Probability, Statistics, and Stochastic Processes

Peter Olofsson

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.

New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

Preface

The Book

In November 2003, I was completing a review of an undergraduate textbook in probability and statistics. In the enclosed evaluation sheet was the question "Have you ever considered writing a textbook?" and I suddenly realized that the answer was "Yes," and had been for quite some time. For several years I had been teaching a course on calculus-based probability and statistics mainly for mathematics, science, and engineering students. Other than the basic probability theory, my goal was to include topics from two areas: statistical inference and stochastic processes. For many students this was the only probability/statistics course they would ever take, and I found it desirable that they were familiar with confidence intervals and the maximum likelihood method, as well as Markov chains and queueing theory. While there were plenty of books covering one area or the other, it was surprisingly difficult to find one that covered both in a satisfying way and on the appropriate level of difficulty. My solution was to choose one textbook and supplement it with lecture notes in the area that was missing. As I changed texts often, plenty of lecture notes accumulated and it seemed like a good idea to organize them into a textbook. I was pleased to learn that the good people at Wiley agreed.

It is now more than a year later, and the book has been written. The first three chapters develop probability theory and introduce the axioms of probability, random variables, and joint distributions. The following two chapters are shorter and of an "introduction to" nature: Chapter 4 on limit theorems and Chapter 5 on simulation. Statistical inference is treated in Chapter 6, which includes a section on Bayesian

statistics, too often a neglected topic in undergraduate texts. Finally, in Chapter 7, Markov chains in discrete and continuous time are introduced. The reference list at the end of the book is by no means intended to be comprehensive; rather, it is a subjective selection of the useful and the entertaining.

Throughout the text I have tried to convey an intuitive understanding of concepts and results, which is why a definition or a proposition is often preceded by a short discussion or a motivating example. I have also attempted to make the exposition entertaining by choosing examples from the rich source of fun and thought-provoking probability problems. The data sets used in the statistics chapter are of three different kinds: real, fake but realistic, and unrealistic but illustrative.

The people

Most textbook authors start by thanking their spouses. I know now that this is far more than a formality, and I would like to thank $A\lambda\kappa\mu\acute{\eta}\nu\eta$ not only for patiently putting up with irregular work hours and an absentmindedness greater than usual but also for valuable comments on the aesthetics of the manuscript.

A number of people have commented on various parts and aspects of the book. First, I would like to thank Olle Häggström at Chalmers University of Technology, Göteborg, Sweden for valuable comments on all chapters. His remarks are always accurate and insightful, and never obscured by unnecessary politeness. Second, I would like to thank Kjell Doksum at the University of Wisconsin for a very helpful review of the statistics chapter. I have also enjoyed the Bayesian enthusiasm of Peter Müller at the University of Texas MD Anderson Cancer Center.

Other people who have commented on parts of the book or been otherwise helpful are my colleagues Dennis Cox, Kathy Ensor, Rudy Guerra, Marek Kimmel, Rolf Riedi, Javier Rojo, David W. Scott, and Jim Thompson at Rice University; Prof. Dr. R.W.J. Meester at Vrije Universiteit, Amsterdam, The Netherlands; Timo Seppäläinen at the University of Wisconsin; Tom English at Behrend College; Robert Lund at Clemson University; and Jared Martin at Shell Exploration and Production. For help with solutions to problems, I am grateful to several bright Rice graduate students: Blair Christian, Julie Cong, Talithia Daniel, Ginger Davis, Li Deng, Gretchen Fix, Hector Flores, Garrett Fox, Darrin Gershman, Jason Gershman, Shu Han, Shannon Neeley, Rick Ott, Galen Papkov, Bo Peng, Zhaoxia Yu, and Jenny Zhang. Thanks to Mikael Andersson at Stockholm University, Sweden for contributions to the problem sections, and to Patrick King at ODS–Petrodata, Inc. for providing data with a distinct Texas flavor: oil rig charter rates. At Wiley, I would like to thank Steve Quigley, Susanne Steitz, and Kellsee Chu for always promptly answering my questions. Finally, thanks to John Haigh, John Allen Paulos, Jeffrey E. Steif, and an anonymous Dutchman for agreeing to appear and be mildly mocked in footnotes.

PETER OLOFSSON

Houston, Texas, 2005

Contents

1

Basic Probability Theory

1.1 INTRODUCTION

Probability theory is the mathematics of randomness. This statement immediately invites the question "What is randomness?" This is a deep question that we cannot attempt to answer without invoking the disciplines of philosophy, psychology, mathematical complexity theory, and quantum physics, and still there would most likely be no completely satisfactory answer. For our purposes, an informal definition of randomness as "what happens in a situation where we cannot predict the outcome with certainty" is sufficient. In many cases, this might simply mean lack of information. For example, if we flip a coin, we might think of the outcome as random. It will be either heads or tails, but we cannot say which, and if the coin is fair, we believe that both outcomes are equally likely. However, if we knew the force from the fingers at the flip, weight and shape of the coin, material and shape of the table surface, and several other parameters, we would be able to predict the outcome with certainty, according to the laws of physics. In this case we use randomness as a way to describe uncertainty due to lack of information.[1]

Next question: "What is probability?" There are two main interpretations of probability, one that could be termed "objective" and the other "subjective." The first is the interpretation of a probability as a *limit of relative frequencies*; the second, as a *degree of belief*. Let us briefly describe each of these.

[1]To quote the French mathematician Pierre-Simon Laplace, one of the first to develop a mathematical theory of probability: "Probability is composed partly of our ignorance, partly of our knowledge."

For the first interpretation, suppose that we have an experiment where we are interested in a particular outcome. We can repeat the experiment over and over and each time record whether we got the outcome of interest. As we proceed, we count the number of times that we got our outcome and divide this number by the number of times that we performed the experiment. The resulting ratio is the *relative frequency* of our outcome. As it can be observed empirically that such relative frequencies tend to stabilize as the number of repetitions of the experiment grows, we might think of the limit of the relative frequencies as the probability of the outcome. In mathematical notation, if we consider n repetitions of the experiment and if S_n of these gave our outcome, then the relative frequency would be $f_n = S_n/n$, and we might say that the probability equals $\lim_{n\to\infty} f_n$. Figure 1.1 shows a plot of the relative frequency of heads in a computer simulation of 100 hundred coin flips. Notice how there is significant variation in the beginning but how the relative frequency settles in toward $\frac{1}{2}$ quickly.

The second interpretation, probability as a degree of belief, is not as easily quantified but has obvious intuitive appeal. In many cases, it overlaps with the previous interpretation, for example, the coin flip. If we are asked to quantify our degree of belief that a coin flip gives heads, where 0 means "impossible" and 1 means "with certainty," we would probably settle for $\frac{1}{2}$ unless we have some specific reason to believe that the coin is not fair. In some cases it is not possible to repeat the experiment in practice, but we can still imagine a sequence of repetitions. For example, in a weather forecast you will often hear statements like "there is a 30% chance of rain tomorrow." Of course, we cannot repeat the experiment; either it rains tomorrow or it does not. The 30% is the meteorologist's measure of the chance of rain. There is still a connection to the relative frequency approach; we can imagine a sequence of days with similar weather conditions, same time of year, and so on, and that in roughly 30% of the cases, it rains the following day.

The "degree of belief" approach becomes less clear for statements such as "the Riemann hypothesis is true" or "there is life on other planets." Obviously these are statements that are either true or false, but we do not know which, and it is not

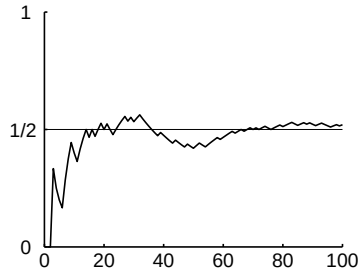

Fig. 1.1 Consecutive relative frequencies of heads in 100 coin flips.

unreasonable to use probabilities to express how strongly we believe in their truth. It is also obvious that different individuals may assign completely different probabilities.

How, then, do we actually *define* a probability? Instead of trying to use any of these interpretations, we will state a strict mathematical definition of probability. The interpretations are still valid to develop intuition for the situation at hand, but instead of, for example, *assuming* that relative frequencies stabilize, we will be able to *prove* that they do, within our theory.

1.2 SAMPLE SPACES AND EVENTS

As mentioned in the introduction, probability theory is a mathematical theory to describe and analyze situations where randomness or uncertainty are present. Any specific such situation will be referred to as a *random experiment*. We use the term "experiment" in a wide sense here; it could mean an actual physical experiment such as flipping a coin or rolling a die, but it could also be a situation where we simply observe something, such as the price of a stock at a given time, the amount of rain in Houston in September, or the number of spam emails we receive in a day. After the experiment is over, we call the result an *outcome*. For any given experiment, there is a set of possible outcomes, and we state the following definition.

***Definition* 1.2.1.** The set of all possible outcomes in a random experiment is called the *sample space*, denoted S.

Here are some examples of random experiments and their associated sample spaces.

***Example* 1.2.1.** Roll a die and observe the number.

Here we can get the numbers 1 through 6, and hence the sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

□

***Example* 1.2.2.** Roll a die repeatedly and count the number of rolls it takes until the first 6 appears.

Since the first 6 may come in the first roll, 1 is a possible outcome. Also, we may fail to get 6 in the first roll and then get 6 in the second, so 2 is also a possible outcome. If we continue this argument we realize that any positive integer is a possible outcome and the sample space is

$$S = \{1, 2, ...\}$$

the set of positive integers. □

***Example* 1.2.3.** Turn on a lightbulb and measure its lifetime, that is, the time until it fails.

Here it is not immediately clear what the sample space should be, since it depends on how accurately we can measure time. The most convenient approach is to note that the lifetime, at least in theory, can assume any nonnegative real number and choose as the sample space

$$S = [0, \infty)$$

where the outcome 0 means that the lightbulb is broken to start with. □

In these three examples, we have sample spaces of three different kinds. The first is *finite*, meaning that it has a finite number of outcomes, whereas the second and third are infinite. Although they are both infinite, they are different in the sense that one has its points separated, $\{1, 2, ...\}$ and the other is an entire continuum of points. We call the first type *countable infinity* and the second *uncountable infinity*. We will return to these concepts later as they turn out to form an important distinction.

In the examples above, the outcomes are always numbers and hence the sample spaces are subsets of the real line. Here are some examples of other types of sample spaces.

***Example* 1.2.4.** Flip a coin twice and observe the sequence of heads and tails.

With H denoting heads and T denoting tails, one possible outcome is HT, which means that we get heads in the first flip and tails in the second. Arguing like this, there are four possible outcomes and the sample space is

$$S = \{HH, HT, TH, TT\}$$

□

***Example* 1.2.5.** Throw a dart at random on a dart board of radius r.

If we think of the board as a disk in the plane with center at the origin, an outcome is an ordered pair of real numbers (x, y), and we can describe the sample space as

$$S = \{(x, y) : x^2 + y^2 \leq r^2\}$$

□

Once we have described an experiment and its sample space, we want to be able to compute probabilities of the various things that may happen. What is the probability that we get 6 when we roll a die? That the first 6 does not come before the fifth roll? That the lightbulb works for at least 1500 hours? That our dart hits the bull's eye? Certainly we need to make further assumptions to be able to answer these questions, but before that, we realize that all these questions have something in common. They all ask for probabilities of either single outcomes or groups of outcomes. Mathematically, we can describe these as subsets of the sample space.

***Definition* 1.2.2.** A subset of S, $A \subseteq S$, is called an *event*.

Note the choice of words here. The terms "outcome" and "event" reflect the fact that we are describing things that may happen in real life. Mathematically, these are described as elements and subsets of the sample space. This duality is typical for probability theory; there is a verbal description and a mathematical description of the same situation. The verbal description is natural when real-world phenomena are described and the mathematical formulation is necessary to develop a consistent theory. See Table 1.1 for a list of set operations and their verbal description.

***Example* 1.2.6.** If we roll a die and observe the number, two possible events are that we get an odd outcome and that we get at least 4. If we view these as subsets of the sample space we get

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

If we want to use the verbal description we might write this as

$$A = \{\text{odd outcome}\} \quad \text{and} \quad B = \{\text{at least } 4\}$$

□

We always use "or" in its nonexclusive meaning; thus, "A or B occurs" includes the possibility that both occur. Note that there are different ways to express combinations of events; for example, $A \setminus B = A \cap B^c$ and $(A \cup B)^c = A^c \cap B^c$. The latter is known as one of *De Morgan's laws*, and we state these without proof together with some other basic set theoretic rules.

Table 1.1 Basic set operations and their verbal description.

Notation	Mathematical description	Verbal description
$A \cup B$	The union of A and B	A or B (or both) occurs
$A \cap B$	The intersection of A and B	Both A and B occur
A^c	The complement of A	A does not occur
$A \setminus B$	The difference between A and B	A occurs but not B
$\emptyset$	The empty set	Impossible event

Proposition 1.2.1. Let A, B, and C be events. Then

(a) (*Distributive Laws*) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

(b) (*De Morgan's Laws*) $(A \cup B)^c = A^c \cap B^c$

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

As usual when dealing with set theory, *Venn diagrams* are useful. See Figure 1.2 for an illustration of some of the set operations introduced above. We will later return to how Venn diagrams can be used to calculate probabilities. If A and B are such that $A \cap B = \emptyset$, they are said to be *disjoint* or *mutually exclusive*. In words, this means that they cannot both occur simultaneously in the experiment.

As we will often deal with unions of more than two or three events, we need more general versions of the results given above. Let us first introduce some notation. If $A_1, A_2, ..., A_n$ is a sequence of events, we denote

$$\bigcup_{k=1}^{n} A_k = A_1 \cup A_2 \cup \cdots \cup A_n$$

the union of all the A_k and

$$\bigcap_{k=1}^{n} A_k = A_1 \cap A_2 \cap \cdots \cap A_n$$

the intersection of all the A_k. In words, these are the events that *at least one* of the A_k occurs and that *all* the A_k occur, respectively. The distributive and De Morgan's

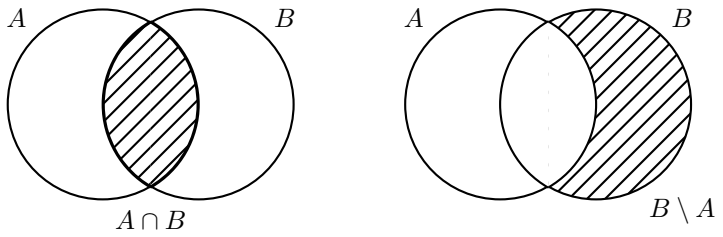

Fig. 1.2 Venn diagrams of the intersection and the difference between events.

laws extend in the obvious way, for example

$$\left(\bigcup_{k=1}^{n} A_k\right)^c = \bigcap_{k=1}^{n} A_k^c$$

It is also natural to consider infinite unions and intersections. For example, in Example 1.2.2, the event that the first 6 comes in an odd roll is the infinite union $\{1\} \cup \{3\} \cup \{5\} \cup \cdots$ and we can use the same type of notation as for finite unions and write

$$\{\text{first } 6 \text{ in odd roll}\} = \bigcup_{k=1}^{\infty} \{2k-1\}$$

For infinite unions and intersections, distributive and De Morgan's laws still extend in the obvious way.

1.3 THE AXIOMS OF PROBABILITY

In the previous section, we laid the basis for a theory of probability by describing random experiments in terms of the sample space, outcomes, and events. As mentioned, we want to be able to compute probabilities of events. In the introduction, we mentioned two different interpretations of probability: as a limit of relative frequencies and as a degree of belief. Since our aim is to build a consistent mathematical theory, as widely applicable as possible, our definition of probability should not depend on any particular interpretation. For example, it makes intuitive sense to require a probability to always be less than or equal to one (or equivalently, less than or equal to 100%). You cannot flip a coin 10 times and get 12 heads. Also, a statement such as "I am 150% sure that it will rain tomorrow" may be used to express extreme pessimism regarding an upcoming picnic but is certainly not sensible from a logical point of view. Also, a probability should be equal to one (or 100%), when there is absolute certainty, regardless of any particular interpretation.

Other properties must hold as well. For example, if you think there is a 20% chance that Bob is in his house, a 30% chance that he is in his backyard, and a 50% chance

that he is at work, then the chance that he is at home is 50%, the sum of 20% and 30%. Relative frequencies are also *additive* in this sense, and it is natural to demand that the same rule apply for probabilities.

We now give a mathematical definition of probability, where it is defined as a real-valued function of the events, satisfying three properties, which we refer to as the *axioms of probability*. In the light of the discussion above, they should be intuitively reasonable.

***Definition* 1.3.1. (*Axioms of Probability*).** A *probability measure* is a function P, which assigns to each event A a number $P(A)$ satisfying

(a) $0 \leq P(A) \leq 1$

(b) $P(S) = 1$

(c) If $A_1, A_2, ...$ is a sequence of *pairwise disjoint* events, that is, if $i \neq j$, then $A_i \cap A_j = \emptyset$, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

We read $P(A)$ as "the probability of A." Note that a probability in this sense is a real number between 0 and 1 but we will occasionally also use percentages so that, for example, the phrases "The probability is 0.2" and "There is a 20% chance" mean the same thing.[2]

The third axiom is the most powerful assumption when it comes to deducing properties and further results. Some texts prefer to state the third axiom for finite unions only, but since infinite unions naturally arise even in simple examples, we choose this more general version of the axioms. As it turns out, the finite case follows as a consequence of the infinite. We next state this in a proposition and also that the empty set has probability zero. Although intuitively obvious, we must prove that it follows from the axioms. We leave this as an exercise.

[2]If the sample space is very large, it may be impossible to assign probabilities to *all* events. The class of events then needs to be restricted to what is called a *σ-field*. For a more advanced treatment of probability theory, this is a necessary restriction, but we can safely disregard this problem.

Proposition 1.3.1. Let P be a probability measure. Then

(a) $P(\emptyset) = 0$

(b) If $A_1, ..., A_n$ are pairwise disjoint events, then

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$$

In particular, if A and B are disjoint, then $P(A \cup B) = P(A) + P(B)$. In general, unions need not be disjoint and we next show how to compute the probability of a union in general, as well as prove some other basic properties of the probability measure.

Proposition 1.3.2. Let P be a probability measure on some sample space S and let A and B be events. Then

(a) $P(A^c) = 1 - P(A)$

(b) $P(A \setminus B) = P(A) - P(A \cap B)$

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(d) If $A \subseteq B$, then $P(A) \leq P(B)$

Proof. We prove (b) and (c), and leave (a) and (d) as exercises. For (b), note that $A = (A \cap B) \cup (A \setminus B)$, which is a disjoint union, and Proposition 1.3.1 gives

$$P(A) = P(A \cap B) + P(A \setminus B)$$

which proves the assertion. For (c), we write $A \cup B = A \cup (B \setminus A)$, which is a disjoint union, and we get

$$P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B)$$

by part (b). ■

Note how we repeatedly used Proposition 1.3.1(b), the finite version of the third axiom. In Proposition 1.3.2(c), for example, the events A and B are not necessarily

disjoint but we can represent their union as a union of other events that are disjoint, thus allowing us to apply the third axiom.

***Example* 1.3.1.** Mrs Boudreaux and Mrs Thibodeaux are chatting over their fence when the new neighbor walks by. He is a man in his sixties with shabby clothes and a distinct smell of cheap whiskey. Mrs B, who has seen him before, tells Mrs T that he is a former Louisiana state senator. Mrs T finds this very hard to believe. "Yes," says Mrs B, "he is a former state senator who got into a scandal long ago, had to resign and started drinking." "Oh," says Mrs T, "that sounds more probable." "No," says Mrs B, "I think you mean less probable."

Actually, Mrs B is right. Consider the following two statements about the shabby man: "He is a former state senator" and "He is a former state senator who got into a scandal long ago, had to resign, and started drinking." It is tempting to think that the second is more probable because it gives a more exhaustive explanation of the situation at hand. However, this is precisely why it is a *less* probable statement. To explain this with probabilities, consider the experiment of observing a person and the two events

$$\begin{aligned} A &= \{\text{he is a former state senator}\} \\ B &= \{\text{he got into a scandal long ago, had to resign and started drinking}\} \end{aligned}$$

The first statement then corresponds to the event A and the second to the event $A \cap B$, and since $A \cap B \subseteq A$, we get $P(A \cap B) \leq P(A)$. Of course, what Mrs T meant was that it was easier to believe that the man was a former state senator once she knew more about his background.

In their book *Judgment under Uncertainty*, Kahneman et al. [5], show empirically how people often make similar mistakes when asked to choose the most probable among a set of statements. With a strict application of the rules of probability we get it right. □

***Example* 1.3.2.** Consider the following statement: "I heard on the news that there is a 50% chance of rain on Saturday and a 50% chance of rain on Sunday. Then there must be a 100% chance of rain during the weekend."

This is, of course, not true. However, it may be harder to point out precisely where the error lies, but we can address it with probability theory. The events of interest are

$$A = \{\text{rain on Saturday}\} \quad \text{and} \quad B = \{\text{rain on Sunday}\}$$

and the event of rain during the weekend is then $A \cup B$. The percentages are reformulated as probabilities so that $P(A) = P(B) = 0.5$ and we get

$$\begin{aligned} P(\text{rain during the weekend}) &= P(A \cup B) \\ &= P(A) + P(B) - P(A \cap B) \\ &= 1 - P(A \cap B) \end{aligned}$$

which is less than 1, that is, the chance of rain during the weekend is less than 100%. The error in the statement lies in that we can add probabilities only when the events are disjoint. In general, we need to subtract the probability of the intersection, which in this case is the probability that it rains both Saturday and Sunday. □

***Example* 1.3.3.** A dart board has area of 143 in.^{2} (square inches). In the center of the board, there is the "bulls eye," which is a disk of area 1 in.^{2}. The rest of the board is divided into 20 sectors numbered $1, 2, ..., 20$. There is also a triple ring that has an area of 10 in.^{2} and a double ring of area 15 in.^{2} (everything rounded to nearest integers). Suppose that you throw a dart at random on the board. What is the probability that you get **(a)** double 14, **(b)** 14 but not double, **(c)** triple or the bull's eye, **(d)** an even number or a double?

Introduce the events $F = \{14\}$, $D = \{\text{double}\}$, $T = \{\text{triple}\}$, $B = \{\text{bull's eye}\}$, and $E = \{\text{even}\}$. We interpret "throw a dart at random" to mean that any region is hit with a probability that equals the fraction of the total area of the board that region occupies. For example, each number has area $(143 - 1)/20 = 7.1$ in.^{2} so the corresponding probability is $7.1/143$. We get

$$P(\text{double } 14) = P(D \cap F) = \frac{0.75}{143} \approx 0.005$$

$$\begin{aligned} P(14 \text{ but not double}) &= P(F \setminus D) = P(F) - P(F \cap D) \\ &= \frac{7.1}{143} - \frac{0.75}{143} \approx 0.044 \end{aligned}$$

$$\begin{aligned} P(\text{triple or bulls eye}) &= P(T \cup B) = P(T) + P(B) \\ &= \frac{10}{143} + \frac{1}{143} \approx 0.077 \end{aligned}$$

$$\begin{aligned} P(\text{even or double}) &= P(E \cup D) = P(E) + P(D) - P(E \cap D) \\ &= \frac{71}{143} + \frac{15}{143} - \frac{7.5}{143} \approx 0.55 \end{aligned}$$

□

Let us say a word here about the interplay between logical statements and events. In the previous example, consider the events $E = \{\text{even}\}$ and $F = \{14\}$. Clearly, if we get 14, we also get an even number. As a logical relation between statements, we would express this as

$$\text{the number is } 14 \Rightarrow \text{the number is even}$$

and in terms of events, we would say "If F occurs, then E must also occur." But this means that $F \subseteq E$ and hence

$$\{\text{the number is } 14\} \subseteq \{\text{the number is even}\}$$

and thus the set-theoretic analog of "$\Rightarrow$" is "$\subseteq$" which is useful to keep in mind.

Venn diagrams turn out to provide a nice and useful interpretation of probabilities. If we imagine the sample space S to be a rectangle of area 1, we can interpret the probability of an event A as the area of A (see Figure 1.3). For example, Proposition 1.3.2(c) says that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. With the interpretation of probabilities as areas, we thus have

$$\begin{aligned} P(A \cup B) &= \text{area of } A \cup B \\ &= \text{area of } A + \text{ area of } B - \text{ area of } A \cap B \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

since when we add the areas of A and B, we count the area of $A \cap B$ twice and must subtract it (think of A and B as overlapping pancakes where we are interested only in how much area they cover). Strictly speaking, this is not a proof but the method can be helpful to find formulas that can then be proved formally. In the case of three events, consider Figure 1.4 to argue that

$$\begin{aligned} \text{Area of } A \cup B \cup C &= \text{area of } A + \text{ area of } B + \text{ area of } C \\ &\quad - \text{ area of } A \cap B - \text{ area of } A \cap C - \text{ area of } B \cap C \\ &\quad + \text{ area of } A \cap B \cap C \end{aligned}$$

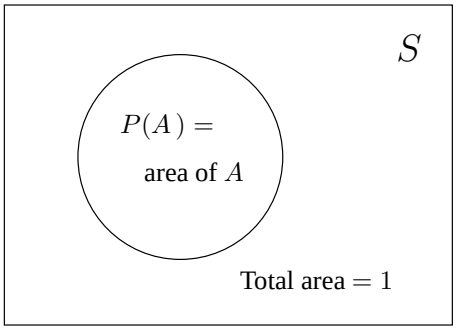

Fig. 1.3 Probabilities with Venn diagrams.

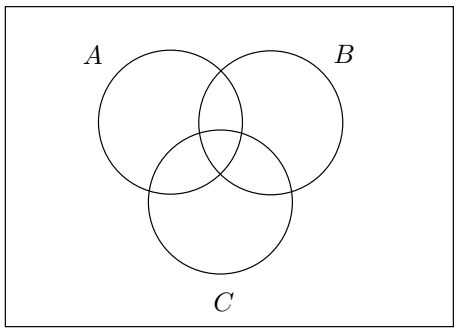

Fig. 1.4 Venn diagram of three events.

since the piece in the middle was first added 3 times and then removed 3 times, so in the end we have to add it again. Note that we must draw the diagram so that we get all possible combinations of intersections between the events. We have argued for the following proposition, which we state and prove formally.

Proposition 1.3.3. Let A, B, and C be three events. Then

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

Proof. By applying Proposition 1.3.2(c) twice — first to the two events $A \cup B$ and C and secondly to the events A and B — we obtain

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cup B) \cap C) \end{aligned}$$

The first four terms are what they should be. To deal with the last term, note that by the distributive laws for set operations, we obtain

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

and yet another application of Proposition 1.3.2(c) gives

$$\begin{aligned} P((A \cup B) \cap C) &= P((A \cap C) \cup (B \cap C)) \\ &= P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) \end{aligned}$$

which gives the desired result. ■

***Example* 1.3.4.** Choose a number at random from the numbers $1, ..., 100$. What is the probability that the chosen number is divisible by either $2, 3$, or 5?

Introduce the events

$$A_k = \{\text{divisible by } k\}, \text{ for } k = 1, 2, ...$$

We interpret "at random" to mean that any set of numbers has a probability that is equal to its relative size, that is, the number of elements divided by 100. We then get

$$P(A_2) = 0.5, \ P(A_3) = 0.33, \text{ and } P(A_5) = 0.2$$

For the intersection, first note that, for example, $A_2 \cap A_3$ is the event that the number is divisible by both 2 and 3, which is the same as saying it is divisible by 6. Hence $A_2 \cap A_3 = A_6$ and

$$P(A_2 \cap A_3) = P(A_6) = 0.16$$

Similarly, we get

$$P(A_2 \cap A_5) = P(A_{10}) = 0.1, \quad P(A_3 \cap A_5) = P(A_{15}) = 0.06$$

and

$$P(A_2 \cap A_3 \cap A_5) = P(A_{30}) = 0.03$$

The event of interest is $A_2 \cup A_3 \cup A_5$, and Proposition 1.3.3 yields

$$P(A_2 \cup A_3 \cup A_5) = 0.5 + 0.33 + 0.2 - (0.16 + 0.1 + 0.06) + 0.03 = 0.74$$

□

It is now easy to believe that the general formula for a union of n events starts by adding the probabilities of the events, then subtracting the probabilities of the pairwise intersections, adding the probabilities of intersections of triples and so on, finishing with either adding or subtracting the intersection of all the n events, depending on whether n is odd or even. We state this in a proposition that is sometimes referred to as the *inclusion–exclusion formula*. It can, for example, be proved by induction, but we leave the proof as an exercise.

Proposition 1.3.4. Let $A_1, A_2, ..., A_n$ be a sequence of n events. Then

$$\begin{aligned} P\left(\bigcup_{k=1}^{n} A_k\right) &= \sum_{k=1}^{n} P(A_k) \\ &\quad - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \\ &\quad \vdots \\ &\quad + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n) \end{aligned}$$

We finish this section with a theoretical result that will be useful from time to time. A sequence of events is said to be *increasing* if

$$A_1 \subseteq A_2 \subseteq \cdots$$

and *decreasing* if

$$A_1 \supseteq A_2 \supseteq \cdots$$

In each case we can define the *limit* of the sequence. If the sequence is increasing, we define

$$\lim_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} A_k$$

and if the sequence is decreasing

$$\lim_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} A_k$$

Note how this is similar to limits of sequences of numbers, with $\subseteq$ and $\supseteq$ corresponding to $\leq$ and $\geq$, respectively, and union and intersection corresponding to supremum and infimum. The following proposition states that the probability measure is a *continuous set function*. The proof is outlined in Problem 15.

Proposition 1.3.5. If $A_1, A_2, ...$ is either increasing or decreasing, then

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

1.4 FINITE SAMPLE SPACES AND COMBINATORICS

The results in the previous section hold for an arbitrary sample space S. In this section we will assume that S is finite, $S = \{s_1, ..., s_n\}$, say. In this case, we can always define the probability measure by assigning probabilities to the individual outcomes.

Proposition 1.4.1. Suppose that $p_1, ..., p_n$ are numbers such that

(a) $p_k \geq 0, \quad k = 1, ..., n$

(b) $\sum_{k=1}^{n} p_k = 1$

and for any event $A \subseteq S$, define

$$P(A) = \sum_{k:s_k \in A} p_k$$

Then P is a probability measure.

Proof. Clearly, the first two axioms of probability are satisfied. For the third, note that in a finite sample space, we cannot have infinitely many disjoint events, so we only have to check this for a disjoint union of two events A and B. We get

$$P(A \cup B) = \sum_{k:s_k \in A \cup B} p_k = \sum_{k:s_k \in A} p_k + \sum_{k:s_k \in B} p_k = P(A) + P(B)$$

and we are done. (Why are two events enough?) ■

Hence, when dealing with finite sample spaces, we do not need to explicitly give the probability of every event, only for each outcome. We refer to the numbers $p_1, ..., p_n$ as a *probability distribution* on S.

***Example* 1.4.1.** Consider the experiment of flipping a fair coin twice and counting the number of heads. We can take the sample space

$$S = \{HH, HT, TH, TT\}$$

and let $p_1 = ... = p_4 = \frac{1}{4}$. Alternatively, since all we are interested in is the number of heads and this can be $0, 1$, or 2, we can use the sample space

$$S = \{0, 1, 2\}$$

and let $p_0 = \frac{1}{4}, p_1 = \frac{1}{2}, p_2 = \frac{1}{4}$. □

Of particular interest is the case when all outcomes are equally likely. If S has n equally likely outcomes, then $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$, which is called a *uniform distribution* on S. The formula for the probability of an event A now simplifies to

$$P(A) = \sum_{k:s_k \in A} \frac{1}{n} = \frac{\#A}{n}$$

where $\#A$ denotes the number of elements in A. This formula is often referred to as the *classical definition of probability*, since historically this was the first context in which probabilities were studied. The outcomes in the event A can be described as *favorable* to A and we get the following formulation.

Corollary 1.4.2. In a finite sample space with uniform probability distribution

$$P(A) = \frac{\#\text{ favorable outcomes}}{\#\text{ possible outcomes}}$$

In daily language, the term "at random" is often used for something that has a uniform distribution. Although our concept of randomness is more general, this colloquial notion is so common that we will also use it (and already have). Thus, if we say "pick a number at random from $1, ..., 10$," we mean "pick a number according to a uniform probability distribution on the sample space $\{1, 2, ..., 10\}$."

***Example* 1.4.2.** Roll a fair die 3 times. What is the probability that all numbers are the same?

The sample space is the set of the 216 ordered triples (i, j, k), and since the die is fair, these are all equally probable and we have a uniform probability distribution. The event of interest is

$$A = \{(1, 1, 1), (2, 2, 2), ..., (6, 6, 6)\}$$

which has six outcomes and probability

$$P(A) = \frac{\#\text{ favorable outcomes}}{\#\text{ possible outcomes}} = \frac{6}{216} = \frac{1}{36}$$ □

***Example* 1.4.3.** Consider a randomly chosen family with three children. What is the probability that they have exactly one daughter?

There are eight possible sequences of boys and girls (in order of birth), and we get the sample space

$$S = \{bbb, bbg, bgb, bgg, gbb, gbg, ggb, ggg\}$$

where, for example, bbg means that the oldest child is a boy, the middle child a boy, and the youngest child a girl. If we assume that all outcomes are equally likely, we get a uniform probability distribution on S, and since there are three outcomes with one girl, we get

$$P(\text{one daughter}) = \frac{3}{8}$$

□

***Example* 1.4.4.** Consider a randomly chosen girl who has two siblings. What is the probability that she has no sisters?

Although this seems like the same problem as in the previous example, it is not. If, for example, the family has three girls, the chosen girl can be any of these three, so there are three different outcomes and the sample space needs to take this into account. Let g^* denote the chosen girl to get the sample space

$$S = \{g^*gg, gg^*g, ggg^*, g^*gb, gg^*b, g^*bg, gbg^*, bg^*g, bgg^*, g^*bb, bg^*b, bbg^*\}$$

and since 3 out of 12 equally likely outcomes have no sisters we get

$$P(\text{no sisters}) = \frac{1}{4}$$

which is smaller than the $\frac{3}{8}$ we got above. On average, 37.5% of families with three children have a single daughter and 25% of girls in three-children families are single daughters.

□

1.4.1 Combinatorics

Combinatorics, "the mathematics of counting," gives rise to a wealth of probability problems. The typical situation is that we have a set of objects from which we draw repeatedly in such a way that all objects are equally likely to be drawn. It is often tedious to list the sample space explicitly, but by counting combinations we can find the total number of cases and the number of favorable cases and apply the methods from the previous section.

The first problem is to find general expressions for the total number of combinations when we draw k times from a set of n distinguishable objects. There are

different ways to interpret this. For example, we can draw *with* or *without replacement*, depending on whether the same object can be drawn more than once. We can also draw *with* or *without regard to order*, depending on whether it matters in which order the objects are drawn. With these distinctions, there are four different cases, illustrated in the following simple example.

***Example* 1.4.5.** Choose two numbers from the set $\{1, 2, 3\}$ and list the possible outcomes.

Let us first choose with regard to order. If we choose with replacement, the possible outcomes are

$$(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)$$

and if we choose without replacement

$$(1,2), (1,3), (2,1), (2,3), (3,1), (3,2)$$

Next, let us choose without regard to order. This means that, for example, the outcomes $(1, 2)$ and $(2, 1)$ are regarded as the same and we denote it by $\{1, 2\}$ to stress that this is the *set* of 1 and 2, *not* the ordered pair. If we choose with replacement, the possible cases are

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

and if we choose without replacement

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

□

To find expressions in the four cases for arbitrary values of n and k, we first need the following result. It is intuitively quite clear, and we state it without proof.

Proposition 1.4.3. If we are to perform r experiments in order, such that there are n_1 possible outcomes of the first experiment, n_2 possible outcomes of the second experiment, ..., n_r possible outcomes of the rth experiment, then there is a total of $n_1 n_2 \cdots n_r$ outcomes of the sequence of the r experiments.

This is called the *fundamental principle of counting* or the *multiplication principle*. Let us illustrate it by a simple example.

***Example* 1.4.6.** A Swedish license plate consists of three letters followed by three digits. How many possible license plates are there?

Although there are 28 letters in the Swedish alphabet, only 23 are used for license plates. Hence we have $r = 6$, $n_1 = n_2 = n_3 = 23$, and $n_4 = n_5 = n_6 = 10$. This gives a total of $23^3 \times 10^3 \approx 12.2$ million different license plates. □

We can now address the problem of drawing k times from a set of n objects. It turns out that choosing with regard to order is the simplest, so let us start with this and first consider the case of choosing with replacement. The first object can be chosen in n ways, and for each such choice, we have n ways to choose also the second object, n ways to choose the third, and so on. The fundamental principle of counting gives

$$n \times n \times \cdots \times n = n^k$$

ways to choose with replacement and with regard to order.

If we instead choose without replacement, the first object can be chosen in n ways, the second in $n - 1$ ways, since the first object has been removed, the third in $n - 2$ ways and so on. The fundamental principle of counting gives

$$n(n-1)\cdots(n-k+1)$$

ways to choose without replacement and with regard to order. Sometimes the notation

$$(n)_k = n(n-1)\cdots(n-k+1)$$

will be used for convenience, but this is not standard.

***Example* 1.4.7.** From a group of 20 students, half of whom are female, a student council president and vice president are chosen at random. What is the probability of getting a female president and a male vice president?

The set of objects is the 20 students. Assuming that the president is drawn first, we need to take order into account, since, for example, (Brenda, Bruce) is a favorable outcome but (Bruce, Brenda) is not. Also, drawing is done without replacement. Thus, we have $k = 2$ and $n = 20$ and there are $20 \times 19 = 380$ equally likely different ways to choose a president and a vice president. The sample space is the set of these 380 combinations and to find the probability, we need the number of favorable cases. By the fundamental principle of counting, this is $10 \times 10 = 100$. The probability of getting a female president and male vice president is $\frac{100}{380} \approx 0.26$. □

***Example* 1.4.8.** A human gene consists of nucleotide base pairs of four different kinds, $A, C, G,$ and T. If a particular region of interest of a gene has 20 base pairs,

what is the probability that a randomly chosen individual has no base pairs in common with a particular reference sequence in a database?

The set of objects is $\{A, C, G, T\}$, and we draw 20 times with replacement and with regard to order. Thus $k = 20$ and $n = 4$, so there are 4^{20} possible outcomes, and let us, for the sake of this example, assume that they are equally likely (which would not be true in reality). For the number of favorable outcomes, $n = 3$ instead of 4 since we need to avoid one particular letter in each choice. Hence the probability is $3^{20}/4^{20} \approx 0.003$. □

***Example* 1.4.9. (*The Birthday Problem*).** This problem is a favorite in the probability literature. In a group of 100 people, what is the probability that at least two have the same birthday?

To simplify the solution, we disregard leap years and assume a uniform distribution of birthdays over the 365 days of the year. To assign birthdays to 100 people, we choose 100 out of 365 with replacement and get 365^{100} different combinations. The sample space is the set of those combinations, and the event of interest is

$$A = \{\text{at least two birthdays are equal}\}$$

and as it turns out, it is easier to deal with its complement

$$A^c = \{\text{all 100 birthdays are different}\}$$

To find the probability of A^c, note that the number of cases favorable to A^c is obtained by choosing 100 days out of 365 *without* replacement and hence

$$P(A) = 1 - P(A^c) = 1 - \frac{365 \times 364 \times \cdots \times 266}{365^{100}} \approx 0.9999997$$

Yes, that is a sequence of six 9s followed by a 7! Hence, we can be almost certain that any group of 100 people has at least two people sharing birthdays. A similar calculation reveals the probability of a shared birthday already exceeds $\frac{1}{2}$ at 23 people, a quite surprising result. About 50% of school classes thus ought to have kids who share birthdays, something that those with idle time on their hands can check empirically. □

A check of real-life birthday distributions will reveal that the assumption of birthdays being uniformly distributed over the year is not true. However, the already high probability of shared birthdays only gets higher with a nonuniform distribution. Intuitively, this is because the less uniform the distribution, the more difficult it becomes to avoid birthdays already taken. For an extreme example, suppose that everybody was born

in January, in which case there would be only 31 days to choose from instead of 365. Thus, in a group of 100 people, there would be absolute certainty of shared birthdays. Generally, it can be shown that the uniform distribution minimizes the probability of shared birthdays (we return to this in Problems 46 and 47).

***Example* 1.4.10. (*The Birthday Problem continued*).** A while ago I was in a group of exactly 100 people and asked for their birthdays. It turned out that nobody had the same birthday as I do. In the light of the previous problem, would this not be a very unlikely coincidence?

No, because here we are only considering the case of avoiding one particular birthday. Hence, with

$$B = \{\text{at least one out of } 99 \text{ birthdays is the same as mine}\}$$

we get

$$B^c = \{99 \text{ birthdays are different from mine}\}$$

and the number of cases favorable to B^c is obtained by choosing with replacement from the 364 days that do not match my birthday. We get

$$P(B) = 1 - P(B^c) = 1 - \frac{364^{99}}{365^{99}} \approx 0.24$$

Thus, it is actually quite likely that nobody shares my birthday, and it is at the same time almost certain that at least somebody shares somebody else's birthday. □

Next we turn to the case of choosing without regard to order. First, suppose that we choose without replacement and let x be the number of possible ways, in which this can be done. Now, there are $n(n-1)\cdots(n-k+1)$ ways to choose with regard to order and each such ordered set can be obtained by first choosing the objects and then order them. Since there are x ways to choose the unordered objects and $k!$ ways to order them, we get the relation

$$n(n-1)\cdots(n-k+1) = x \times k!$$

and hence there are

$$x = \frac{n(n-1)\cdots(n-k+1)}{k!} \tag{1.4.1}$$

ways to choose without replacement, without regard to order. In other words, this is the number of subsets of size k of a set of size n, called the *binomial coefficient*, read "n choose k" and usually denoted and defined as

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

but we use the expression in Equation (1.4.1) for computations. By convention,

$$\binom{n}{0} = 1$$

and from the definition it follows immediately that

$$\binom{n}{k} = \binom{n}{n-k}$$

which is useful for computations. For some further properties, see Problem 21.

***Example* 1.4.11.** In Texas Lotto, you choose five of the numbers $1, ..., 44$ and one bonus ball number, also from $1, ..., 44$. Winning numbers are chosen randomly. Which is more likely: that you match the first five numbers but not the bonus ball or that you match four of the first five numbers and the bonus ball?

Since we have to match five of our six numbers in each case, are the two not equally likely? Let us compute the probabilities and see. The set of objects is $\{1, 2, ..., 44\}$ and the first five numbers are drawn without replacement and without regard to order. Hence there are $\binom{44}{5}$ combinations and for each of these there are then 44 possible choices of the bonus ball. Thus, there is a total of $\binom{44}{5} \times 44 = 47,784,352$ different combinations. Introduce the events

$$\begin{aligned} A &= \{\text{match the first five numbers but not the bonus ball}\} \\ B &= \{\text{match four of the first five numbers and the bonus ball}\} \end{aligned}$$

For A, the number of favorable cases is 1×43 (only one way to match the first five numbers, 43 ways to avoid the winning bonus ball). Hence

$$P(A) = \frac{1 \times 43}{\binom{44}{5} \times 44} \approx 9 \times 10^{-7}$$

To find the number of cases favorable to B, note that there are $\binom{5}{4} = 5$ ways to match four out of five winning numbers and then $\binom{39}{1} = 39$ ways to avoid the fifth winning number. There is only one choice for the bonus ball and we get

$$P(B) = \frac{5 \times 39 \times 1}{\binom{44}{5} \times 44} \approx 4 \times 10^{-6}$$

so B is more than 4 times as likely as A. □

***Example* 1.4.12.** You are dealt a poker hand (5 cards out of 52 without replacement). **(a)** What is the probability that you get no hearts? **(b)** What is the probability that you get exactly k hearts? **(c)** What is the most likely number of hearts?

We will solve this by disregarding order. The number of possible cases is the number of ways in which we can choose 5 out of 52 cards, which equals $\binom{52}{5}$. In (a), to get a favorable case, we need to choose all 5 cards from the 39 that are not hearts. Since this can be done in $\binom{39}{5}$ ways, we get

$$P(\text{no hearts}) = \frac{\binom{39}{5}}{\binom{52}{5}} \approx 0.22$$

In (b), we need to choose k cards among the 13 hearts, and for each such choice, the remaining $5-k$ cards are chosen among the remaining 39 that are not hearts. This gives

$$P(k \text{ hearts}) = \frac{\binom{13}{k}\binom{39}{5-k}}{\binom{52}{5}}, \quad k = 0, 1, ..., 5$$

and for (c), direct computation gives the most likely number as 1, which has probability 0.41. □

The problem in the previous example can also be solved by taking order into account. Hence, we imagine that we get the cards one by one and list them in order and note that there are $(52)_5$ different cases. There are $(13)_k(39)_{5-k}$ ways to choose so that we get k hearts and $5-k$ nonhearts in a particular order. Since there are $\binom{5}{k}$ ways to choose position for the k hearts, we get

$$P(k \text{ hearts}) = \frac{\binom{5}{k}(13)_k(39)_{5-k}}{(52)_5}$$

which is the same as we got when we disregarded order above. It does not matter to the solution of the problem whether we take order into account, but we must be consistent and count the same way for the total and the favorable number of cases. In this particular example, it is probably easier to disregard order.

***Example* 1.4.13.** An urn contains 10 white balls, 10 red balls, and 10 black balls. You draw 5 balls at random without replacement. What is the probability that you do

not get all colors?

Introduce the events

$$R = \{\text{no red balls}\},\ \ W = \{\text{no white balls}\},\ \ B = \{\text{no black balls}\}$$

The event of interest is then $R \cup W \cup B$, and we will apply Proposition 1.3.3. First note that by symmetry, $P(R) = P(W) = P(B)$. Also, each intersection of any two events has the same probability and finally $R \cap W \cap B = \emptyset$. We get

$$P(\text{not all colors}) = 3P(R) - 3P(R \cap W)$$

In order to get no red balls, the 5 balls must be chosen among the 20 balls that are not red and hence

$$P(R) = \binom{20}{5} \Big/ \binom{30}{5}$$

Similarly, to get neither red, nor white balls, the 5 balls must be chosen among the black balls and

$$P(R \cap W) = \binom{10}{5} \Big/ \binom{30}{5}$$

We get

$$P(\text{not all colors}) \quad = \quad 3\left(\binom{20}{5} - \binom{10}{5}\right) \Big/ \binom{30}{5} \ \approx\ 0.32$$

□

The final case, choosing with replacement and without regard to order, turns out to be the trickiest. As we noted above, when we choose without replacement, each unordered set of k objects corresponds to exactly $k!$ ordered sets. The relation is not so simple when we choose with replacement. For example, the unordered set $\{1, 1\}$ corresponds to one ordered set $(1, 1)$, whereas the unordered set $\{1, 2\}$ corresponds to two ordered sets $(1, 2)$ and $(2, 1)$. To find the general expression, we need to take a less direct route.

Imagine a row of n slots, numbered from 1 to n and separated by single walls where slot number j represents the jth object.. Whenever object j is drawn, a ball is put in slot number j. After k draws, we will thus have k balls distributed over the n slots (and slots corresponding to objects never drawn are empty). The question now reduces to how many ways there are to distribute k balls over n slots. This is equivalent to rearranging the $n - 1$ inner walls and the k balls, which in turn is equivalent to choosing positions for the k balls from a total of $n - 1 + k$ positions. But this can be done in $\binom{n-1+k}{k}$ ways, and hence this is the number of ways to choose with replacement and without regard to order.

***Example* 1.4.14.** The Texas Lottery game "Pick 3" is played by picking three numbers with replacement from the numbers $0, 1, ..., 9$. You can play "exact order" or

"any order." With the "exact order" option, you win when your numbers match the winning numbers in the exact order they are drawn. With the "any order" option, you win whenever your numbers match the winning numbers in any order. How many possible winning combinations are there with the "any order" option?

We have $n = 10, k = 3$, and the winning numbers are chosen with replacement and without regard to order and hence there are

$$\binom{10-1+3}{3} = \binom{12}{3} = 220$$

possible winning combinations. □

***Example* 1.4.15.** Draw twice from the set $\{1, ..., 9\}$ at random with replacement. What is the probability that the two drawn numbers are equal?

We have $n = 9$ and $k = 2$. Taking order into account, there are $9 \times 9 = 81$ possible cases, 9 of which are favorable. Hence the probability is $\frac{9}{81} = \frac{1}{9}$. If we disregard order, we have $\binom{9-1+2}{2} = 45$ possible cases and still 9 favorable and the probability is $\frac{9}{45} = \frac{1}{5}$. Since whether we draw with or without regard to order does not seem to matter to the question, why do we get different results?

The problem is that in the second case, when we draw without regard to order, the *distribution is not uniform*. For example, the outcome $\{1, 2\}$ corresponds to the two equally likely ordered outcomes $(1, 2)$ and $(2, 1)$ and is thus twice as likely as the outcome $\{1, 1\}$, which corresponds to only one ordered outcome $(1, 1)$. Thus, the first solution $\frac{1}{9}$ is correct. □

Thus, when we draw with replacement but without regard to order, we must be careful when we compute probabilities, since the distribution is not uniform, as it is in the other three cases. Luckily, this case is far more uncommon in applications than are the other three cases. There is one interesting application, though, that has to do with the number of integer solutions to a certain type of equation. If we look again at the way in which we arrived at the formula and let x_j denote the number of balls in slot j, we realize that we must have $x_1 + \cdots + x_n = k$ and get the following observation.

Corollary 1.4.4. There are $\binom{n-1+k}{k}$ non-negative integer solutions $(x_1, ..., x_n)$ to the equation $x_1 + \cdots + x_n = k$.

The four different ways of choosing k out of n objects are summarized in Table 1.2. Note that when we choose without replacement, k must be less than or equal to n, but when we choose with replacement, there is no such restriction.

We finish with another favorite problem from the probability literature. It combines combinatorics with previous results concerning the probability of a union.

***Example* 1.4.16. (*The Matching Problem*).** The numbers $1, 2, ..., n$ are listed in random order. Whenever a number remains in its original position in the permutation, we call this a "match." For example, if $n = 5$, then there are two matches in the permutation 32541 and none in 23451. **(a)** What is the probability that there are no matches? **(b)** What happens to the probability in (a) as $n \to \infty$?

Before we solve this, let us try to think about part (b). Does it get easier or harder to avoid matches when n is large? It seems possible to argue for both. With so many choices, it is easy to avoid a match in each particular position. On the other hand, there are many positions to try, so it should not be too hard to get at least one match. It is not easy to have good intuition for what happens here.

To solve the problem, we first consider the complement of no matches and introduce the events

$$\begin{aligned} A &= \{\text{at least one match}\} \\ A_k &= \{\text{match in the } k\text{th draw}\}, \quad k = 1, 2, ..., n \end{aligned}$$

so that

$$A = \bigcup_{k=1}^{n} A_k$$

We will apply Proposition 1.3.4, so we need to figure out the probabilities of the events A_k as well as all intersections of two events, three events and so on.

First note that there are $n!$ different permutations of the numbers $1, 2, ..., n$. To get a match in position k, there is only one choice for that number and the rest can be

Table 1.2 Choosing k out of n objects

	With replacement	Without replacement
With regard to order	n^k	$n(n-1)\cdots(n-k+1)$
Without regard to order	$\binom{n-1+k}{k}$	$\binom{n}{k}$

ordered in $(n-1)!$ different ways. We get the probability

$$P(A_k) = \frac{\#\text{ favorable outcomes}}{\#\text{ possible outcomes}} = \frac{(n-1)!}{n!} = \frac{1}{n}$$

which means that the first sum in Proposition 1.3.4 equals 1. To get a match in both the ith and jth positions, we have only one choice for each of these two positions and the remaining $n-2$ numbers can be ordered in $(n-2)!$ ways and

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

Since there are $\binom{n}{2}$ ways to select two events A_i and A_j, we get, the following equation for the second sum in Proposition 1.3.4:

$$\begin{aligned} \sum_{i<j} P(A_i \cap A_j) &= \binom{n}{2} \frac{1}{n(n-1)} \\ &= \frac{n(n-1)}{2!} \times \frac{1}{n(n-1)} = \frac{1}{2!} \end{aligned}$$

Proceeding to the third sum, a similar argument gives that, for fixed $i < j < k$

$$\sum_{i<j<k} P(A_i \cap A_j \cap A_k) = \binom{n}{3} \times \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}$$

and the pattern emerges. The jth sum in Proposition 1.3.4 equals $1/j!$, and with the alternating signs we get

$$P(\text{at least one match}) = 1 - \sum_{j=2}^{n} \frac{(-1)^j}{j!} = 1 - \sum_{j=0}^{n} \frac{(-1)^j}{j!}$$

which finally gives

$$P(\text{no matches}) = \sum_{j=0}^{n} \frac{(-1)^j}{j!}$$

This is interesting. First, the probability is not monotone in n, so we cannot say that it gets easier or harder to avoid matches as n increases. Second, as $n \to \infty$, we recognize the limit as the Taylor expansion of e^{-1} and hence the probability of no matches converges to $e^{-1} \approx 0.37$ as $n \to \infty$. We can also note how rapid the convergence is; already for $n = 4$, the probability is 0.375. Thus, for all practical purposes, the probability to get no matches is 0.37 regardless of n. In Problem 32, you are asked to find the probability of exactly j matches. □