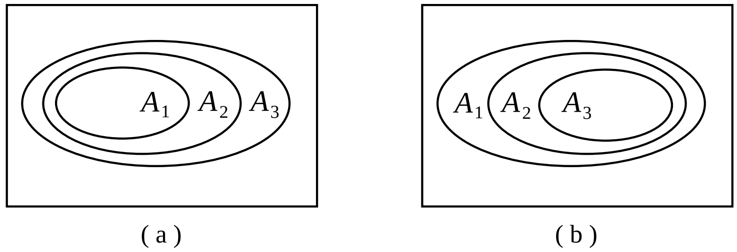


Figure 1.9. (a) For increasing events $A_1 \subset A_2 \subset A_3$, the union $A_1 \cup A_2 \cup A_3 = A_3$. (b) For decreasing events $A_1 \supset A_2 \supset A_3$, the intersection $A_1 \cap A_2 \cap A_3 = A_3$.

Formulas (1.12) and (1.13) together imply that for any sequence of events A_n,

$$\boxed{\wp\biggl(\bigcup_{n=1}^{\infty} A_n\biggr) \leq \sum_{n=1}^{\infty} \wp(A_n).} \tag{1.17}$$

This formula is known as the **union bound** in engineering and as **countable subadditivity** in mathematics. It is derived in Problems 29 and 30 at the end of the chapter.

1.4. Conditional Probability

Suppose we want to find out if there is a relationship between cancer and smoking. We survey n people and find out if they have cancer and if they have smoked. For each person, there are four possible outcomes, depending on whether the person has cancer (c) or does not have cancer (nc) and whether the person is a smoker (s) or a nonsmoker (ns). We denote these pairs of outcomes by $O_{\text{c,s}}$, $O_{\text{c,ns}}$, $O_{\text{nc,s}}$, and $O_{\text{nc,ns}}$. The numbers of each outcome can be arranged in the matrix

$$\begin{bmatrix} N(O_{\text{c,s}}) & N(O_{\text{c,ns}}) \\ N(O_{\text{nc,s}}) & N(O_{\text{nc,ns}}) \end{bmatrix}. \tag{1.18}$$

The sum of the first column is the number of smokers, which we denote by $N(O_{\text{s}})$. The sum of the second column is the number of nonsmokers, which we denote by $N(O_{\text{ns}})$.

The relative frequency of smokers who have cancer is $N(O_{\text{c,s}})/N(O_{\text{s}})$, and the relative frequency of nonsmokers who have cancer is $N(O_{\text{c,ns}})/N(O_{\text{ns}})$. If $N(O_{\text{c,s}})/N(O_{\text{s}})$ is substantially greater than $N(O_{\text{c,ns}})/N(O_{\text{ns}})$, we would conclude that smoking increases the occurrence of cancer.

Notice that the relative frequency of smokers who have cancer can also be written as the quotient of relative frequencies,

$$\frac{N(O_{\text{c,s}})}{N(O_{\text{s}})} = \frac{N(O_{\text{c,s}})/n}{N(O_{\text{s}})/n}. \tag{1.19}$$

This suggests the following definition of conditional probability. Let Ω be a sample space. Let the event S model a person's being a smoker, and let the event

C model a person's having cancer. In our model, the **conditional probability** that a person has cancer given that the person is a smoker is defined by

$$\boxed{\wp(C|S) \;:=\; \frac{\wp(C \cap S)}{\wp(S)},}$$

where the probabilities model the relative frequencies on the right-hand side of (1.19). This definition makes sense only if $\wp(S) > 0$. If $\wp(S) = 0$, $\wp(C|S)$ is not defined.

Given any two events A and B of positive probability,

$$\wp(A|B) \;=\; \frac{\wp(A \cap B)}{\wp(B)} \tag{1.20}$$

and

$$\wp(B|A) \;=\; \frac{\wp(A \cap B)}{\wp(A)}.$$

From (1.20), we see that

$$\boxed{\wp(A \cap B) \;=\; \wp(A|B)\,\wp(B).} \tag{1.21}$$

Substituting this into the numerator above yields

$$\wp(B|A) \;=\; \frac{\wp(A|B)\,\wp(B)}{\wp(A)}. \tag{1.22}$$

We next turn to the problem of computing the denominator $\wp(A)$.

The Law of Total Probability and Bayes' Rule

From the identity (recall Figure 1.7(a))

$$A \;=\; (A \cap B) \cup (A \cap B^c),$$

it follows that

$$\wp(A) \;=\; \wp(A \cap B) + \wp(A \cap B^c).$$

Using (1.21), we have

$$\boxed{\wp(A) \;=\; \wp(A|B)\,\wp(B) + \wp(A|B^c)\,\wp(B^c).} \tag{1.23}$$

This formula is the simplest version of the **law of total probability**. Substituting (1.23) into (1.22), we obtain

$$\wp(B|A) \;=\; \frac{\wp(A|B)\,\wp(B)}{\wp(A|B)\,\wp(B) + \wp(A|B^c)\,\wp(B^c)}. \tag{1.24}$$

This formula is the simplest version of **Bayes' rule**.

As illustrated in the following example, it is not necessary to remember Bayes' rule as long as you know the definition of conditional probability and the law of total probability.

***Example* 1.19.** Polychlorinated biphenyls (PCBs) are toxic chemicals. Given that you are exposed to PCBs, suppose that your conditional probability of developing cancer is 1/3. Given that you are *not* exposed to PCBs, suppose that your conditional probability of developing cancer is 1/4. Suppose that the probability of being exposed to PCBs is 3/4. Find the probability that you were exposed to PCBs given that you do not develop cancer.

Solution. To solve this problem, we use the notation**

$$E = \{\text{exposed to PCBs}\} \quad \text{and} \quad C = \{\text{develop cancer}\}.$$

With this notation, it is easy to interpret the problem as telling us that

$$\wp(C|E) = 1/3, \quad \wp(C|E^c) = 1/4, \quad \text{and} \quad \wp(E) = 3/4, \tag{1.25}$$

and asking us to find $\wp(E|C^c)$. Before solving the problem, note that the above data implies three additional equations as follows. First, recall that $\wp(E^c) = 1 - \wp(E)$. Similarly, since conditional probability is a probability as a function of its first argument, we can write $\wp(C^c|E) = 1 - \wp(C|E)$ and $\wp(C^c|E^c) = 1 - \wp(C|E^c)$. Hence,

$$\wp(C^c|E) = 2/3, \quad \wp(C^c|E^c) = 3/4, \quad \text{and} \quad \wp(E^c) = 1/4. \tag{1.26}$$

To find the desired conditional probability, we write

$$\begin{aligned}
\wp(E|C^c) &= \frac{\wp(E \cap C^c)}{\wp(C^c)} \\
&= \frac{\wp(C^c|E)\wp(E)}{\wp(C^c)} \\
&= \frac{(2/3)(3/4)}{\wp(C^c)} \\
&= \frac{1/2}{\wp(C^c)}.
\end{aligned}$$

To find the denominator, we use the law of total probability to write

$$\begin{aligned}
\wp(C^c) &= \wp(C^c|E)\wp(E) + \wp(C^c|E^c)\wp(E^c) \\
&= (2/3)(3/4) + (3/4)(1/4) = 11/16.
\end{aligned}$$

**In working this example, we follow common practice and do not explicitly specify the sample space Ω or the probability measure $\wp$. Hence, the expression "let $E = \{\text{exposed to PCBs}\}$" is shorthand for "let E be the subset of Ω that models being exposed to PCBs." The curious reader may find one possible choice for Ω and $\wp$, along with precise mathematical definitions of the events E and C, in Note 3.

Hence,

$$\wp(E|C^c) = \frac{1/2}{11/16} = \frac{8}{11}.$$

We now generalize the law of total probability. Let B_n be a sequence of pairwise disjoint events such that $\sum_n \wp(B_n) = 1$. Then for any event A,

$$\wp(A) = \sum_n \wp(A|B_n)\wp(B_n).$$

To derive this result, put $B := \bigcup_n B_n$, and observe that[††]

$$\wp(B) = \sum_n \wp(B_n) = 1.$$

It follows that $\wp(B^c) = 1 \Leftrightarrow \wp(B) = 0$. Next, for any event A, $A \cap B^c \subset B^c$, and so

$$0 \le \wp(A \cap B^c) \le \wp(B^c) = 0.$$

Hence, $\wp(A \cap B^c) = 0$. Writing (recall Figure 1.7(a))

$$A = (A \cap B) \cup (A \cap B^c),$$

it follows that

$$\begin{aligned} \wp(A) &= \wp(A \cap B) + \wp(A \cap B^c) \\ &= \wp(A \cap B) \\ &= \wp\left(A \cap \left[\bigcup_n B_n\right]\right) \\ &= \wp\left(\bigcup_n [A \cap B_n]\right) \\ &= \sum_n \wp(A \cap B_n). \end{aligned} \tag{1.27}$$

To compute $\wp(B_k|A)$, write

$$\wp(B_k|A) = \frac{\wp(A \cap B_k)}{\wp(A)} = \frac{\wp(A|B_k)\wp(B_k)}{\wp(A)}.$$

Applying the law of total probability to $\wp(A)$ in the denominator yields the general form of Bayes' rule,

$$\wp(B_k|A) = \frac{\wp(A|B_k)\wp(B_k)}{\sum_n \wp(A|B_n)\wp(B_n)}.$$

[††]Notice that since we do not require $\bigcup_n B_n = \Omega$, the B_n do not, strictly speaking, form a partition. However, since $\wp(B) = 1$, the remainder set (cf. (1.8)), which in this case is B^c, has probability zero.

In formulas like this, A is an event that we observe, while the B_n are events that we cannot observe but would like to make some inference about. Before making any observations, we know the **prior probabilities** $\wp(B_n)$, and we know the conditional probabilities $\wp(A|B_n)$. After we observe A, we compute the **posterior probabilities** $\wp(B_k|A)$ for each k.

***Example* 1.20.** In Example 1.19, before we learn any information about a person, that person's prior probability of being exposed to PCBs is $\wp(E) = 3/4 = 0.75$. After we observe that the person does not develop cancer, the posterior probability that the person was exposed to PCBs is $\wp(E|C^c) = 8/11 \approx 0.73$, which is lower than the prior probability.

1.5. Independence

In the previous section, we discussed how we might determine if there is a relationship between cancer and smoking. We said that if the relative frequency of smokers who have cancer is substantially larger than the relative frequency of nonsmokers who have cancer, we would conclude that smoking affects the occurrence of cancer. On the other hand, if the relative frequencies of cancer among smokers and nonsmokers are about the same, we would say that the occurrence of cancer does not depend on whether or not a person smokes.

In probability theory, if events A and B satisfy $\wp(A|B) = \wp(A|B^c)$, we say A does not depend on B. This condition says that

$$\frac{\wp(A \cap B)}{\wp(B)} = \frac{\wp(A \cap B^c)}{\wp(B^c)}. \tag{1.28}$$

Applying the formulas $\wp(B^c) = 1 \Leftrightarrow \wp(B)$ and

$$\wp(A) = \wp(A \cap B) + \wp(A \cap B^c)$$

to the right-hand side yields

$$\frac{\wp(A \cap B)}{\wp(B)} = \frac{\wp(A) \Leftrightarrow \wp(A \cap B)}{1 \Leftrightarrow \wp(B)}.$$

Cross multiplying to eliminate the denominators gives

$$\wp(A \cap B)[1 \Leftrightarrow \wp(B)] = \wp(B)[\wp(A) \Leftrightarrow \wp(A \cap B)].$$

Subtracting common terms from both sides shows that $\wp(A \cap B) = \wp(A)\,\wp(B)$. Since this sequence of calculations is reversible, and since the condition $\wp(A \cap B) = \wp(A)\,\wp(B)$ is symmetric in A and B, it follows that A does not depend on B if and only if B does not depend on A.

When events A and B satisfy

$$\wp(A \cap B) = \wp(A)\,\wp(B), \tag{1.29}$$

we say they are **statistically independent**, or just **independent**.

Caution: The reader is warned to make sure he or she understands the difference between disjoint sets and independent events. Recall that A and B are disjoint if $A \cap B = \varnothing$. This concept does not involve $\wp$ in any way; to determine if A and B are disjoint requires only knowledge of A and B themselves. On the other hand, (1.29) implies that independence *does* depend on $\wp$ and not just on A and B. To determine if A and B are independent requires not only knowledge of A and B, but also knowledge of $\wp$. See Problem 48.

In arriving at (1.29) as the definition of independent events, we noted that (1.29) is equivalent to (1.28). Hence, if A and B are independent, $\wp(A|B) = \wp(A|B^c)$. What is this common value? Write

$$\wp(A|B) = \frac{\wp(A \cap B)}{\wp(B)} = \frac{\wp(A)\,\wp(B)}{\wp(B)} = \wp(A).$$

We now make some further observations about independence. First, it is a simple exercise to show that if A and B are independent events, then so are A and B^c, A^c and B, and A^c and B^c. For example, writing

$$\begin{aligned} \wp(A) &= \wp(A \cap B) + \wp(A \cap B^c) \\ &= \wp(A)\,\wp(B) + \wp(A \cap B^c), \end{aligned}$$

we have

$$\begin{aligned} \wp(A \cap B^c) &= \wp(A) - \wp(A)\,\wp(B) \\ &= \wp(A)[1 - \wp(B)] \\ &= \wp(A)\,\wp(B^c). \end{aligned}$$

By interchanging the roles of A and A^c and/or B and B^c, it follows that if any one of the four pairs is independent, then so are the other three.

***Example* 1.21.** An Internet packet travels from its source to router 1, from router 1 to router 2, and from router 2 to its destination. If routers drop packets independently with probability p, what is the probability that a packet is successfully transmitted from its source to its destination?

Solution. For $i = 1, 2$, let D_i denote the event that the packet is dropped by router i. Let S denote the event that the packet is successfully transmitted from the source to the destination. Observe that S occurs if and only if the packet is not dropped by router 1 and it is not dropped by router 2. We can write this symbolically as

$$S = D_1^c \cap D_2^c.$$

Since the problem tells us that D_1 and D_2 are independent events, so are D_1^c and D_2^c. Hence,

$$\begin{aligned} \wp(S) &= \wp(D_1^c \cap D_2^c) \\ &= \wp(D_1^c)\,\wp(D_2^c) \\ &= [1 - \wp(D_1)]\,[1 - \wp(D_2)] \\ &= (1 - p)^2. \end{aligned}$$

Now suppose that A and B are any two events. If $\wp(B) = 0$, then we claim that A and B are independent. We must show that

$$\wp(A \cap B) \;=\; \wp(A)\,\wp(B) \;=\; 0.$$

To show that the left-hand side is zero, observe that since probabilities are nonnegative, and since $A \cap B \subset B$,

$$0 \;\le\; \wp(A \cap B) \;\le\; \wp(B) \;=\; 0. \tag{1.30}$$

We now show that if $\wp(B) = 1$, then A and B are independent. Since $\wp(B) = 1$, $\wp(B^c) = 1 - \wp(B) = 0$, and it follows that A and B^c are independent. But then so are A and B.

Independence for More Than Two Events

Suppose that for $j = 1, 2, \ldots$, A_j is an event. When we say that the A_j are independent, we certainly want that for any $i \neq j$,

$$\wp(A_i \cap A_j) \;=\; \wp(A_i)\,\wp(A_j).$$

And for any distinct i, j, k, we want

$$\wp(A_i \cap A_j \cap A_k) \;=\; \wp(A_i)\,\wp(A_j)\,\wp(A_k).$$

We want analogous equations to hold for any four events, five events, and so on. In general, we want that for every *finite* subset J containing two or more positive integers,

$$\wp\biggl(\bigcap_{j \in J} A_j\biggr) \;=\; \prod_{j \in J} \wp(A_j).$$

In other words, we want the probability of every intersection involving finitely many of the A_j to be equal to the product of the probabilities of the individual events. If the above equation holds for all *finite* subsets of two or more positive integers, then we say that the A_j are **mutually independent**, or just independent. If the above equation holds for all subsets J containing exactly two positive integers but not necessarily for all finite subsets of 3 or more positive integers, we say that the A_j are **pairwise independent**.

***Example* 1.22.** Given three events, say A, B, and C, they are mutually independent if and only if the following equations *all* hold,

$$\begin{aligned}
\wp(A \cap B \cap C) &= \wp(A)\,\wp(B)\,\wp(C) \\
\wp(A \cap B) &= \wp(A)\,\wp(B) \\
\wp(A \cap C) &= \wp(A)\,\wp(C) \\
\wp(B \cap C) &= \wp(B)\,\wp(C).
\end{aligned}$$

It is possible to construct events A, B, and C such that the last three equations hold (pairwise independence), but the first one does not.[4] It is also possible for the first equation to hold while the last three fail.[5]

***Example* 1.23.** A coin is tossed three times and the number of heads is noted. Find the probability that the number of heads is two, assuming the tosses are mutually independent and that on each toss the probability of heads is λ for some fixed $0 \leq \lambda \leq 1$.

Solution. To solve the problem, let H_i denote the event that the ith toss is heads (so $\wp(H_i) = \lambda$), and let S_2 denote the event that the number of heads in three tosses is 2.* Then

$$S_2 = (H_1 \cap H_2 \cap H_3^c) \cup (H_1 \cap H_2^c \cap H_3) \cup (H_1^c \cap H_2 \cap H_3).$$

This is a disjoint union, and so $\wp(S_2)$ is equal to

$$\wp(H_1 \cap H_2 \cap H_3^c) + \wp(H_1 \cap H_2^c \cap H_3) + \wp(H_1^c \cap H_2 \cap H_3). \tag{1.31}$$

Next, since H_1, H_2, and H_3 are mutually independent, so are $(H_1 \cap H_2)$ and H_3. Hence, $(H_1 \cap H_2)$ and H_3^c are also independent. Thus,

$$\begin{aligned} \wp(H_1 \cap H_2 \cap H_3^c) &= \wp(H_1 \cap H_2)\,\wp(H_3^c) \\ &= \wp(H_1)\,\wp(H_2)\,\wp(H_3^c) \\ &= \lambda^2(1 \Leftrightarrow \lambda). \end{aligned}$$

Treating the last two terms in (1.31) similarly, we have $\wp(S_2) = 3\lambda^2(1 \Leftrightarrow \lambda)$. If the coin is fair, i.e., $\lambda = 1/2$, then $\wp(S_2) = 3/8$.

***Example* 1.24.** If $A_1, A_2, \ldots$ are mutually independent, show that

$$\wp\biggl(\bigcap_{n=1}^{\infty} A_n\biggr) = \prod_{n=1}^{\infty} \wp(A_n).$$

Solution. Write

$$\begin{aligned} \wp\biggl(\bigcap_{n=1}^{\infty} A_n\biggr) &= \lim_{N\to\infty} \wp\biggl(\bigcap_{n=1}^{N} A_n\biggr), \quad \text{by limit property (1.14)}, \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} \wp(A_n), \quad \text{by independence}, \\ &= \prod_{n=1}^{\infty} \wp(A_n), \end{aligned}$$

*In working this example, we again do not explicitly specify the sample space Ω or the probability measure $\wp$. The interested reader can find one possible choice for Ω and $\wp$ in Note 6.

where the last step is just the definition of the infinite product.

***Example* 1.25.** Consider an infinite sequence of independent coin tosses. Assume that the probability of heads is $0 < p < 1$. What is the probability of seeing all heads? What is the probability of ever seeing heads?

Solution. We use the result of the preceding example as follows. Let Ω be a sample space equipped with a probability measure $\wp$ and events $A_n, n = 1, 2, \ldots$, with $\wp(A_n) = p$, where the A_n are mutually independent.[7] The event A_n corresponds to, or models, the outcome that the nth toss results in a heads. The outcome of seeing all heads corresponds to the event $\bigcap_{n=1}^{\infty} A_n$, and its probability is

$$\wp\Big(\bigcap_{n=1}^{\infty} A_n\Big) \;=\; \lim_{N\to\infty} \prod_{n=1}^{N} \wp(A_n) \;=\; \lim_{N\to\infty} p^N \;=\; 0.$$

The outcome of ever seeing heads corresponds to the event $A := \bigcup_{n=1}^{\infty} A_n$. Since $\wp(A) = 1 - \wp(A^c)$, it suffices to compute the probability of $A^c = \bigcap_{n=1}^{\infty} A_n^c$. Arguing exactly as above, we have

$$\wp\Big(\bigcap_{n=1}^{\infty} A_n^c\Big) \;=\; \lim_{N\to\infty} \prod_{n=1}^{N} \wp(A_n^c) \;=\; \lim_{N\to\infty} (1 - p)^N \;=\; 0.$$

Thus, $\wp(A) = 1 - 0 = 1$.

1.6. Combinatorics and Probability

Note to the Reader. The material in this section is not required anywhere else in the book, and can be covered at any time.

There are many probability problems, especially those concerned with gambling, that can ultimately be reduced to questions about cardinalities of various sets. We saw several examples in Section 1.2. Those examples were simple, and they were chosen so that it was easy to determine the cardinalities of the required sets. However, in more complicated problems, it is extremely helpful to have some systematic methods for finding cardinalities of sets. **Combinatorics** is the study of systematic counting methods, which we will be using to find the cardinalities of various sets that arise in probability. The four kinds of counting problems we discuss are

(*i*) ordered sampling with replacement;
(*ii*) ordered sampling without replacement;
(*iii*) unordered sampling without replacement;
(*iv*) unordered sampling with replacement.

Of these, the first two are rather straightforward, and the last two are somewhat complicated.

Ordered Sampling with Replacement

Before stating the problem, we begin with some examples to illustrate the concepts to be used.

***Example* 1.26.** Let A, B, and C be finite sets. How many triples are there of the form (a, b, c), where $a \in A$, $b \in B$, and $c \in C$?

Solution. Since there are $|A|$ choices for a, $|B|$ choices for b, and $|C|$ choices for c, the total number of triples is $|A| \cdot |B| \cdot |C|$.

Similar reasoning shows that for k finite sets $A_1, \ldots, A_k$, there are $|A_1| \cdots |A_k|$ k-tuples of the form $(a_1, \ldots, a_k)$ where each $a_i \in A_i$.

***Example* 1.27.** How many k-digit numbers are there?

Solution. Taking "digit" to mean "decimal digit," we put each $A_i := \{0, \ldots, 9\}$. Since each $|A_i| = 10$, there are 10^k k-digit numbers.

As the previous example shows, it is often convenient to take all of the A_i to be the same set.

***Example* 1.28** (Ordered Sampling with Replacement). From a deck of n cards, we draw k cards with replacement; i.e., we draw each card, make a note of it, put the card back in the deck and re-shuffle the deck before choosing the next card. How many different sequences of k cards can be drawn in this way?

Solution. Although there is only one deck of cards, because we draw with replacement and re-shuffle, we may as well think of ourselves as having k decks of n cards each. Think of each $A_i := \{1, \ldots, n\}$. Then we need to know how many k-tuples there are of the form $(a_1, \ldots, a_k)$ where each $a_i \in A_i$. We see that the answer is n^k.

Ordered Sampling without Replacement

In Example 1.26, we formed triples (a, b, c) where no matter which $a \in A$ we chose, it did not affect which elements we were allowed to choose from the sets B or C. We next consider the construction of k-tuples in which our choice for the each entry affects the choices available for the remaining entries.

***Example* 1.29** (Ordered Sampling without Replacement). From a deck of n cards, we draw $k \le n$ cards without replacement. How many sequences can be drawn in this way?

Solution. There are n cards for the first draw, $n \Leftrightarrow 1$ cards for the second draw, and so on. Hence, there are

$$n(n \Leftrightarrow 1) \cdots (n \Leftrightarrow [k \Leftrightarrow 1]) \;=\; \frac{n!}{(n \Leftrightarrow k)!}$$

different sequences.

***Example* 1.30.** Let A be a finite set of n elements. How may k-tuples $(a_1, \ldots, a_k)$ of *distinct* entries $a_i \in A$ can be formed?

Solution. There are n choices for a_1, but only $n \Leftrightarrow 1$ choices for a_2 since repeated entries are not allowed. Similarly, there are only $n \Leftrightarrow 2$ choices for a_3, and so on. This is the same argument used in the previous example. Hence, there are $n!/(n \Leftrightarrow k)!$ k-tuples with distinct elements of A.

Given a set A, we let A^k denote the set of all k-tuples $(a_1, \ldots, a_k)$ where each $a_i \in A$. We denote by A_*^k the subset of all k-tuples with *distinct* entries. If $|A| = n$, then $|A^k| = |A|^k = n^k$, and $|A_*^k| = n!/(n \Leftrightarrow k)!$.

***Example* 1.31** (The Birthday Problem). In a group of k people, what is the probability that two or more people have the same birthday?

Solution. The first step in the solution is to specify the sample space Ω and the probability $\wp$. Let $D := \{1, \ldots, 365\}$ denote the days of the year, and let

$$\Omega \;:=\; \{(d_1, \ldots, d_k) : d_i \in D\}$$

denote the set of all possible sequences of k birthdays. Then $|\Omega| = |D|^k$. Assuming all sequences are equally likely, we take $\wp(E) := |E|/|\Omega|$ for arbitrary events $E \subset \Omega$.

Let Q denote the set of sequences $(d_1, \ldots, d_k)$ that have at least one pair of repeated entries. For example, if $k = 9$, one of the sequences in Q would be

$$(364, 17, 201, 17, 51, 171, 51, 33, 51).$$

Notice that 17 appears twice and 51 appears 3 times. The set Q is very complicated. On the other hand, consider Q^c, which is the set of sequences $(d_1, \ldots, d_k)$ that have *no* repeated entries. Then

$$|Q^c| \;=\; \frac{|D|!}{(|D| \Leftrightarrow k)!},$$

and

$$\wp(Q^c) \;=\; \frac{|Q^c|}{|\Omega|} \;=\; \frac{|D|!}{|D|^k\,(|D| \Leftrightarrow k)!},$$

where $|D| = 365$. A plot of $\wp(Q) = 1 \Leftrightarrow \wp(Q^c)$ as a function of k is shown in Figure 1.10. As the dashed line indicates, for $k \geq 23$, the probability of two more more people having the same birthday is greater than $1/2$.

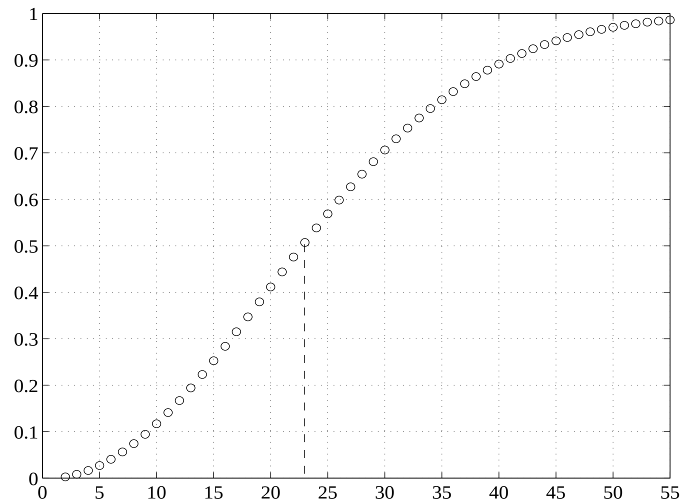


Figure 1.10. A plot of $\wp(Q)$ as a function of k. For $k \geq 23$, the probability of two or more people having the same birthday is greater than $1/2$.

Unordered Sampling without Replacement

Before stating the problem, we begin with a simple example to illustrate the concept to be used.

***Example* 1.32.** Let $A = \{1,2,3,4,5\}$. Then A^3 contains $5^3 = 125$ triples. The set of triples with distinct entries, A_*^3, contains $5!/2! = 60$ triples. We can write A_*^3 as the disjoint union

$$\begin{aligned} A_*^3 &= G_{123} \cup G_{124} \cup G_{125} \cup G_{134} \cup G_{135} \\ &\quad \cup G_{145} \cup G_{234} \cup G_{235} \cup G_{245} \cup G_{345}, \end{aligned}$$

where for distinct i, j, k,

$$G_{ijk} := \{(i,j,k),(i,k,j),(j,i,k),(j,k,i),(k,i,j),(k,j,i)\}.$$

Each triple in G_{ijk} is a rearrangement, or **permutation**, of the same 3 elements.

The above decomposition works in general. Write A_*^k as the union of disjoint sets,

$$A_*^k = \bigcup G, \tag{1.32}$$

where each subset G consists of k-tuples that contain the same elements. In general, for a k-tuple built from k distinct elements, there are k choices for the first entry, $k \Leftrightarrow 1$ choices for the second entry, and so on. Hence, there are $k!$

k-tuples that can be built. In other words, each G in (1.32) has $|G| = k!$. It follows from (1.32) that

$$|A_*^k| \;=\; (\text{number of different sets } G) \cdot k!, \tag{1.33}$$

and so the number of different subsets G is

$$\frac{|A_*^k|}{k!} \;=\; \frac{n!}{k!(n \Leftrightarrow k)!}.$$

The standard notation for the above right-hand side is

$$\binom{n}{k} \;:=\; \frac{n!}{k!(n \Leftrightarrow k)!}$$

and is read "n choose k." In MATLAB, $\binom{n}{k}$ = `nchoosek(n,k)`. The symbol $\binom{n}{k}$ is also called the **binomial coefficient** because it arises in the **binomial theorem**, which is discussed in Chapter 2.

***Example* 1.33** (Unordered Sampling without Replacement). In many card games, we are dealt a hand of k cards, but the order in which the cards are dealt is not important. From a deck of n cards, how many k-card hands are possible?

Solution. First think about ordered hands corresponding to k-tuples with distinct entries. The set of all such hands corresponds to A_*^k. Now group together k-tuples composed of the same elements into sets G as in (1.32). All the ordered k-tuples in a particular G represent rearrangements of a single hand. So it is really the number of different sets G that corresponds to the number of unordered hands. Thus, the number of k-card hands is $\binom{n}{k}$.

***Example* 1.34.** A 12-person jury is to be selected from a group of 20 potential jurors. How many different juries are possible?

Solution. There are

$$\binom{20}{12} \;=\; \frac{20!}{12!\,8!} \;=\; 125970$$

different juries.

***Example* 1.35.** A 12-person jury is to be selected from a group of 20 potential jurors of which 11 are men and 9 are women. How many 12-person juries are there with 5 men and 7 women?

Solution. There are $\binom{11}{5}$ ways to choose the 5 men, and there are $\binom{9}{7}$ ways to choose the 7 women. Hence, there are

$$\binom{11}{5}\binom{9}{7} \;=\; \frac{11!}{5!\,6!} \cdot \frac{9!}{7!\,2!} \;=\; 16632$$

possible juries with 5 men and 7 women.

***Example* 1.36.** An urn contains 11 green balls and 9 red balls. If 12 balls are chosen at random, what is the probability of choosing exactly 5 green balls and 7 red balls?

Solution. Since balls are chosen at random, the desired probability is

$$\frac{\text{number of ways to choose 5 green balls and 7 red balls}}{\text{number of ways to choose 12 balls}}.$$

In the numerator, the 5 green balls must be chosen from the 11 available green balls, and the 7 red balls must be chosen from the 9 available red balls. In the denominator, the total of $5+7=12$ balls must be chosen from the $11+9=20$ available balls. So the required probability is

$$\frac{\binom{11}{5}\binom{9}{7}}{\binom{20}{12}} = \frac{16632}{125970} \approx 0.132.$$

***Example* 1.37.** Consider a collection of N items, of which d are defective (and $N \Leftrightarrow d$ work properly). Suppose we test $n \leq N$ items at random. Show that the probability that k of the n tested items are defective is

$$\frac{\binom{d}{k}\binom{N \Leftrightarrow d}{n \Leftrightarrow k}}{\binom{N}{n}}. \tag{1.34}$$

Solution. Since items are chosen at random, the desired probability is

$$\frac{\text{number of ways to choose } k \text{ defective and } n \Leftrightarrow k \text{ working items}}{\text{number of ways to choose } n \text{ items}}.$$

In the numerator, the k defective items are chosen from the total of d defective ones, and the $n \Leftrightarrow k$ working items are chosen from the total of $N \Leftrightarrow d$ ones that work. In the denominator, the n items to be tested are chosen from the total of N items. Hence, the desired numerator is $\binom{d}{k}\binom{N-d}{n-k}$, and the desired denominator is $\binom{N}{n}$.

***Example* 1.38** (Lottery). In some state lottery games, a player chooses n distinct numbers from the set $\{1, \ldots, N\}$. At the lottery drawing, balls numbered from 1 to N are mixed, and n balls withdrawn. What is the probability that k of the n balls drawn match the player's choices?

Solution. Let D denote the subset of n numbers chosen by the player. Then $\{1, \ldots, N\} = D \cup D^c$. We need to find the probability that the lottery drawing chooses k numbers from D and $n \Leftrightarrow k$ numbers from D^c. Since $|D| = n$, this probability is

$$\frac{\binom{n}{k}\binom{N \Leftrightarrow n}{n \Leftrightarrow k}}{\binom{N}{n}}.$$

Notice that this is just (1.34) with $d = n$. In other words, we regard the numbers chosen by the player as "defective," and we are finding the probability that the lottery drawing chooses k defective and $n \Leftrightarrow k$ non-defective numbers.

***Example* 1.39** (Binomial Probabilities). A certain coin has probability p of turning up heads. If the coin is tossed n times, what is the probability that k of the n tosses result in heads? Assume tosses are independent.

Solution. Let H_i denote the event that the ith toss is heads. We call i the toss index, which takes values $1, \ldots, n$. A typical sequence of n tosses would be

$$H_1 \cap H_2^c \cap H_3 \cap \cdots \cap H_{n-1} \cap H_n^c,$$

where H_i^c is the event that the ith toss is tails. The probability that n tosses result in k heads and $n \Leftrightarrow k$ tails is

$$\wp\Big(\bigcup \widetilde{H}_1 \cap \cdots \cap \widetilde{H}_n\Big),$$

where $\widetilde{H}_i$ is either H_i or H_i^c, and the union is over all such intersections for which $\widetilde{H}_i = H_i$ occurs k times and $\widetilde{H}_i = H_i^c$ occurs $n \Leftrightarrow k$ times. Since this is a disjoint union,

$$\wp\Big(\bigcup \widetilde{H}_1 \cap \cdots \cap \widetilde{H}_n\Big) = \sum \wp(\widetilde{H}_1 \cap \cdots \cap \widetilde{H}_n).$$

By independence,

$$\begin{aligned} \wp(\widetilde{H}_1 \cap \cdots \cap \widetilde{H}_n) &= \wp(\widetilde{H}_1) \cdots \wp(\widetilde{H}_n) \\ &= p^k (1 \Leftrightarrow p)^{n-k} \end{aligned}$$

is the same for every term in the sum. The number of terms in the sum is the number of ways of selecting k out of n toss indexes to assign to heads. Since this number is $\binom{n}{k}$, the probability that k of n tosses result in heads is

$$\binom{n}{k} p^k (1 \Leftrightarrow p)^{n-k}.$$

***Example* 1.40** (Bridge). In bridge, 52 cards are dealt to 4 players; hence, each player has 13 cards. The order in which the cards are dealt is not important, just the final 13 cards each player ends up with. How many different bridge games can be dealt?

Solution. There are $\binom{52}{13}$ ways to choose the 13 cards of the first player. Now there are only $52-13 = 39$ cards left. Hence, there are $\binom{39}{13}$ ways to choose the 13 cards for the second player. Similarly, there are $\binom{26}{13}$ ways to choose the second player's cards, and $\binom{13}{13} = 1$ way to choose the fourth player's cards. It follows that there are

$$\begin{aligned}\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} &= \frac{52!}{13!\,39!}\cdot\frac{39!}{13!\,26!}\cdot\frac{26!}{13!\,13!}\cdot\frac{13!}{13!\,0!} \\ &= \frac{52!}{(13!)^4} \approx 5.36\times 10^{28}\end{aligned}$$

games that can be dealt.

***Example* 1.41.** Traditionally, computers use binary arithmetic, and store n-bit words composed of zeros and ones. The new m–Computer uses m-ary arithmetic, and stores n-symbol words in which the symbols (m-ary digits) come from the set $\{0, 1, \ldots, m-1\}$. How many n-symbol words are there with k_0 zeros, k_1 ones, k_2 twos, ..., and k_{m-1} copies of symbol $m-1$, where $k_0 + k_1 + k_2 + \cdots + k_{m-1} = n$?

Solution. To answer this question, we build a typical n-symbol word of the required form as follows. We begin with an empty word,

$$\underbrace{(\ ,\ ,\ldots,\ )}_{n\text{ empty positions}}\ .$$

From these n available positions, there are $\binom{n}{k_0}$ ways to select positions to put the k_0 zeros. For example, if $k_0 = 3$, we might have

$$\underbrace{(\ ,0,\ ,0,\ ,\ldots,\ ,0)}_{n-3\text{ empty positions}}.$$

Now there are only $n-k_0$ empty positions. From these, there are $\binom{n-k_0}{k_1}$ ways to select positions to put the k_1 ones. For example, if $k_1 = 1$, we might have

$$\underbrace{(\ ,0,1,0,\ ,\ldots,\ ,0)}_{n-4\text{ empty positions}}.$$

Now there are only $n-k_0-k_1$ empty positions. From these, there are $\binom{n-k_0-k_1}{k_2}$ ways to select positions to put the k_2 twos. Continuing in this way, we find that the number of n-symbol words with the required numbers of zeros, ones, twos, etc., is

$$\binom{n}{k_0}\binom{n-k_0}{k_1}\binom{n-k_0-k_1}{k_2}\cdots\binom{n-k_0-k_1-\cdots-k_{m-2}}{k_{m-1}},$$

which expands to

$$\frac{n!}{k_0!\,(n-k_0)!}\cdot\frac{(n-k_0)!}{k_1!\,(n-k_0-k_1)!}\cdot\frac{(n-k_0-k_1)!}{k_2!\,(n-k_0-k_1-k_2)!}\cdots$$
$$\cdots\frac{(n-k_0-k_1-\cdots-k_{m-2})!}{k_{m-1}!\,(n-k_0-k_1-\cdots-k_{m-1})!}.$$

Canceling common factors and noting that $(n-k_0-k_1-\cdots-k_{m-1})! = 0! = 1$, we obtain

$$\frac{n!}{k_0!\,k_1!\cdots k_{m-1}!}$$

as the number of n-symbol words with k_0 zeros, k_1 ones, etc.

We call

$$\binom{n}{n_0,\ldots,n_{m-1}} := \frac{n!}{k_0!\,k_1!\cdots k_{m-1}!}$$

the **multinomial coefficient**. When $m = 2$,

$$\binom{n}{k_0,k_1} = \binom{n}{k_0,n-k_0} = \frac{n!}{k_0!\,(n-k_0)!} = \binom{n}{k_0}$$

becomes the binomial coefficient.

Unordered Sampling with Replacement

Before stating the problem, we begin with a simple example to illustrate the concepts involved.

***Example* 1.42.** An automated snack machine dispenses apples, bananas, and carrots. For a fixed price, the customer gets 5 items from among the 3 possible choices. For example, a customer could choose 1 apple, 2 bananas, and 2 carrots. To record the customer's choices electronically, 7-bit sequences are used. For example, $(0,1,0,0,1,0,0)$ means 1 apple, 2 bananas, and 2 carrots. The first group of zeros tells how many apples, the second group of zeros tells how many bananas, and the third group of zeros tells how many carrots. The ones are used to separate the groups of zeros. As another example, $(0,0,0,1,0,1,0)$ means 3 apples, 1 banana, and 1 carrot. How many customer choices are there?

Solution. The question is equivalent to asking how many 7-bit sequences there are with 5 zeros and 2 ones. From Example 1.41, the answer is $\binom{7}{5,2} = \binom{7}{5} = \binom{7}{2}$.

***Example* 1.43** (Unordered Sampling with Replacement). From the set $A = \{1, 2, \ldots, n\}$, k numbers are drawn with replacement. How many different sets of k numbers can be obtained in this way?

Solution. We identify the sets drawn in this way with binary sequences with k zeros and $n-1$ ones, where the ones break up the groups of zeros as in the previous example. Hence, there are $\binom{k+n-1}{k} = \binom{k+n-1}{n-1}$ ways to draw k items with replacement from n items.

Just as we partitioned A^k_* in (1.32), we can partition A^k using

$$A^k = \bigcup G,$$

where each G contains all k-tuples with the same elements. Unfortunately, different Gs may contain different numbers of k-tuples. For example, if $n = 3$ and $k = 3$, one of the sets G would be

$$\{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\},$$

while another G would be

$$\{(1,2,2),(2,1,2),(2,2,1)\}.$$

How many different sets G are there? Although we cannot find the answer by using an equation like (1.33), we see from the above analysis that there are $\binom{k+n-1}{k}$ sets G.

Notes

§*1.3: Axioms and Properties of Probability*

***Note* 1.** When the sample space Ω is finite or countably infinite, $\wp(A)$ is usually defined for all subsets of Ω by taking

$$\wp(A) := \sum_{\omega \in A} p(\omega)$$

for some nonnegative function p that sums to one; i.e., $p(\omega) \geq 0$ and $\sum_{\omega\in\Omega} p(\omega) = 1$. (It is easy to check that if $\wp$ is defined in this way, then it satisfies the axioms of a probability measure.) However, for larger sample spaces, such as when Ω is an interval of the real line, e.g., Example 1.16, and we want the probability of an interval to be proportional to its length, it is not possible to define $\wp(A)$ for all subsets and still have $\wp$ satisfy all four axioms. (A proof of this fact can be found in advanced texts, e.g., [4, p. 45].) The way around this difficulty is to define $\wp(A)$ only for some subsets of Ω, but not all subsets of Ω. It is indeed fortunate that this can be done in such a way that $\wp(A)$ is defined for all subsets of interest that occur in practice. A set A for which $\wp(A)$

is defined is called an **event**, and the collection of all events is denoted by $\mathcal{A}$. The triple $(\Omega, \mathcal{A}, \wp)$ is called a **probability space**.

Given that $\wp(A)$ is defined only for $A \in \mathcal{A}$, in order for the probability axioms stated in Section 1.3 to make sense, $\mathcal{A}$ must have certain properties. First, axiom (i) requires that $\varnothing \in \mathcal{A}$. Second, axiom (iv) requires that $\Omega \in \mathcal{A}$. Third, axiom (iii) requires that if $A_1, A_2, \ldots$ are mutually exclusive events, then their union, $\bigcup_{n=1}^\infty A_n$, is also an event. Additionally, we need that if $A_1, A_2, \ldots$ are arbitrary events, then so is their intersection, $\bigcap_{n=1}^\infty A_n$. We show below that these four requirements are satisfied if we assume only that $\mathcal{A}$ is a σ-field.

If $\mathcal{A}$ is a collection of subsets of Ω with the following properties, then $\mathcal{A}$ is called a ***σ*-field** or a ***σ*-algebra**.

(i) The empty set $\varnothing$ belongs to $\mathcal{A}$, i.e., $\varnothing \in \mathcal{A}$.

(ii) If $A \in \mathcal{A}$, then so does its complement, A^c, i.e., $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$.

(iii) If $A_1, A_2, \ldots$ belong to $\mathcal{A}$, then so does their union, $\bigcup_{n=1}^\infty A_n$.

If $\mathcal{A}$ is a σ-field, then $\varnothing$ and $\varnothing^c = \Omega$ belong to $\mathcal{A}$. Also, for any $A_n \in \mathcal{A}$, their union is in $\mathcal{A}$, and by De Morgan's law, so is their intersection.

Given any set Ω, let 2^Ω denote the collection of all subsets of Ω. We call 2^Ω the **power set** of Ω. This notation is used for both finite and infinite sets. The notation is motivated by the fact that if Ω is a finite set, then there are $2^{|\Omega|}$ different subsets of Ω.† Since the power set obviously satisfies the three properties above, the power set is a σ-field.

Let $\mathcal{C}$ be any collection of subsets of Ω. We do not assume $\mathcal{C}$ is a σ-field. Define $\sigma(\mathcal{C})$ to be the **smallest *σ*-field that contains $\mathcal{C}$**. By this we mean that if $\mathcal{D}$ is any σ-field with $\mathcal{C} \subset \mathcal{D}$, then $\sigma(\mathcal{C}) \subset \mathcal{D}$.

***Example* 1.44.** Let A be a nonempty subset of Ω, and put $\mathcal{C} = \{A\}$ so that the collection $\mathcal{C}$ consists of a single subset. Find $\sigma(\mathcal{C})$.

Solution. From the three properties of a σ-field, any σ-field that contains A must also contain A^c, $\varnothing$, and Ω. We claim

$$\sigma(\mathcal{C}) = \{\varnothing, A, A^c, \Omega\}.$$

†Suppose $\Omega = \{\omega_1, \ldots, \omega_n\}$. Each subset of Ω can be associated with an n-bit word. A point ω_i is in the subset if and only if the ith bit in the word is a 1. For example, if $n = 5$, we associate 01011 with the subset $\{\omega_2, \omega_4, \omega_5\}$ since bits 2, 4, and 5 are ones. In particular, 00000 corresponds to the empty set and 11111 corresponds to Ω itself. Since there are 2^n n-bit words, there are 2^n subsets of Ω.

Since $A \cup A^c = \Omega$, it is easy to see that our choice satisfies the three properties of a σ-field. It is also clear that if $\mathcal{D}$ is any σ-field such that $\mathcal{C} \subset \mathcal{D}$, then every subset in our choice for $\sigma(\mathcal{C})$ must belong to $\mathcal{D}$; i.e., $\sigma(\mathcal{C}) \subset \mathcal{D}$.

More generally, if $A_1, \ldots, A_n$ is a partition of Ω, then $\sigma(\{A_1, \ldots, A_n\})$ consists of the empty set along with the $2^n \Leftrightarrow 1$ subsets constructed by taking all possible unions of the A_i. See Problem 33.

For general collections $\mathcal{C}$ of subsets of Ω, all we can say is that (Problem 37)

$$\sigma(\mathcal{C}) \;=\; \bigcap_{\mathcal{C} \subset \mathcal{A}} \mathcal{A},$$

where the intersection is over all σ-fields $\mathcal{A}$ that contain $\mathcal{C}$. Note that there is always at least one σ-field $\mathcal{A}$ that contains $\mathcal{C}$; e.g., the power set.

***Note* 2.** In light of the preceding note, we see that to guarantee that $\wp(\{\omega_n\})$ is defined in Example 1.18, it is necessary to assume that the singleton sets $\{\omega_n\}$ are events, i.e., $\{\omega_n\} \in \mathcal{A}$.

§*1.4: Conditional Probability*

***Note* 3.** Here is a choice for Ω and $\wp$ for Example 1.19. Let

$$\Omega \;:=\; \{(e,c) : e, c = 0 \text{ or } 1\},$$

where $e = 1$ corresponds to exposure to PCBs, and $c = 1$ corresponds to developing cancer. We then take

$$E \;:=\; \{(e,c) : e = 1\} \;=\; \{\ (1,0)\ ,\ (1,1)\ \},$$

and

$$C \;:=\; \{(e,c) : c = 1\} \;=\; \{\ (0,1)\ ,\ (1,1)\ \}.$$

It follows that

$$E^c \;=\; \{\ (0,1)\ ,\ (0,0)\ \} \quad \text{and} \quad C^c \;=\; \{\ (1,0)\ ,\ (0,0)\ \}.$$

Hence, $E \cap C = \{(1,1)\}$, $E \cap C^c = \{(1,0)\}$, $E^c \cap C = \{(0,1)\}$, and $E^c \cap C^c = \{(0,0)\}$.

In order to specify a suitable probability measure on Ω, we work backwards. First, *if* a measure $\wp$ on Ω exists such that (1.25) and (1.26) hold, then

$$\begin{aligned}
\wp(\{(1,1)\}) &= \wp(E \cap C) = \wp(C|E)\wp(E) = 1/4, \\
\wp(\{(0,1)\}) &= \wp(E^c \cap C) = \wp(C|E^c)\wp(E^c) = 1/16, \\
\wp(\{(1,0)\}) &= \wp(E \cap C^c) = \wp(C^c|E)\wp(E) = 1/2, \\
\wp(\{(0,0)\}) &= \wp(E^c \cap C^c) = \wp(C^c|E^c)\wp(E^c) = 3/16.
\end{aligned}$$

This suggests that we *define* $\wp$ by

$$\wp(A) \; := \; \sum_{\omega \in A} p(\omega),$$

where $p(\omega) = p(e, c)$ is given by $p(1, 1) := 1/4$, $p(0, 1) := 1/16$, $p(1, 0) := 1/2$, and $p(0, 0) := 3/16$. Starting from this definition of $\wp$, it is not hard to check that (1.25) and (1.26) hold.

§*1.5: Independence*

***Note* 4.** Here is an example of three events that are pairwise independent, but not mutually independent. Let

$$\Omega \; := \; \{1, 2, 3, 4, 5, 6, 7\},$$

and put $\wp(\{\omega\}) := 1/8$ for $\omega \neq 7$, and $\wp(\{7\}) := 1/4$. Take $A := \{1, 2, 7\}$, $B := \{3, 4, 7\}$, and $C := \{5, 6, 7\}$. Then $\wp(A) = \wp(B) = \wp(C) = 1/2$. and $\wp(A \cap B) = \wp(A \cap C) = \wp(B \cap C) = \wp(\{7\}) = 1/4$. Hence, A and B, A and C, and B and C are pairwise independent. However, since $\wp(A \cap B \cap C) = \wp(\{7\}) = 1/4$, and since $\wp(A)\,\wp(B)\,\wp(C) = 1/8$, A, B, and C are not mutually independent.

***Note* 5.** Here is an example of three events for which $\wp(A \cap B \cap C) = \wp(A)\,\wp(B)\,\wp(C)$ but no pair is independent. Let $\Omega := \{1, 2, 3, 4\}$. Put $\wp(\{1\}) = \wp(\{2\}) = \wp(\{3\}) = p$ and $\wp(\{4\}) = q$, where $3p + q = 1$ and $0 \leq p, q \leq 1$. Put $A := \{1, 4\}$, $B := \{2, 4\}$, and $C := \{3, 4\}$. Then the intersection of any pair is $\{4\}$, as is the intersection of all three sets. Also, $\wp(\{4\}) = q$. Since $\wp(A) = \wp(B) = \wp(C) = p + q$, we require $(p + q)^3 = q$ and $(p + q)^2 \neq q$. Solving $3p + q = 1$ and $(p + q)^3 = q$ for q reduces to solving $8q^3 + 12q^2 \Leftrightarrow 21q + 1 = 0$. Now, $q = 1$ is obviously a root, but this results in $p = 0$, which implies mutual independence. However, since $q = 1$ is a root, it is easy to verify that

$$8q^3 + 12q^2 \Leftrightarrow 21q + 1 \; = \; (q \Leftrightarrow 1)(8q^2 + 20q \Leftrightarrow 1).$$

By the quadratic formula, the desired root is $q = \big(\Leftrightarrow 5 + 3\sqrt{3}\,\big)/4$. It then follows that $p = \big(3 \Leftrightarrow \sqrt{3}\,\big)/4$ and that $p + q = \big(\Leftrightarrow 1 + \sqrt{3}\,\big)/2$. Now just observe that $(p + q)^2 \neq q$.

***Note* 6.** Here is a choice for Ω and $\wp$ for Example 1.23. Let

$$\Omega \; := \; \{(i, j, k) : i, j, k = 0 \text{ or } 1\},$$

with 1 corresponding to heads and 0 to tails. Now put

$$\begin{aligned} H_1 &:= \{(i, j, k) : i = 1\}, \\ H_2 &:= \{(i, j, k) : j = 1\}, \\ H_3 &:= \{(i, j, k) : k = 1\}, \end{aligned}$$

and observe that

$$\begin{aligned}
H_1 &= \{\ (1,0,0)\ ,\ (1,0,1)\ ,\ (1,1,0)\ ,\ (1,1,1)\ \}, \\
H_2 &= \{\ (0,1,0)\ ,\ (0,1,1)\ ,\ (1,1,0)\ ,\ (1,1,1)\ \}, \\
H_3 &= \{\ (0,0,1)\ ,\ (0,1,1)\ ,\ (1,0,1)\ ,\ (1,1,1)\ \}.
\end{aligned}$$

Next, let $\wp(\{(i,j,k)\}) := \lambda^{i+j+k}(1 \Leftrightarrow \lambda)^{3-(i+j+k)}$. Since

$$H_3^c \ = \ \{\ (0,0,0)\ ,\ (1,0,0)\ ,\ (0,1,0)\ ,\ (1,1,0)\ \},$$

$H_1 \cap H_2 \cap H_3^c = \{(1,1,0)\}$. Similarly, $H_1 \cap H_2^c \cap H_3 = \{(1,0,1)\}$, and $H_1^c \cap H_2 \cap H_3 = \{(0,1,1)\}$. Hence,

$$\begin{aligned}
S_2 &= \{\ (1,1,0)\ ,\ (1,0,1)\ ,\ (0,1,1)\ \} \\
&= \{(1,1,0)\} \cup \{(1,0,1)\} \cup \{(0,1,1)\},
\end{aligned}$$

and thus, $\wp(S_2) = 3\lambda^2(1 \Leftrightarrow \lambda)$.

Note 7. To show the *existence* of a sample space and probability measure with such independent events is beyond the scope of this book. Such constructions can be found in more advanced texts such as [4, Section 36].

Problems

§*1.1: Review of Set Notation*

1. For real numbers $\Leftrightarrow\infty < a < b < \infty$, we use the following notation.

$$\begin{aligned}
(a,b] &:= \{x : a < x \leq b\} \\
(a,b) &:= \{x : a < x < b\} \\
[a,b) &:= \{x : a \leq x < b\} \\
[a,b] &:= \{x : a \leq x \leq b\}.
\end{aligned}$$

We also use

$$\begin{aligned}
(\Leftrightarrow\infty, b] &:= \{x : x \leq b\} \\
(\Leftrightarrow\infty, b) &:= \{x : x < b\} \\
(a, \infty) &:= \{x : x > a\} \\
[a, \infty) &:= \{x : x \geq a\}.
\end{aligned}$$

For example, with this notation, $(0,1]^c = (\Leftrightarrow\infty, 0] \cup (1,\infty)$ and $(0,2] \cup [1,3) = (0,3)$. Now analyze

(a) $[2,3]^c$,

(b) $(1,3) \cup (2,4)$,

(c) $(1,3) \cap [2,4)$,

(d) $(3,6] \setminus (5,7)$.

2. Sketch the following subsets of the x-y plane.

(a) $B_z := \{(x,y) : x + y \le z\}$ for $z = 0, -1, +1$.

(b) $C_z := \{(x,y) : x > 0,\ y > 0,\ \text{and } xy \le z\}$ for $z = 1$.

(c) $H_z := \{(x,y) : x \le z\}$ for $z = 3$.

(d) $J_z := \{(x,y) : y \le z\}$ for $z = 3$.

(e) $H_z \cap J_z$ for $z = 3$.

(f) $H_z \cup J_z$ for $z = 3$.

(g) $M_z := \{(x,y) : \max(x,y) \le z\}$ for $z = 3$, where $\max(x,y)$ is the larger of x and y. For example, $\max(7,9) = 9$. Of course, $\max(9,7) = 9$ too.

(h) $N_z := \{(x,y) : \min(x,y) \le z\}$ for $z = 3$, where $\min(x,y)$ is the smaller of x and y. For example, $\min(7,9) = 7 = \min(9,7)$.

(i) $M_2 \cap N_3$.

(j) $M_4 \cap N_3$.

3. Let Ω denote the set of real numbers, $\Omega = (-\infty, \infty)$.

(a) Use the distributive law to simplify

$$[1,4] \cap \Big([0,2] \cup [3,5]\Big).$$

(b) Use De Morgan's law to simplify $\Big([0,1] \cup [2,3]\Big)^c$.

(c) Simplify $\displaystyle\bigcap_{n=1}^{\infty} (-1/n, 1/n)$.

(d) Simplify $\displaystyle\bigcap_{n=1}^{\infty} [0, 3 + 1/(2n))$.

(e) Simplify $\displaystyle\bigcup_{n=1}^{\infty} [5, 7 - (3n)^{-1}]$.

(f) Simplify $\displaystyle\bigcup_{n=1}^{\infty} [0, n]$.

4. Fix two sets A and C. If $C \subset A$, show that for every subset B,

$$(A \cap B) \cup C \ = \ A \cap (B \cup C). \tag{1.35}$$

Also show that if (1.35) holds for any set B, then $C \subset A$ (and thus (1.35) holds for all sets B).

*5. Let A and B be subsets of Ω. Put

$$I := \{\omega \in \Omega : \omega \in A \text{ implies } \omega \in B\}.$$

Show that $A \cap I = A \cap B$.

*6. Explain why $f: (\Leftrightarrow\infty, \infty) \to [0, \infty)$ with $f(x) = x^3$ is not well defined.

*7. Sketch $f(x) = \sin(x)$ for $x \in [\Leftrightarrow\pi/2, \pi/2]$.

(a) Determine, if possible, a choice of co-domain Y such that $f: [\Leftrightarrow\pi/2, \pi/2] \to Y$ is invertible.

(b) Find the inverse image $\{x : f(x) \leq 1/2\}$.

(c) Find the inverse image $\{x : f(x) < 0\}$.

*8. Sketch $f(x) = \sin(x)$ for $x \in [0, \pi]$.

(a) Determine, if possible, a choice of co-domain Y such that $f: [0, \pi] \to Y$ is invertible.

(b) Find the inverse image $\{x : f(x) \leq 1/2\}$.

(c) Find the inverse image $\{x : f(x) < 0\}$.

*9. Let X be any set, and let $A \subset X$. Define the real-valued function f by

$$f(x) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Thus, $f: X \to \mathbb{R}$, where $\mathbb{R} := (\Leftrightarrow\infty, \infty)$ denotes the real numbers. For arbitrary $B \subset \mathbb{R}$, find $f^{-1}(B)$. *Hint:* There are four cases to consider, depending on whether 0 or 1 belong to B.

*10. Let $f: X \to Y$ be a function such that f takes only n distinct values, say $y_1, \ldots, y_n$. Define

$$A_i := f^{-1}(\{y_i\}) = \{x \in X : f(x) = y_i\}.$$

Show that for arbitrary $B \subset Y$, $f^{-1}(B)$ can be expressed as a union of the A_i. (It then follows that there are only 2^n possibilities for $f^{-1}(B)$.)

*11. If $f: X \to Y$, show that inverse images preserve the following set operations.

(a) If $B \subset Y$, show that $f^{-1}(B^c) = f^{-1}(B)^c$.

(b) If B_n is a sequence of subsets of Y, show that

$$f^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(B_n).$$

(c) If B_n is a sequence of subsets of Y, show that

$$f^{-1}\left(\bigcap_{n=1}^{\infty} B_n\right) = \bigcap_{n=1}^{\infty} f^{-1}(B_n).$$

*12. Show that if B and C are countable sets, then so is $B \cup C$. *Hint:* If $B = \bigcup_i \{b_i\}$ and $C = \bigcup_i \{c_i\}$, put $a_{2i} := \{b_i\}$ and $a_{2i-1} := \{c_i\}$.

*13. Let C_1, C_2, ... be countable sets. Show that

$$B := \bigcup_{i=1}^{\infty} C_i$$

is a countable set.

*14. Show that any subset of a countable set is countable.

*15. Show that if $A \subset B$ and A is uncountable, then so is B.

*16. Show that the union of a countable set and an uncountable set is uncountable.

§*1.2: Probability Models*

17. A letter of the alphabet (a–z) is generated at random. Specify a sample space Ω and a probability measure $\wp$. Compute the probability that a vowel (a, e, i, o, u) is generated.

18. A collection of plastic letters, a–z, is mixed in a jar. Two letters are drawn at random, one after the other. What is the probability of drawing a vowel (a, e, i, o, u) and a consonant in either order? Two vowels in any order? Specify your sample space Ω and probability $\wp$.

19. A new baby wakes up exactly once every night. The time at which the baby wakes up occurs at random between 9 pm and 7 am. If the parents go to sleep at 11 pm, what is the probability that the parents are not awakened by the baby before they would normally get up at 7 am? Specify your sample space Ω and probability $\wp$.

20. For any real or complex number $z \neq 1$ and any positive integer N, derive the **geometric series** formula

$$\boxed{\sum_{k=0}^{N-1} z^k = \frac{1 \Leftrightarrow z^N}{1 \Leftrightarrow z}, \quad z \neq 1.}$$

Hint: Let $S_N := 1 + z + \cdots + z^{N-1}$, and show that $S_N \Leftrightarrow z S_N = 1 \Leftrightarrow z^N$. Then solve for S_N.

Remark. If $|z| < 1$, $|z|^N \to 0$ as $N \to \infty$. Hence,

$$\boxed{\sum_{k=0}^{\infty} z^k \;=\; \frac{1}{1 \Leftrightarrow z}, \quad \text{for } |z| < 1.}$$

21. Let $\Omega := \{1, \ldots, 6\}$. If $p(\omega) = 2\,p(\omega \Leftrightarrow 1)$ for $\omega = 2, \ldots, 6$, and if $\sum_{\omega=1}^{6} p(\omega) = 1$, show that $p(\omega) = 2^{\omega-1}/63$. *Hint:* Use Problem 20.

§*1.3: Axioms and Properties of Probability*

22. Let A and B be events for which $\wp(A)$, $\wp(B)$, and $\wp(A \cup B)$ are known. Express the following in terms of these probabilities:

    (a) $\wp(A \cap B)$.

    (b) $\wp(A \cap B^c)$.

    (c) $\wp(B \cup (A \cap B^c))$.

    (d) $\wp(A^c \cap B^c)$.

23. Let Ω be a sample space equipped with two probability measures, $\wp_1$ and $\wp_2$. Given any $0 \le \lambda \le 1$, show that if $\wp(A) := \lambda \wp_1(A) + (1 \Leftrightarrow \lambda)\wp_2(A)$, then $\wp$ satisfies the four axioms of a probability measure.

24. Let Ω be a sample space, and fix any point $\omega_0 \in \Omega$. For any event A, put

$$\mu(A) \;:=\; \begin{cases} 1, & \omega_0 \in A, \\ 0, & \text{otherwise.} \end{cases}$$

    Show that μ satisfies the axioms of a probability measure.

25. Suppose that instead of axiom (iii) of Section 1.3, we assume only that for any two disjoint events A and B, $\wp(A \cup B) = \wp(A) + \wp(B)$. Use this assumption and induction‡ on N to show that for any *finite* sequence of pairwise disjoint events $A_1, \ldots, A_N$,

$$\wp\biggl(\bigcup_{n=1}^{N} A_n\biggr) \;=\; \sum_{n=1}^{N} \wp(A_n).$$

    Using this result for finite N, it is not possible to derive axiom (iii), which is the assumption needed to derive the limit results of Section 1.3.

‡In this case, using induction on N means that you first verify the desired result for $N = 2$. Second, you assume the result is true for some arbitrary $N \ge 2$ and then prove the desired result is true for $N + 1$.

*26. The purpose of this problem is to show that any countable union can be written as a union of pairwise disjoint sets. Given any sequence of sets F_n, define a new sequence by $A_1 := F_1$, and

$$A_n := F_n \cap F_{n-1}^c \cap \cdots \cap F_1^c, \quad n \geq 2.$$

Note that the A_n are pairwise disjoint. For finite $N \geq 1$, show that

$$\bigcup_{n=1}^{N} F_n = \bigcup_{n=1}^{N} A_n.$$

Also show that

$$\bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} A_n.$$

*27. Use the preceding problem to show that for any sequence of events F_n,

$$\wp\biggl(\bigcup_{n=1}^{\infty} F_n\biggr) = \lim_{N\to\infty} \wp\biggl(\bigcup_{n=1}^{N} F_n\biggr).$$

*28. Use the preceding problem to show that for any sequence of events G_n,

$$\wp\biggl(\bigcap_{n=1}^{\infty} G_n\biggr) = \lim_{N\to\infty} \wp\biggl(\bigcap_{n=1}^{N} G_n\biggr).$$

29. *The Finite Union Bound.* Show that for any finite sequence of events $F_1, \ldots, F_N$,

$$\wp\biggl(\bigcup_{n=1}^{N} F_n\biggr) \leq \sum_{n=1}^{N} \wp(F_n).$$

Hint: Use the inclusion-exclusion formula (1.12) and induction on N. See the last footnote for information on induction.

*30. *The Infinite Union Bound.* Show that for any infinite sequence of events F_n,

$$\wp\biggl(\bigcup_{n=1}^{\infty} F_n\biggr) \leq \sum_{n=1}^{\infty} \wp(F_n).$$

Hint: Combine Problems 27 and 29.

*31. *First Borel–Cantelli Lemma.* Show that if B_n is a sequence of events for which

$$\sum_{n=1}^{\infty} \wp(B_n) < \infty, \tag{1.36}$$

then

$$\wp\biggl(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} B_k\biggr) \;=\; 0.$$

Hint: Let $G := \bigcap_{n=1}^{\infty} G_n$, where $G_n := \bigcup_{k=n}^{\infty} B_k$. Now use Problem 28, the union bound of the preceding problem, and the fact that (1.36) implies

$$\lim_{N\to\infty} \sum_{n=N}^{\infty} \wp(B_n) \;=\; 0.$$

*32. *Second Borel–Cantelli Lemma.* Show that if B_n is a sequence of *independent* events for which

$$\sum_{n=1}^{\infty} \wp(B_n) \;=\; \infty,$$

then

$$\wp\biggl(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} B_k\biggr) \;=\; 1.$$

Hint: The inequality $1 \Leftrightarrow \wp(B_k) \leq \exp[\Leftrightarrow \wp(B_k)]$ may be helpful.[§]

*33. This problem assumes you have read Note 1. Let $A_1, \ldots, A_n$ be a partition of Ω. If $\mathcal{C} := \{A_1, \ldots, A_n\}$, show that $\sigma(\mathcal{C})$ consists of the empty set along with all unions of the form

$$\bigcup_i A_{k_i}$$

where k_i is a finite subsequence of distinct elements from $\{1, \ldots, n\}$.

*34. This problem assumes you have read Note 1. Let Ω be a sample space, and let $X\colon \Omega \to \mathrm{I\!R}$, where $\mathrm{I\!R}$ denotes the set of real numbers. Suppose the mapping X takes finitely many distinct values $x_1, \ldots, x_n$. Find the smallest σ-field $\mathcal{A}$ of subsets of Ω such that for all $B \subset \mathrm{I\!R}$, $X^{-1}(B) \in \mathcal{A}$. *Hint:* Problems 10 and 11.

*35. This problem assumes you have read Note 1. Let $\Omega := \{1, 2, 3, 4, 5\}$, and put $A := \{1, 2, 3\}$ and $B := \{3, 4, 5\}$. Put $\wp(A) := 5/8$ and $\wp(B) := 7/8$.

(a) Find $\mathcal{F} := \sigma(\{A, B\})$, the smallest σ-field containing the sets A and B.

(b) Compute $\wp(F)$ for all $F \in \mathcal{F}$.

(c) *Trick Question.* What is $\wp(\{1\})$?

*36. This problem assumes you have read Note 1. Show that a σ-field cannot be countably infinite; i.e., show that if a σ-field contains an infinite number of sets, then it contains an uncountable number of sets.

[§]The inequality $1 - x \leq e^{-x}$ for $x \geq 0$ can be derived by showing that the function $f(x) := e^{-x} - (1 - x)$ satisfies $f(0) \geq 0$ and is nondecreasing, e.g., $f'(x) \geq 0$.

*37. This problem assumes you have read Note 1.

(a) Let $\mathcal{A}_\alpha$ be any indexed collection of σ-fields. Show that $\bigcap_\alpha \mathcal{A}_\alpha$ is also a σ-field.

(b) Given a collection of subsets of Ω, show that

$$\mathcal{C}_0 \; := \; \bigcap_{\mathcal{C} \subset \mathcal{A}} \mathcal{A},$$

where the intersection is over all σ-fields $\mathcal{A}$ that contain $\mathcal{C}$, is the smallest σ-field containing $\mathcal{C}$.

*38. *The Borel σ-Field.* This problem assumes you have read Note 1. Let $\mathcal{B}$ denote the smallest σ-field containing all the open subsets of $\mathbb{R} := (\Leftrightarrow\infty, \infty)$. This collection $\mathcal{B}$ is called the **Borel $\boldsymbol{\sigma}$-field**. The sets in $\mathcal{B}$ are called **Borel sets**. Hence, every open set, and every open interval, is a Borel set.

(a) Show that every interval of the form $(a, b]$ is also a Borel set. *Hint:* Write $(a, b]$ as a countable intersection of open intervals and use the properties of a σ-field.

(b) Show that every singleton set $\{a\}$ is a Borel set.

(c) Let $a_1, a_2, \ldots$ be distinct real numbers. Put

$$A \; := \; \bigcup_{k=1}^{\infty} \{a_k\}.$$

Determine whether or not A is a Borel set.

(d) **Lebesgue measure** λ on the Borel subsets of $(0, 1)$ is a probability measure that is completely characterized by the property that the Lebesgue measure of an open interval $(a, b) \subset (0, 1)$ is its length; i.e., $\lambda\big((a, b)\big) = b \Leftrightarrow a$. Show that $\lambda\big((a, b]\big)$ is also equal to $b \Leftrightarrow a$. Find $\lambda(\{a\})$ for any singleton set. If the set A in part (c) is a Borel set, compute $\lambda(A)$.

Remark. Note 5 in Chapter 4 contains more details on the constuction of probability measures on the Borel subsets of $\mathbb{R}$.

*39. *The Borel σ-Field, Continued.* This problem assumes you have read Note 1.

Background: Recall that a set $U \subset \mathbb{R}$ is **open** if for every $x \in U$, there is a positive number ε_x, depending on x, such that $(x \Leftrightarrow \varepsilon_x, x + \varepsilon_x) \subset U$. Hence, an open set U can always be written in the form

$$U \; = \; \bigcup_{x \in U} (x \Leftrightarrow \varepsilon_x, x + \varepsilon_x).$$

Now observe that if $(x \Leftrightarrow \varepsilon_x, x + \varepsilon_x) \subset U$, we can find a rational number q_x close to x and a rational number $\rho_x < \varepsilon_x$ such that

$$x \in (q_x \Leftrightarrow \rho_x, q_x + \rho_x) \subset (x \Leftrightarrow \varepsilon_x, x + \varepsilon_x) \subset U.$$

Thus, every open set can be written in the form

$$U = \bigcup_{x \in U} (q_x \Leftrightarrow \rho_x, q_x + \rho_x),$$

where each q_x and each ρ_x is a rational number. Since the rational numbers form a countable set, there are only countably many such intervals with rational centers and rational lengths; hence, the union is really a countable one.

Problem: Show that the smallest σ-field containing all the open intervals is equal to the Borel σ-field defined in Problem 38.

§*1.4: Conditional Probability*

40. If

$$\frac{N(O_{\mathrm{c,ns}})}{N(O_{\mathrm{ns}})}, \quad N(O_{\mathrm{ns}}), \quad \frac{N(O_{\mathrm{c,s}})}{N(O_{\mathrm{s}})}, \quad \text{and} \quad N(O_{\mathrm{s}})$$

are given, compute $N(O_{\mathrm{nc,s}})$ and $N(O_{\mathrm{nc,ns}})$ in terms of them.

41. If $\wp(C)$ and $\wp(B \cap C)$ are positive, derive the **chain rule of conditional probability**,

$$\wp(A \cap B|C) = \wp(A|B \cap C)\,\wp(B|C).$$

Also show that

$$\wp(A \cap B \cap C) = \wp(A|B \cap C)\,\wp(B|C)\,\wp(C).$$

42. The university buys workstations from two different suppliers, Mini Micros (MM) and Highest Technology (HT). On delivery, 10% of MM's workstations are defective, while 20% of HT's workstations are defective. The university buys 140 MM workstations and 60 HT workstations for its computer lab. Suppose you walk into the computer lab and randomly sit down at a workstation.

    (a) What is the probability that your workstation is from MM? From HT?

    (b) What is the probability that your workstation is defective? *Answer:* 0.13.

    (c) Given that your workstation is defective, what is the probability that it came from Mini Micros? *Answer:* 7/13.

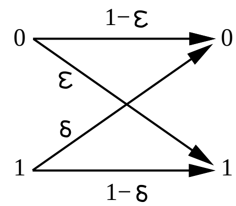


Figure 1.11. Binary channel with crossover probabilities ε and δ. If $\delta = \varepsilon$, this is called a binary symmetric channel.

43. The probability that a cell in a wireless system is overloaded is 1/3. Given that it is overloaded, the probability of a blocked call is 0.3. Given that it is not overloaded, the probability of a blocked call is 0.1. Find the conditional probability that the system is overloaded given that your call is blocked. *Answer:* 0.6.

44. The binary channel shown in Figure 1.11 operates as follows. Given that a 0 is transmitted, the conditional probability that a 1 is received is ε. Given that a 1 is transmitted, the conditional probability that a 0 is received is δ. Assume that the probability of transmitting a 0 is the same as the probability of transmitting a 1. Given that a 1 is received, find the conditional probability that a 1 was transmitted. *Hint:* Use the notation
$$T_i \;:=\; \{i \text{ is transmitted}\}, \quad i = 0, 1,$$
and
$$R_j \;:=\; \{j \text{ is received}\}, \quad j = 0, 1.$$
Remark. If $\delta = \varepsilon$, this channel is called the **binary symmetric channel**.

45. Professor Random has taught probability for many years. She has found that 80% of students who do the homework pass the exam, while 10% of students who don't do the homework pass the exam. If 60% of the students do the homework, what percent of students pass the exam? Of students who pass the exam, what percent did the homework? *Answer:* 12/13.

46. A certain jet aircraft's autopilot has conditional probability 1/3 of failure given that it employs a faulty microprocessor chip. The autopilot has conditional probability 1/10 of failure given that it employs nonfaulty chip. According to the chip manufacturer, the probability of a customer's receiving a faulty chip is 1/4. Given that an autopilot failure has occurred, find the conditional probability that a faulty chip was used. Use the following notation:
$$\begin{aligned} A_F &= \{\text{autopilot fails}\} \\ C_F &= \{\text{chip is faulty}\}. \end{aligned}$$

Answer: 10/19.

*47. You have five computer chips, two of which are known to be defective.

(a) You test one of the chips; what is the probability that it is defective?

(b) Your friend tests two chips at random and reports that one is defective and one is not. Given this information, you test one of the three remaining chips at random; what is the conditional probability that the chip you test is defective?

(c) Consider the following modification of the preceding scenario. Your friend takes away two chips at random for testing; before your friend tells you the results, you test one of the three remaining chips at random; given this (lack of) information, what is the conditional probability that the chip you test is defective? Since you have not yet learned the results of your friend's tests, intuition suggests that your conditional probability should be the same as your answer to part (a). Is your intuition correct?

§*1.5: Independence*

48. (a) If two sets A and B are disjoint, what equation must they satisfy?

(b) If two events A and B are independent, what equation must they satisfy?

(c) Suppose two events A and B are disjoint. Give conditions under which they are also independent. Give conditions under which they are not independent.

49. A certain binary communication system has a bit-error rate of 0.1; i.e., in transmitting a single bit, the probability of receiving the bit in error is 0.1. To transmit messages, a three-bit repetition code is used. In other words, to send the message **1**, 111 is transmitted, and to send the message **0**, 000 is transmitted. At the receiver, if two or more 1s are received, the decoder decides that message **1** was sent; otherwise, i.e., if two or more zeros are received, it decides that message **0** was sent. Assuming bit errors occur independently, find the probability that the decoder puts out the wrong message. *Answer:* 0.028.

50. You and your neighbor attempt to use your cordless phones at the same time. Your phones independently select one of ten channels at random to connect to the base unit. What is the probability that both phones pick the same channel?

51. A new car is equipped with dual airbags. Suppose that they fail independently with probability p. What is the probability that at least one airbag functions properly?

52. A discrete-time FIR filter is to be found satisfying certain constraints, such as energy, phase, sign changes of the coefficients, etc. FIR filters can be thought of as vectors in some finite-dimensional space. The energy constraint implies that all suitable vectors lie in some hypercube; i.e., a square in $\mathbb{R}^2$, a cube in $\mathbb{R}^3$, etc. It is easy to generate random vectors uniformly in a hypercube. This suggests the following Monte–Carlo procedure for finding a filter that satisfies all the desired properties. Suppose we generate vectors independently in the hypercube until we find one that has all the desired properties. What is the probability of ever finding such a filter?

53. A dart is repeatedly thrown at random toward a circular dartboard of radius 10 cm. Assume the thrower never misses the board. Let A_n denote the event that the dart lands within 2 cm of the center on the nth throw. Suppose that the A_n are mutually independent and that $\wp(A_n) = p$ for some $0 < p < 1$. Find the probability that the dart never lands within 2 cm of the center.

54. Each time you play the lottery, your probability of winning is p. You play the lottery n times, and plays are independent. How large should n be to make the probability of winning at least once more than $1/2$? *Answer:* For $p = 1/10^6$, $n \geq 693147$.

55. Alex and Bill go fishing. Find the conditional probability that Alex catches no fish given that at least one of them catches no fish. Assume they catch fish independently and that each has probability $0 < p < 1$ of catching no fish.

56. Consider the sample space $\Omega = [0, 1]$ equipped with the probability measure

$$\wp(A) \;:=\; \int_A 1\,d\omega, \quad A \subset \Omega.$$

For $A = [0, 1/2]$, $B = [0, 1/4] \cup [1/2, 3/4]$, and $C = [0, 1/8] \cup [1/4, 3/8] \cup [1/2, 5/8] \cup [3/4, 7/8]$, determine whether or not A, B, and C are mutually independent.

57. Given events A, B, and C, show that

$$\wp(A \cap C|B) \;=\; \wp(A|B)\,\wp(C|B)$$

if and only if

$$\wp(A|B \cap C) \;=\; \wp(A|B).$$

In this case, A and C are **conditionally independent** given B.

§*1.6: Combinatorics and Probability*

58. An electronics store carries 3 brands of computers, 5 brands of flat screens, and 7 brands of printers. How many different systems (computer, flat screen, and printer) can the store sell?

59. If we use binary digits, how many n-bit numbers are there?

60. A certain Internet message consists of 4 header packets plus 96 data packets. Unfortunately, a faulty router randomly re-orders the packets. What is the probability that the first header packet to arrive at the destination is the 10th packet received? *Answer:* 0.02996.

61. In a pick-4 lottery game, a player selects 4 digits, each one from $0, \ldots, 9$. If the 4 digits selected by the player match the random 4 digits of the lottery drawing in any order, the player wins. If the player has selected 4 distinct digits, what is the probability of winning? *Answer:* 0.0024.

62. How many 8-bit words are there with 3 ones (and 5 zeros)? *Answer:* 56.

63. A faulty computer memory location reads out random 8-bit bytes. What is the probability that a random word has 4 ones and 4 zeros? *Answer:* 0.2734.

64. Suppose 41 people enter a contest in which 3 winners are chosen at random. The first contestant chosen wins \$500, the second contestant chosen wins \$400, and the third contestant chosen wins \$250. How many different outcomes are possible? If all three winners receive \$250, how many different outcomes are possible? *Answers:* 63,960 and 10,660.

65. From a well-shuffled deck of 52 playing cards you are dealt 14 cards. What is the probability that 2 cards are spades, 3 are hearts, 4 are diamonds, and 5 are clubs? *Answer:* 0.0116.

66. From a well-shuffled deck of 52 playing cards you are dealt 5 cards. What is the probability that all 5 cards are of the same suit? *Answer:* 0.00198.

67. A finite set D of n elements is to be partitioned into m disjoint subsets, $D_1, \ldots, D_m$ in which $|D_i| = k_i$. How many different partitions are possible?

68. *m-ary Pick n Lottery.* In this game, a player chooses n m-ary digits. In the lottery drawing, n m-ary digits are chosen at random. If the n digits selected by the player match the random n digits of the lottery drawing in any order, the player wins. If the player has selected n digits with k_0 zeros, k_1 ones, $\ldots$, and k_{m-1} copies of digit $m \Leftrightarrow 1$, where $k_0 + \cdots + k_{m-1} = n$, what is the probability of winning? In the case of $n = 4$, $m = 10$, and a player's choice of the form $xxyz$, what is the probability of winning; for $xxyy$; for $xxxy$? *Answers:* 0.0012, 0.0006, 0.0004.

69. In Example 1.42, what 7-bit sequence corresponds to 2 apples and 3 carrots? What sequence corresponds to 2 apples and 3 bananas? What sequence corresponds to 5 apples?

Exam Preparation

You may use the following suggestions to prepare a study sheet, including formulas mentioned that you have trouble remembering. You may also want to ask your instructor for additional suggestions.

1.1. Review of Set Notation. Be familiar with set notation, operations, and identities. Graduate students should in addition be familiar with the precise definition of a function and the notions of countable and uncountable sets.

1.2. Probability Models. Know how to construct and use probability models for simple problems.

1.3. Axioms and Properties of Probability. Know axioms and properties of probability. Important formulas include (1.9) for disjoint unions, and (1.10) – (1.12). Graduate students should understand and know how to use (1.13)–(1.17); in addition, your instructor may also require familiarity with Note 1 and related problems concerning σ-fields.

1.4. Conditional Probability. What is important is the **law of total probability** (1.23) and *and being able to use it to solve problems.*

1.5. Independence. Do not confuse independent sets with disjoint sets. If $A_1, A_2, \ldots$ are independent, then so are $\tilde{A}_1, \tilde{A}_2, \ldots$, where each $\tilde{A}_i$ is either A_i or A_i^c.

1.6. Combinatorics and Probability. The four kinds of counting problems are

- (i) ordered sampling of k out of n items with replacement: n^k.
- (ii) ordered sampling of $k \leq n$ out of n items without replacement: $n!/(n \Leftrightarrow k)!$.
- (iii) unordered sampling of $k \leq n$ out of n items without replacement: $\binom{n}{k}$.
- (iv) unordered sampling of k out of n items with replacement: $\binom{k+n-1}{k}$.

Know also the multinomial coefficient.

Work any review problems assigned by your instructor. If you finish them, re-work your homework assignments.