Chapter 1

Discrete Probability Distributions

1.1 Simulation of Discrete Probabilities

Probability

In this chapter, we shall first consider chance experiments with a finite number of possible outcomes $\omega_1, \omega_2, \ldots, \omega_n$. For example, we roll a die and the possible outcomes are 1, 2, 3, 4, 5, 6 corresponding to the side that turns up. We toss a coin with possible outcomes H (heads) and T (tails).

It is frequently useful to be able to refer to an outcome of an experiment. For example, we might want to write the mathematical expression which gives the sum of four rolls of a die. To do this, we could let X_i, $i = 1, 2, 3, 4$, represent the values of the outcomes of the four rolls, and then we could write the expression

$$X_1 + X_2 + X_3 + X_4$$

for the sum of the four rolls. The X_i's are called *random variables*. A random variable is simply an expression whose value is the outcome of a particular experiment. Just as in the case of other types of variables in mathematics, random variables can take on different values.

Let X be the random variable which represents the roll of one die. We shall assign probabilities to the possible outcomes of this experiment. We do this by assigning to each outcome ω_j a nonnegative number $m(\omega_j)$ in such a way that

$$m(\omega_1) + m(\omega_2) + \cdots + m(\omega_6) = 1 \ .$$

The function $m(\omega_j)$ is called the *distribution function* of the random variable X. For the case of the roll of the die we would assign equal probabilities or probabilities 1/6 to each of the outcomes. With this assignment of probabilities, one could write

$$P(X \le 4) = \frac{2}{3}$$

to mean that the probability is 2/3 that a roll of a die will have a value which does not exceed 4.

Let Y be the random variable which represents the toss of a coin. In this case, there are two possible outcomes, which we can label as H and T. Unless we have reason to suspect that the coin comes up one way more often than the other way, it is natural to assign the probability of 1/2 to each of the two outcomes.

In both of the above experiments, each outcome is assigned an equal probability. This would certainly not be the case in general. For example, if a drug is found to be effective 30 percent of the time it is used, we might assign a probability .3 that the drug is effective the next time it is used and .7 that it is not effective. This last example illustrates the intuitive *frequency concept of probability.* That is, if we have a probability p that an experiment will result in outcome A, then if we repeat this experiment a large number of times we should expect that the fraction of times that A will occur is about p. To check intuitive ideas like this, we shall find it helpful to look at some of these problems experimentally. We could, for example, toss a coin a large number of times and see if the fraction of times heads turns up is about 1/2. We could also simulate this experiment on a computer.

Simulation

We want to be able to perform an experiment that corresponds to a given set of probabilities; for example, $m(\omega_1) = 1/2$, $m(\omega_2) = 1/3$, and $m(\omega_3) = 1/6$. In this case, one could mark three faces of a six-sided die with an ω_1, two faces with an ω_2, and one face with an ω_3.

In the general case we assume that $m(\omega_1)$, $m(\omega_2)$, $\ldots$, $m(\omega_n)$ are all rational numbers, with least common denominator n. If $n > 2$, we can imagine a long cylindrical die with a cross-section that is a regular n-gon. If $m(\omega_j) = n_j/n$, then we can label n_j of the long faces of the cylinder with an ω_j, and if one of the end faces comes up, we can just roll the die again. If $n = 2$, a coin could be used to perform the experiment.

We will be particularly interested in repeating a chance experiment a large number of times. Although the cylindrical die would be a convenient way to carry out a few repetitions, it would be difficult to carry out a large number of experiments. Since the modern computer can do a large number of operations in a very short time, it is natural to turn to the computer for this task.

Random Numbers

We must first find a computer analog of rolling a die. This is done on the computer by means of a *random number generator.* Depending upon the particular software package, the computer can be asked for a real number between 0 and 1, or an integer in a given set of consecutive integers. In the first case, the real numbers are chosen in such a way that the probability that the number lies in any particular subinterval of this unit interval is equal to the length of the subinterval. In the second case, each integer has the same probability of being chosen.

.203309	.762057	.151121	.623868
.932052	.415178	.716719	.967412
.069664	.670982	.352320	.049723
.750216	.784810	.089734	.966730
.946708	.380365	.027381	.900794

Table 1.1: Sample output of the program **RandomNumbers**.

Let X be a random variable with distribution function $m(\omega)$, where ω is in the set $\{\omega_1, \omega_2, \omega_3\}$, and $m(\omega_1) = 1/2$, $m(\omega_2) = 1/3$, and $m(\omega_3) = 1/6$. If our computer package can return a random integer in the set $\{1, 2, ..., 6\}$, then we simply ask it to do so, and make 1, 2, and 3 correspond to ω_1, 4 and 5 correspond to ω_2, and 6 correspond to ω_3. If our computer package returns a random real number r in the interval $(0,\ 1)$, then the expression

$$\lfloor 6r \rfloor + 1$$

will be a random integer between 1 and 6. (The notation $\lfloor x \rfloor$ means the greatest integer not exceeding x, and is read "floor of x.")

The method by which random real numbers are generated on a computer is described in the historical discussion at the end of this section. The following example gives sample output of the program **RandomNumbers**.

Example 1.1 (Random Number Generation) The program **RandomNumbers** generates n random real numbers in the interval $[0, 1]$, where n is chosen by the user. When we ran the program with $n = 20$, we obtained the data shown in Table 1.1. □

Example 1.2 (Coin Tossing) As we have noted, our intuition suggests that the probability of obtaining a head on a single toss of a coin is 1/2. To have the computer toss a coin, we can ask it to pick a random real number in the interval $[0, 1]$ and test to see if this number is less than 1/2. If so, we shall call the outcome *heads*; if not we call it *tails.* Another way to proceed would be to ask the computer to pick a random integer from the set $\{0, 1\}$. The program **CoinTosses** carries out the experiment of tossing a coin n times. Running this program, with $n = 20$, resulted in:

THTTTHTTTTHTTTTTHHTT.

Note that in 20 tosses, we obtained 5 heads and 15 tails. Let us toss a coin n times, where n is much larger than 20, and see if we obtain a proportion of heads closer to our intuitive guess of 1/2. The program **CoinTosses** keeps track of the number of heads. When we ran this program with $n = 1000$, we obtained 494 heads. When we ran it with $n = 10000$, we obtained 5039 heads.

We notice that when we tossed the coin 10,000 times, the proportion of heads was close to the "true value" .5 for obtaining a head when a coin is tossed. A mathematical model for this experiment is called Bernoulli Trials (see Chapter 3). The *Law of Large Numbers,* which we shall study later (see Chapter 8), will show that in the Bernoulli Trials model, the proportion of heads should be near .5, consistent with our intuitive idea of the frequency interpretation of probability.

Of course, our program could be easily modified to simulate coins for which the probability of a head is p, where p is a real number between 0 and 1. □

In the case of coin tossing, we already knew the probability of the event occurring on each experiment. The real power of simulation comes from the ability to estimate probabilities when they are not known ahead of time. This method has been used in the recent discoveries of strategies that make the casino game of blackjack favorable to the player. We illustrate this idea in a simple situation in which we can compute the true probability and see how effective the simulation is.

Example 1.3 (Dice Rolling) We consider a dice game that played an important role in the historical development of probability. The famous letters between Pascal and Fermat, which many believe started a serious study of probability, were instigated by a request for help from a French nobleman and gambler, Chevalier de Méré. It is said that de Méré had been betting that, in four rolls of a die, at least one six would turn up. He was winning consistently and, to get more people to play, he changed the game to bet that, in 24 rolls of two dice, a pair of sixes would turn up. It is claimed that de Méré lost with 24 and felt that 25 rolls were necessary to make the game favorable. It was *un grand scandale* that mathematics was wrong.

We shall try to see if de Méré is correct by simulating his various bets. The program **DeMere1** simulates a large number of experiments, seeing, in each one, if a six turns up in four rolls of a die. When we ran this program for 1000 plays, a six came up in the first four rolls 48.6 percent of the time. When we ran it for 10,000 plays this happened 51.98 percent of the time.

We note that the result of the second run suggests that de Méré was correct in believing that his bet with one die was favorable; however, if we had based our conclusion on the first run, we would have decided that he was wrong. *Accurate results by simulation require a large number of experiments.* □

The program **DeMere2** simulates de Méré's second bet that a pair of sixes will occur in n rolls of a pair of dice. The previous simulation shows that it is important to know how many trials we should simulate in order to expect a certain degree of accuracy in our approximation. We shall see later that in these types of experiments, a rough rule of thumb is that, at least 95% of the time, the error does not exceed the reciprocal of the square root of the number of trials. Fortunately, for this dice game, it will be easy to compute the exact probabilities. We shall show in the next section that for the first bet the probability that de Méré wins is $1 - (5/6)^4 = .518$.

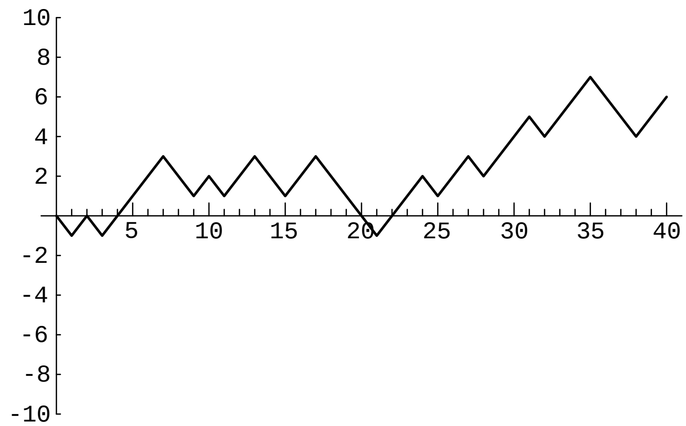


Figure 1.1: Peter's winnings in 40 plays of heads or tails.

One can understand this calculation as follows: The probability that no 6 turns up on the first toss is (5/6). The probability that no 6 turns up on either of the first two tosses is $(5/6)^2$. Reasoning in the same way, the probability that no 6 turns up on any of the first four tosses is $(5/6)^4$. Thus, the probability of at least one 6 in the first four tosses is $1 - (5/6)^4$. Similarly, for the second bet, with 24 rolls, the probability that de Méré wins is $1 - (35/36)^{24} = .491$, and for 25 rolls it is $1 - (35/36)^{25} = .506$.

Using the rule of thumb mentioned above, it would require 27,000 rolls to have a reasonable chance to determine these probabilities with sufficient accuracy to assert that they lie on opposite sides of .5. It is interesting to ponder whether a gambler can detect such probabilities with the required accuracy from gambling experience. Some writers on the history of probability suggest that de Méré was, in fact, just interested in these problems as intriguing probability problems.

Example 1.4 (Heads or Tails) For our next example, we consider a problem where the exact answer is difficult to obtain but for which simulation easily gives the qualitative results. Peter and Paul play a game called *heads or tails.* In this game, a fair coin is tossed a sequence of times—we choose 40. Each time a head comes up Peter wins 1 penny from Paul, and each time a tail comes up Peter loses 1 penny to Paul. For example, if the results of the 40 tosses are

THTHHHHTTHTHHTTHHTTTTHHHTHHTHHHTHHHTTTHH.

Peter's winnings may be graphed as in Figure 1.1.

Peter has won 6 pennies in this particular game. It is natural to ask for the probability that he will win j pennies; here j could be any even number from -40 to 40. It is reasonable to guess that the value of j with the highest probability is $j = 0$, since this occurs when the number of heads equals the number of tails. Similarly, we would guess that the values of j with the lowest probabilities are $j = \pm 40$.

A second interesting question about this game is the following: How many times in the 40 tosses will Peter be in the lead? Looking at the graph of his winnings (Figure 1.1), we see that Peter is in the lead when his winnings are positive, but we have to make some convention when his winnings are 0 if we want all tosses to contribute to the number of times in the lead. We adopt the convention that, when Peter's winnings are 0, he is in the lead if he was ahead at the previous toss and not if he was behind at the previous toss. With this convention, Peter is in the lead 34 times in our example. Again, our intuition might suggest that the most likely number of times to be in the lead is 1/2 of 40, or 20, and the least likely numbers are the extreme cases of 40 or 0.

It is easy to settle this by simulating the game a large number of times and keeping track of the number of times that Peter's final winnings are j, and the number of times that Peter ends up being in the lead by k. The proportions over all games then give estimates for the corresponding probabilities. The program **HTSimulation** carries out this simulation. Note that when there are an even number of tosses in the game, it is possible to be in the lead only an even number of times. We have simulated this game 10,000 times. The results are shown in Figures 1.2 and 1.3. These graphs, which we call spike graphs, were generated using the program **Spikegraph**. The vertical line, or spike, at position x on the horizontal axis, has a height equal to the proportion of outcomes which equal x. Our intuition about Peter's final winnings was quite correct, but our intuition about the number of times Peter was in the lead was completely wrong. The simulation suggests that the least likely number of times in the lead is 20 and the most likely is 0 or 40. This is indeed correct, and the explanation for it is suggested by playing the game of heads or tails with a large number of tosses and looking at a graph of Peter's winnings. In Figure 1.4 we show the results of a simulation of the game, for 1000 tosses and in Figure 1.5 for 10,000 tosses.

In the second example Peter was ahead most of the time. It is a remarkable fact, however, that, if play is continued long enough, Peter's winnings will continue to come back to 0, but there will be very long times between the times that this happens. These and related results will be discussed in Chapter 12. □

In all of our examples so far, we have simulated equiprobable outcomes. We illustrate next an example where the outcomes are not equiprobable.

Example 1.5 (Horse Races) Four horses (Acorn, Balky, Chestnut, and Dolby) have raced many times. It is estimated that Acorn wins 30 percent of the time, Balky 40 percent of the time, Chestnut 20 percent of the time, and Dolby 10 percent of the time.

We can have our computer carry out one race as follows: Choose a random number x. If $x < .3$ then we say that Acorn won. If $.3 \leq x < .7$ then Balky wins. If $.7 \leq x < .9$ then Chestnut wins. Finally, if $.9 \leq x$ then Dolby wins.

The program **HorseRace** uses this method to simulate the outcomes of n races. Running this program for $n = 10$ we found that Acorn won 40 percent of the time, Balky 20 percent of the time, Chestnut 10 percent of the time, and Dolby 30 percent

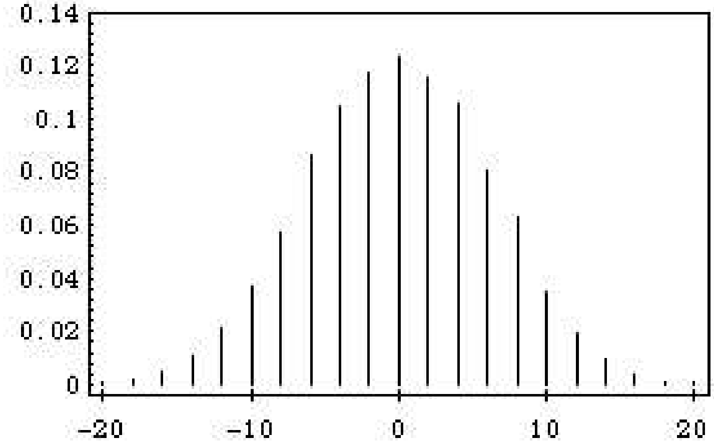


Figure 1.2: Distribution of winnings.

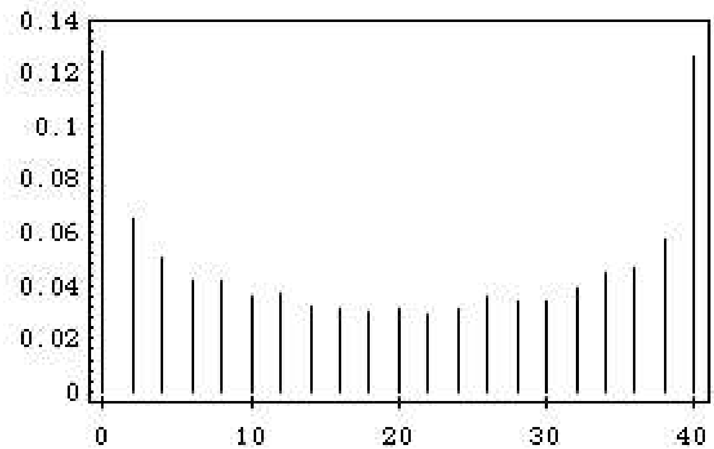


Figure 1.3: Distribution of number of times in the lead.

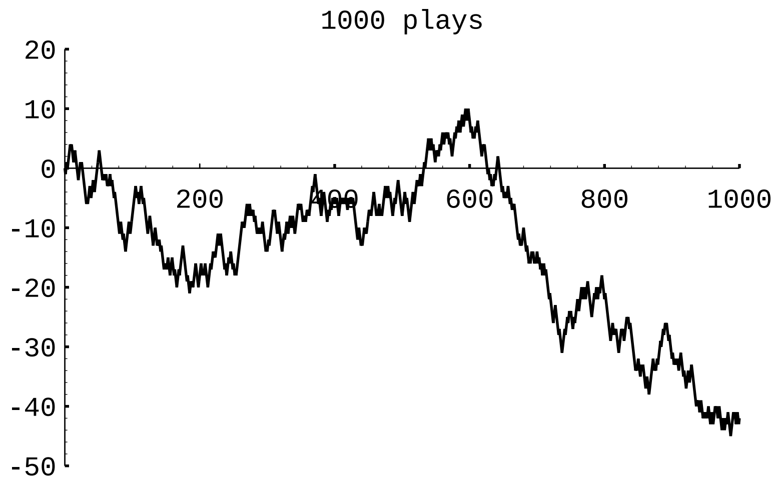


Figure 1.4: Peter's winnings in 1000 plays of heads or tails.

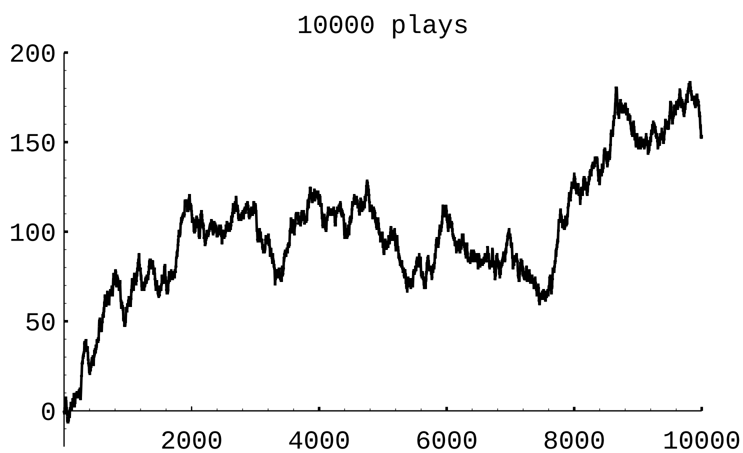


Figure 1.5: Peter's winnings in 10,000 plays of heads or tails.

of the time. A larger number of races would be necessary to have better agreement with the past experience. Therefore we ran the program to simulate 1000 races with our four horses. Although very tired after all these races, they performed in a manner quite consistent with our estimates of their abilities. Acorn won 29.8 percent of the time, Balky 39.4 percent, Chestnut 19.5 percent, and Dolby 11.3 percent of the time.

The program **GeneralSimulation** uses this method to simulate repetitions of an arbitrary experiment with a finite number of outcomes occurring with known probabilities. □

Historical Remarks

Anyone who plays the same chance game over and over is really carrying out a simulation, and in this sense the process of simulation has been going on for centuries. As we have remarked, many of the early problems of probability might well have been suggested by gamblers' experiences.

It is natural for anyone trying to understand probability theory to try simple experiments by tossing coins, rolling dice, and so forth. The naturalist Buffon tossed a coin 4040 times, resulting in 2048 heads and 1992 tails. He also estimated the number π by throwing needles on a ruled surface and recording how many times the needles crossed a line (see Section 2.1). The English biologist W. F. R. Weldon[1] recorded 26,306 throws of 12 dice, and the Swiss scientist Rudolf Wolf[2] recorded 100,000 throws of a single die without a computer. Such experiments are very time-consuming and may not accurately represent the chance phenomena being studied. For example, for the dice experiments of Weldon and Wolf, further analysis of the recorded data showed a suspected bias in the dice. The statistician Karl Pearson analyzed a large number of outcomes at certain roulette tables and suggested that the wheels were biased. He wrote in 1894:

> Clearly, since the Casino does not serve the valuable end of huge laboratory for the preparation of probability statistics, it has no scientific *raison d'être*. Men of science cannot have their most refined theories disregarded in this shameless manner! The French Government must be urged by the hierarchy of science to close the gaming-saloons; it would be, of course, a graceful act to hand over the remaining resources of the Casino to the Académie des Sciences for the endowment of a laboratory of orthodox probability; in particular, of the new branch of that study, the application of the theory of chance to the biological problems of evolution, which is likely to occupy so much of men's thoughts in the near future.[3]

However, these early experiments were suggestive and led to important discoveries in probability and statistics. They led Pearson to the *chi-squared test,* which

[1] T. C. Fry, *Probability and Its Engineering Uses,* 2nd ed. (Princeton: Van Nostrand, 1965).

[2] E. Czuber, *Wahrscheinlichkeitsrechnung,* 3rd ed. (Berlin: Teubner, 1914).

[3] K. Pearson, "Science and Monte Carlo," *Fortnightly Review*, vol. 55 (1894), p. 193; cited in S. M. Stigler, *The History of Statistics* (Cambridge: Harvard University Press, 1986).

is of great importance in testing whether observed data fit a given probability distribution.

By the early 1900s it was clear that a better way to generate random numbers was needed. In 1927, L. H. C. Tippett published a list of 41,600 digits obtained by selecting numbers haphazardly from census reports. In 1955, RAND Corporation printed a table of 1,000,000 random numbers generated from electronic noise. The advent of the high-speed computer raised the possibility of generating random numbers directly on the computer, and in the late 1940s John von Neumann suggested that this be done as follows: Suppose that you want a random sequence of four-digit numbers. Choose any four-digit number, say 6235, to start. Square this number to obtain 38,875,225. For the second number choose the middle four digits of this square (i.e., 8752). Do the same process starting with 8752 to get the third number, and so forth.

More modern methods involve the concept of modular arithmetic. If a is an integer and m is a positive integer, then by $a \pmod{m}$ we mean the remainder when a is divided by m. For example, $10 \pmod 4 = 2$, $8 \pmod 2 = 0$, and so forth. To generate a random sequence $X_0, X_1, X_2, \ldots$ of numbers choose a starting number X_0 and then obtain the numbers X_{n+1} from X_n by the formula

$$X_{n+1} = (aX_n + c) \pmod{m} ,$$

where a, c, and m are carefully chosen constants. The sequence $X_0, X_1, X_2, \ldots$ is then a sequence of integers between 0 and $m - 1$. To obtain a sequence of real numbers in $[0, 1)$, we divide each X_j by m. The resulting sequence consists of rational numbers of the form j/m, where $0 \leq j \leq m - 1$. Since m is usually a very large integer, we think of the numbers in the sequence as being random real numbers in $[0, 1)$.

For both von Neumann's squaring method and the modular arithmetic technique the sequence of numbers is actually completely determined by the first number. Thus, there is nothing really random about these sequences. However, they produce numbers that behave very much as theory would predict for random experiments. To obtain different sequences for different experiments the initial number X_0 is chosen by some other procedure that might involve, for example, the time of day.[4]

During the Second World War, physicists at the Los Alamos Scientific Laboratory needed to know, for purposes of shielding, how far neutrons travel through various materials. This question was beyond the reach of theoretical calculations. Daniel McCracken, writing in the *Scientific American*, states:

> The physicists had most of the necessary data: they knew the average distance a neutron of a given speed would travel in a given substance before it collided with an atomic nucleus, what the probabilities were that the neutron would bounce off instead of being absorbed by the nucleus, how much energy the neutron was likely to lose after a given

[4]For a detailed discussion of random numbers, see D. E. Knuth, *The Art of Computer Programming,* vol. II (Reading: Addison-Wesley, 1969).

> collision and so on.[5]

John von Neumann and Stanislas Ulam suggested that the problem be solved by modeling the experiment by chance devices on a computer. Their work being secret, it was necessary to give it a code name. Von Neumann chose the name "Monte Carlo." Since that time, this method of simulation has been called the *Monte Carlo Method.*

William Feller indicated the possibilities of using computer simulations to illustrate basic concepts in probability in his book *An Introduction to Probability Theory and Its Applications.* In discussing the problem about the number of times in the lead in the game of "heads or tails" Feller writes:

> The results concerning fluctuations in coin tossing show that widely held beliefs about the law of large numbers are fallacious. These results are so amazing and so at variance with common intuition that even sophisticated colleagues doubted that coins actually misbehave as theory predicts. The record of a simulated experiment is therefore included.[6]

Feller provides a plot showing the result of 10,000 plays of *heads or tails* similar to that in Figure 1.5.

The martingale betting system described in Exercise 10 has a long and interesting history. Russell Barnhart pointed out to the authors that its use can be traced back at least to 1754, when Casanova, writing in his memoirs, *History of My Life,* writes

> She [Casanova's mistress] made me promise to go to the casino [the Ridotto in Venice] for money to play in partnership with her. I went there and took all the gold I found, and, determinedly doubling my stakes according to the system known as the martingale, I won three or four times a day during the rest of the Carnival. I never lost the sixth card. If I had lost it, I should have been out of funds, which amounted to two thousand zecchini.[7]

Even if there were no zeros on the roulette wheel so the game was perfectly fair, the martingale system, or any other system for that matter, cannot make the game into a favorable game. The idea that a fair game remains fair and unfair games remain unfair under gambling systems has been exploited by mathematicians to obtain important results in the study of probability. We will introduce the general concept of a martingale in Chapter 6.

The word *martingale* itself also has an interesting history. The origin of the word is obscure. A recent version of the *Oxford English Dictionary* gives examples

[5]D. D. McCracken, "The Monte Carlo Method," *Scientific American,* vol. 192 (May 1955), p. 90.

[6]W. Feller, *Introduction to Probability Theory and its Applications,* vol. 1, 3rd ed. (New York: John Wiley & Sons, 1968), p. xi.

[7]G. Casanova, *History of My Life,* vol. IV, Chap. 7, trans. W. R. Trask (New York: Harcourt-Brace, 1968), p. 124.

of its use in the early 1600s and says that its probable origin is the reference in Rabelais's Book One, Chapter 20:

> Everything was done as planned, the only thing being that Gargantua doubted if they would be able to find, right away, breeches suitable to the old fellow's legs; he was doubtful, also, as to what cut would be most becoming to the orator—the martingale, which has a draw-bridge effect in the seat, to permit doing one's business more easily; the sailor-style, which affords more comfort for the kidneys; the Swiss, which is warmer on the belly; or the codfish-tail, which is cooler on the loins.[8]

Dominic Lusinchi noted an earlier occurrence of the word martingale. According to the French dictionary *Le Petit Robert*, the word comes from the Provençal word "martegalo," which means "from Martigues." Martigues is a town due west of Merseille. The dictionary gives the example of "chausses à la martinguale" (which means Martigues-style breeches) and the date 1491.

In modern uses martingale has several different meanings, all related to *holding down,* in addition to the gambling use. For example, it is a strap on a horse's harness used to hold down the horse's head, and also part of a sailing rig used to hold down the bowsprit.

The Labouchere system described in Exercise 9 is named after Henry du Pre Labouchere (1831–1912), an English journalist and member of Parliament. Labouchere attributed the system to Condorcet. Condorcet (1743–1794) was a political leader during the time of the French revolution who was interested in applying probability theory to economics and politics. For example, he calculated the probability that a jury using majority vote will give a correct decision if each juror has the same probability of deciding correctly. His writings provided a wealth of ideas on how probability might be applied to human affairs.[9]

Exercises

1 Modify the program **CoinTosses** to toss a coin n times and print out after every 100 tosses the proportion of heads minus 1/2. Do these numbers appear to approach 0 as n increases? Modify the program again to print out, every 100 times, both of the following quantities: the proportion of heads minus 1/2, and the number of heads minus half the number of tosses. Do these numbers appear to approach 0 as n increases?

2 Modify the program **CoinTosses** so that it tosses a coin n times and records whether or not the proportion of heads is within .1 of .5 (i.e., between .4 and .6). Have your program repeat this experiment 100 times. About how large must n be so that approximately 95 out of 100 times the proportion of heads is between .4 and .6?

[8]Quoted in the *Portable Rabelais,* ed. S. Putnam (New York: Viking, 1946), p. 113.

[9]Le Marquise de Condorcet, *Essai sur l'Application de l'Analyse à la Probabilité dès Décisions Rendues a la Pluralité des Voix* (Paris: Imprimerie Royale, 1785).

3 In the early 1600s, Galileo was asked to explain the fact that, although the number of triples of integers from 1 to 6 with sum 9 is the same as the number of such triples with sum 10, when three dice are rolled, a 9 seemed to come up less often than a 10—supposedly in the experience of gamblers.

(a) Write a program to simulate the roll of three dice a large number of times and keep track of the proportion of times that the sum is 9 and the proportion of times it is 10.

(b) Can you conclude from your simulations that the gamblers were correct?

4 In raquetball, a player continues to serve as long as she is winning; a point is scored only when a player is serving and wins the volley. The first player to win 21 points wins the game. Assume that you serve first and have a probability .6 of winning a volley when you serve and probability .5 when your opponent serves. Estimate, by simulation, the probability that you will win a game.

5 Consider the bet that all three dice will turn up sixes at least once in n rolls of three dice. Calculate $f(n)$, the probability of at least one triple-six when three dice are rolled n times. Determine the smallest value of n necessary for a favorable bet that a triple-six will occur when three dice are rolled n times. (DeMoivre would say it should be about $216 \log 2 = 149.7$ and so would answer 150—see Exercise 1.2.17. Do you agree with him?)

6 In Las Vegas, a roulette wheel has 38 slots numbered 0, 00, 1, 2, ..., 36. The 0 and 00 slots are green and half of the remaining 36 slots are red and half are black. A croupier spins the wheel and throws in an ivory ball. If you bet 1 dollar on red, you win 1 dollar if the ball stops in a red slot and otherwise you lose 1 dollar. Write a program to find the total winnings for a player who makes 1000 bets on red.

7 Another form of bet for roulette is to bet that a specific number (say 17) will turn up. If the ball stops on your number, you get your dollar back plus 35 dollars. If not, you lose your dollar. Write a program that will plot your winnings when you make 500 plays of roulette at Las Vegas, first when you bet each time on red (see Exercise 6), and then for a second visit to Las Vegas when you make 500 plays betting each time on the number 17. What differences do you see in the graphs of your winnings on these two occasions?

8 An astute student noticed that, in our simulation of the game of heads or tails (see Example 1.4), the proportion of times the player is always in the lead is very close to the proportion of times that the player's total winnings end up 0. Work out these probabilities by enumeration of all cases for two tosses and for four tosses, and see if you think that these probabilities are, in fact, the same.

9 The *Labouchere system* for roulette is played as follows. Write down a list of numbers, usually 1, 2, 3, 4. Bet the sum of the first and last, $1 + 4 = 5$, on red. If you win, delete the first and last numbers from your list. If you lose, add the amount that you last bet to the end of your list. Then use the new list and bet the sum of the first and last numbers (if there is only one number, bet that amount). Continue until your list becomes empty. Show that, if this happens, you win the sum, $1 + 2 + 3 + 4 = 10$, of your original list. Simulate this system and see if you do always stop and, hence, always win. If so, why is this not a foolproof gambling system?

10 Another well-known gambling system is the *martingale doubling system*. Suppose that you are betting on red to turn up in roulette. Every time you win, bet 1 dollar next time. Every time you lose, double your previous bet. Suppose that you use this system until you have won at least 5 dollars or you have lost more than 100 dollars. Write a program to simulate this and play it a number of times and see how you do. In his book *The Newcomes,* W. M. Thackeray remarks "You have not played as yet? Do not do so; above all avoid a martingale if you do."[10] Was this good advice?

11 Modify the program **HTSimulation** so that it keeps track of the maximum of Peter's winnings in each game of 40 tosses. Have your program print out the proportion of times that your total winnings take on values $0,\ 2,\ 4,\ \ldots,\ 40$. Calculate the corresponding exact probabilities for games of two tosses and four tosses.

12 In an upcoming national election for the President of the United States, a pollster plans to predict the winner of the popular vote by taking a random sample of 1000 voters and declaring that the winner will be the one obtaining the most votes in his sample. Suppose that 48 percent of the voters plan to vote for the Republican candidate and 52 percent plan to vote for the Democratic candidate. To get some idea of how reasonable the pollster's plan is, write a program to make this prediction by simulation. Repeat the simulation 100 times and see how many times the pollster's prediction would come true. Repeat your experiment, assuming now that 49 percent of the population plan to vote for the Republican candidate; first with a sample of 1000 and then with a sample of 3000. (The Gallup Poll uses about 3000.) (This idea is discussed further in Chapter 9, Section 9.1.)

13 The psychologist Tversky and his colleagues[11] say that about four out of five people will answer (a) to the following question:

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital 15 babies are born each day. Although the overall proportion of boys is about 50 percent, the actual proportion at either hospital may be more or less than 50 percent on any day.

[10] W. M. Thackerey, *The Newcomes* (London: Bradbury and Evans, 1854–55).

[11] See K. McKean, "Decisions, Decisions," *Discover,* June 1985, pp. 22–31. Kevin McKean, Discover Magazine, ©1987 Family Media, Inc. Reprinted with permission. This popular article reports on the work of Tverksy et. al. in *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982).

At the end of a year, which hospital will have the greater number of days on which more than 60 percent of the babies born were boys?

(a) the large hospital

(b) the small hospital

(c) neither—the number of days will be about the same.

Assume that the probability that a baby is a boy is .5 (actual estimates make this more like .513). Decide, by simulation, what the right answer is to the question. Can you suggest why so many people go wrong?

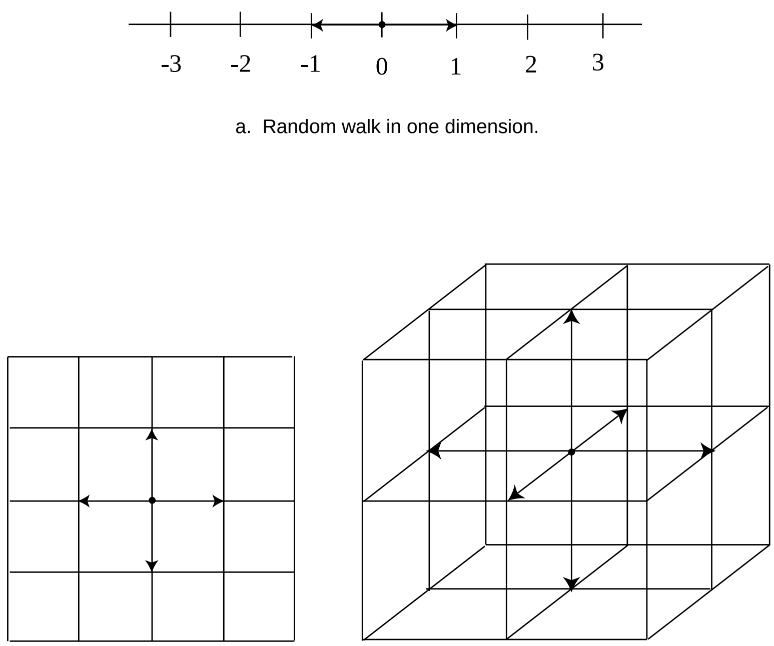
14 You are offered the following game. A fair coin will be tossed until the first time it comes up heads. If this occurs on the jth toss you are paid 2^j dollars. You are sure to win at least 2 dollars so you should be willing to pay to play this game—but how much? Few people would pay as much as 10 dollars to play this game. See if you can decide, by simulation, a reasonable amount that you would be willing to pay, per game, if you will be allowed to make a large number of plays of the game. Does the amount that you would be willing to pay per game depend upon the number of plays that you will be allowed?

15 Tversky and his colleagues[12] studied the records of 48 of the Philadelphia 76ers basketball games in the 1980–81 season to see if a player had times when he was hot and every shot went in, and other times when he was cold and barely able to hit the backboard. The players estimated that they were about 25 percent more likely to make a shot after a hit than after a miss. In fact, the opposite was true—the 76ers were 6 percent more likely to score after a miss than after a hit. Tversky reports that the number of hot and cold streaks was about what one would expect by purely random effects. Assuming that a player has a fifty-fifty chance of making a shot and makes 20 shots a game, estimate by simulation the proportion of the games in which the player will have a streak of 5 or more hits.

16 Estimate, by simulation, the average number of children there would be in a family if all people had children until they had a boy. Do the same if all people had children until they had at least one boy and at least one girl. How many more children would you expect to find under the second scheme than under the first in 100,000 families? (Assume that boys and girls are equally likely.)

17 Mathematicians have been known to get some of the best ideas while sitting in a cafe, riding on a bus, or strolling in the park. In the early 1900s the famous mathematician George Pólya lived in a hotel near the woods in Zurich. He liked to walk in the woods and think about mathematics. Pólya describes the following incident:

[12] ibid.



a. Random walk in one dimension.

b. Random walk in two dimensions.

c. Random walk in three dimensions.

Figure 1.6: Random walk.

> At the hotel there lived also some students with whom I usually took my meals and had friendly relations. On a certain day one of them expected the visit of his fiancée, what (sic) I knew, but I did not foresee that he and his fiancée would also set out for a stroll in the woods, and then suddenly I met them there. And then I met them the same morning repeatedly, I don't remember how many times, but certainly much too often and I felt embarrassed: It looked as if I was snooping around which was, I assure you, not the case.[13]

This set him to thinking about whether random walkers were destined to meet.

Pólya considered random walkers in one, two, and three dimensions. In one dimension, he envisioned the walker on a very long street. At each intersection the walker flips a fair coin to decide which direction to walk next (see Figure 1.6a). In two dimensions, the walker is walking on a grid of streets, and at each intersection he chooses one of the four possible directions with equal probability (see Figure 1.6b). In three dimensions (we might better speak of a random climber), the walker moves on a three-dimensional grid, and at each intersection there are now six different directions that the walker may choose, each with equal probability (see Figure 1.6c).

The reader is referred to Section 12.1, where this and related problems are discussed.

(a) Write a program to simulate a random walk in one dimension starting at 0. Have your program print out the lengths of the times between returns to the starting point (returns to 0). See if you can guess from this simulation the answer to the following question: Will the walker always return to his starting point eventually or might he drift away forever?

(b) The paths of two walkers in two dimensions who meet after n steps can be considered to be a single path that starts at $(0, 0)$ and returns to $(0, 0)$ after $2n$ steps. This means that the probability that two random walkers in two dimensions meet is the same as the probability that a single walker in two dimensions ever returns to the starting point. Thus the question of whether two walkers are sure to meet is the same as the question of whether a single walker is sure to return to the starting point.

Write a program to simulate a random walk in two dimensions and see if you think that the walker is sure to return to $(0, 0)$. If so, Pólya would be sure to keep meeting his friends in the park. Perhaps by now you have conjectured the answer to the question: Is a random walker in one or two dimensions sure to return to the starting point? Pólya answered

[13]G. Pólya, "Two Incidents," *Scientists at Work: Festschrift in Honour of Herman Wold,* ed. T. Dalenius, G. Karlsson, and S. Malmquist (Uppsala: Almquist & Wiksells Boktryckeri AB, 1970).

this question for dimensions one, two, and three. He established the remarkable result that the answer is *yes* in one and two dimensions and *no* in three dimensions.

(c) Write a program to simulate a random walk in three dimensions and see whether, from this simulation and the results of (a) and (b), you could have guessed Pólya's result.

1.2 Discrete Probability Distributions

In this book we shall study many different experiments from a probabilistic point of view. What is involved in this study will become evident as the theory is developed and examples are analyzed. However, the overall idea can be described and illustrated as follows: to each experiment that we consider there will be associated a random variable, which represents the outcome of any particular experiment. The set of possible outcomes is called the *sample space*. In the first part of this section, we will consider the case where the experiment has only finitely many possible outcomes, i.e., the sample space is finite. We will then generalize to the case that the sample space is either finite or countably infinite. This leads us to the following definition.

Random Variables and Sample Spaces

Definition 1.1 Suppose we have an experiment whose outcome depends on chance. We represent the outcome of the experiment by a capital Roman letter, such as X, called a *random variable*. The *sample space* of the experiment is the set of all possible outcomes. If the sample space is either finite or countably infinite, the random variable is said to be *discrete*. □

We generally denote a sample space by the capital Greek letter Ω. As stated above, in the correspondence between an experiment and the mathematical theory by which it is studied, the sample space Ω corresponds to the set of possible outcomes of the experiment.

We now make two additional definitions. These are subsidiary to the definition of sample space and serve to make precise some of the common terminology used in conjunction with sample spaces. First of all, we define the elements of a sample space to be *outcomes*. Second, each subset of a sample space is defined to be an *event*. Normally, we shall denote outcomes by lower case letters and events by capital letters.

Example 1.6 A die is rolled once. We let X denote the outcome of this experiment. Then the sample space for this experiment is the 6-element set

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

where each outcome i, for $i = 1, \ldots, 6$, corresponds to the number of dots on the face which turns up. The event

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

corresponds to the statement that the result of the roll is an even number. The event E can also be described by saying that X is even. Unless there is reason to believe the die is loaded, the natural assumption is that every outcome is equally likely. Adopting this convention means that we assign a probability of 1/6 to each of the six outcomes, i.e., $m(i) = 1/6$, for $1 \leq i \leq 6$. □

Distribution Functions

We next describe the assignment of probabilities. The definitions are motivated by the example above, in which we assigned to each outcome of the sample space a nonnegative number such that the sum of the numbers assigned is equal to 1.

Definition 1.2 Let X be a random variable which denotes the value of the outcome of a certain experiment, and assume that this experiment has only finitely many possible outcomes. Let Ω be the sample space of the experiment (i.e., the set of all possible values of X, or equivalently, the set of all possible outcomes of the experiment.) A *distribution function* for X is a real-valued function m whose domain is Ω and which satisfies:

1. $m(\omega) \geq 0 \ , \qquad \text{for all } \ \omega \in \Omega \ , \text{ and}$

2. $\sum_{\omega \in \Omega} m(\omega) = 1 \ .$

For any subset E of Ω, we define the *probability* of E to be the number $P(E)$ given by

$$P(E) = \sum_{\omega \in E} m(\omega) \ .$$

□

Example 1.7 Consider an experiment in which a coin is tossed twice. Let X be the random variable which corresponds to this experiment. We note that there are several ways to record the outcomes of this experiment. We could, for example, record the two tosses, in the order in which they occurred. In this case, we have $\Omega =$\{HH,HT,TH,TT\}. We could also record the outcomes by simply noting the number of heads that appeared. In this case, we have $\Omega =$\{0,1,2\}. Finally, we could record the two outcomes, without regard to the order in which they occurred. In this case, we have $\Omega =$\{HH,HT,TT\}.

We will use, for the moment, the first of the sample spaces given above. We will assume that all four outcomes are equally likely, and define the distribution function $m(\omega)$ by

$$m(\mathrm{HH}) = m(\mathrm{HT}) = m(\mathrm{TH}) = m(\mathrm{TT}) = \frac{1}{4} \ .$$

Let E ={HH,HT,TH} be the event that at least one head comes up. Then, the probability of E can be calculated as follows:

$$\begin{aligned} P(E) &= m(\mathrm{HH}) + m(\mathrm{HT}) + m(\mathrm{TH}) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \ . \end{aligned}$$

Similarly, if F ={HH,HT} is the event that heads comes up on the first toss, then we have

$$\begin{aligned} P(F) &= m(\mathrm{HH}) + m(\mathrm{HT}) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \ . \end{aligned}$$

□

Example 1.8 (Example 1.6 continued) The sample space for the experiment in which the die is rolled is the 6-element set $\Omega = \{1, 2, 3, 4, 5, 6\}$. We assumed that the die was fair, and we chose the distribution function defined by

$$m(i) = \frac{1}{6}, \qquad \text{for } i = 1, \ldots, 6 \ .$$

If E is the event that the result of the roll is an even number, then $E = \{2, 4, 6\}$ and

$$\begin{aligned} P(E) &= m(2) + m(4) + m(6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \ . \end{aligned}$$

□

Notice that it is an immediate consequence of the above definitions that, for every $\omega \in \Omega$,

$$P(\{\omega\}) = m(\omega) \ .$$

That is, the probability of the elementary event $\{\omega\}$, consisting of a single outcome ω, is equal to the value $m(\omega)$ assigned to the outcome ω by the distribution function.

Example 1.9 Three people, A, B, and C, are running for the same office, and we assume that one and only one of them wins. The sample space may be taken as the 3-element set Ω ={A,B,C} where each element corresponds to the outcome of that candidate's winning. Suppose that A and B have the same chance of winning, but that C has only 1/2 the chance of A or B. Then we assign

$$m(\mathrm{A}) = m(\mathrm{B}) = 2m(\mathrm{C}) \ .$$

Since

$$m(\mathrm{A}) + m(\mathrm{B}) + m(\mathrm{C}) = 1 \ ,$$

we see that

$$2m(\mathrm{C}) + 2m(\mathrm{C}) + m(\mathrm{C}) = 1 \ ,$$

which implies that $5m(\mathrm{C}) = 1$. Hence,

$$m(\mathrm{A}) = \frac{2}{5}\ , \qquad m(\mathrm{B}) = \frac{2}{5}\ , \qquad m(\mathrm{C}) = \frac{1}{5}\ .$$

Let E be the event that either A or C wins. Then $E = \{\mathrm{A,C}\}$, and

$$P(E) = m(\mathrm{A}) + m(\mathrm{C}) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5}\ .$$

□

In many cases, events can be described in terms of other events through the use of the standard constructions of set theory. We will briefly review the definitions of these constructions. The reader is referred to Figure 1.7 for Venn diagrams which illustrate these constructions.

Let A and B be two sets. Then the union of A and B is the set

$$A \cup B = \{x \,|\, x \in A \text{ or } x \in B\}\ .$$

The intersection of A and B is the set

$$A \cap B = \{x \,|\, x \in A \text{ and } x \in B\}\ .$$

The difference of A and B is the set

$$A - B = \{x \,|\, x \in A \text{ and } x \notin B\}\ .$$

The set A is a subset of B, written $A \subset B$, if every element of A is also an element of B. Finally, the complement of A is the set

$$\tilde{A} = \{x \,|\, x \in \Omega \text{ and } x \notin A\}\ .$$

The reason that these constructions are important is that it is typically the case that complicated events described in English can be broken down into simpler events using these constructions. For example, if A is the event that "it will snow tomorrow and it will rain the next day," B is the event that "it will snow tomorrow," and C is the event that "it will rain two days from now," then A is the intersection of the events B and C. Similarly, if D is the event that "it will snow tomorrow or it will rain the next day," then $D = B \cup C$. (Note that care must be taken here, because sometimes the word "or" in English means that exactly one of the two alternatives will occur. The meaning is usually clear from context. In this book, we will always use the word "or" in the inclusive sense, i.e., A or B means that at least one of the two events A, B is true.) The event $\tilde{B}$ is the event that "it will not snow tomorrow." Finally, if E is the event that "it will snow tomorrow but it will not rain the next day," then $E = B - C$.

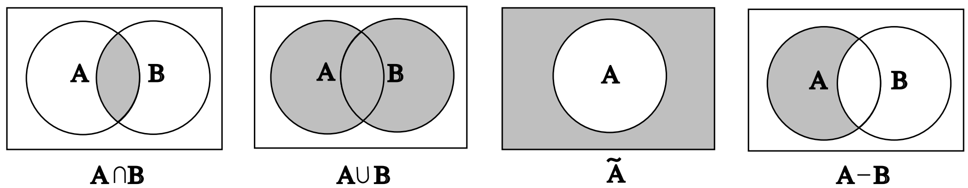


Figure 1.7: Basic set operations.

Properties

Theorem 1.1 The probabilities assigned to events by a distribution function on a sample space Ω satisfy the following properties:

1. $P(E) \geq 0$ for every $E \subset \Omega$.
2. $P(\Omega) = 1$.
3. If $E \subset F \subset \Omega$, then $P(E) \leq P(F)$.
4. If A and B are *disjoint* subsets of Ω, then $P(A \cup B) = P(A) + P(B)$.
5. $P(\tilde{A}) = 1 - P(A)$ for every $A \subset \Omega$.

Proof. For any event E the probability $P(E)$ is determined from the distribution m by

$$P(E) = \sum_{\omega \in E} m(\omega) ,$$

for every $E \subset \Omega$. Since the function m is nonnegative, it follows that $P(E)$ is also nonnegative. Thus, Property 1 is true.

Property 2 is proved by the equations

$$P(\Omega) = \sum_{\omega \in \Omega} m(\omega) = 1 .$$

Suppose that $E \subset F \subset \Omega$. Then every element ω that belongs to E also belongs to F. Therefore,

$$\sum_{\omega \in E} m(\omega) \leq \sum_{\omega \in F} m(\omega) ,$$

since each term in the left-hand sum is in the right-hand sum, and all the terms in both sums are non-negative. This implies that

$$P(E) \leq P(F) ,$$

and Property 3 is proved.

Suppose next that A and B are disjoint subsets of Ω. Then every element ω of $A \cup B$ lies either in A and not in B or in B and not in A. It follows that

$$\begin{aligned} P(A \cup B) &= \textstyle\sum_{\omega \in A \cup B} m(\omega) = \sum_{\omega \in A} m(\omega) + \sum_{\omega \in B} m(\omega) \\ &= P(A) + P(B) \ , \end{aligned}$$

and Property 4 is proved.

Finally, to prove Property 5, consider the disjoint union

$$\Omega = A \cup \tilde{A} \ .$$

Since $P(\Omega) = 1$, the property of disjoint additivity (Property 4) implies that

$$1 = P(A) + P(\tilde{A}) \ ,$$

whence $P(\tilde{A}) = 1 - P(A)$. $\square$

It is important to realize that Property 4 in Theorem 1.1 can be extended to more than two sets. The general finite additivity property is given by the following theorem.

Theorem 1.2 If $A_1, \ldots, A_n$ are pairwise disjoint subsets of Ω (i.e., no two of the A_i's have an element in common), then

$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) \ .$$

Proof. Let ω be any element in the union

$$A_1 \cup \cdots \cup A_n \ .$$

Then $m(\omega)$ occurs exactly once on each side of the equality in the statement of the theorem. $\square$

We shall often use the following consequence of the above theorem.

Theorem 1.3 Let $A_1, \ldots, A_n$ be pairwise disjoint events with $\Omega = A_1 \cup \cdots \cup A_n$, and let E be any event. Then

$$P(E) = \sum_{i=1}^{n} P(E \cap A_i) \ .$$

Proof. The sets $E \cap A_1, \ldots, E \cap A_n$ are pairwise disjoint, and their union is the set E. The result now follows from Theorem 1.2. $\square$

Corollary 1.1 For any two events A and B,

$$P(A) = P(A \cap B) + P(A \cap \tilde{B}) \ .$$

□

Property 4 can be generalized in another way. Suppose that A and B are subsets of Ω which are not necessarily disjoint. Then:

Theorem 1.4 If A and B are subsets of Ω, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \ . \tag{1.1}$$

Proof. The left side of Equation 1.1 is the sum of $m(\omega)$ for ω in either A or B. We must show that the right side of Equation 1.1 also adds $m(\omega)$ for ω in A or B. If ω is in exactly one of the two sets, then it is counted in only one of the three terms on the right side of Equation 1.1. If it is in both A and B, it is added twice from the calculations of $P(A)$ and $P(B)$ and subtracted once for $P(A \cap B)$. Thus it is counted exactly once by the right side. Of course, if $A \cap B = \emptyset$, then Equation 1.1 reduces to Property 4. (Equation 1.1 can also be generalized; see Theorem 3.8.) □

Tree Diagrams

Example 1.10 Let us illustrate the properties of probabilities of events in terms of three tosses of a coin. When we have an experiment which takes place in stages such as this, we often find it convenient to represent the outcomes by a *tree diagram* as shown in Figure 1.8.

A *path* through the tree corresponds to a possible outcome of the experiment. For the case of three tosses of a coin, we have eight paths ω_1, ω_2, ..., ω_8 and, assuming each outcome to be equally likely, we assign equal weight, 1/8, to each path. Let E be the event "at least one head turns up." Then $\tilde{E}$ is the event "no heads turn up." This event occurs for only one outcome, namely, $\omega_8 = \mathrm{TTT}$. Thus, $\tilde{E} = \{\mathrm{TTT}\}$ and we have

$$P(\tilde{E}) = P(\{\mathrm{TTT}\}) = m(\mathrm{TTT}) = \frac{1}{8} \ .$$

By Property 5 of Theorem 1.1,

$$P(E) = 1 - P(\tilde{E}) = 1 - \frac{1}{8} = \frac{7}{8} \ .$$

Note that we shall often find it is easier to compute the probability that an event does not happen rather than the probability that it does. We then use Property 5 to obtain the desired probability.

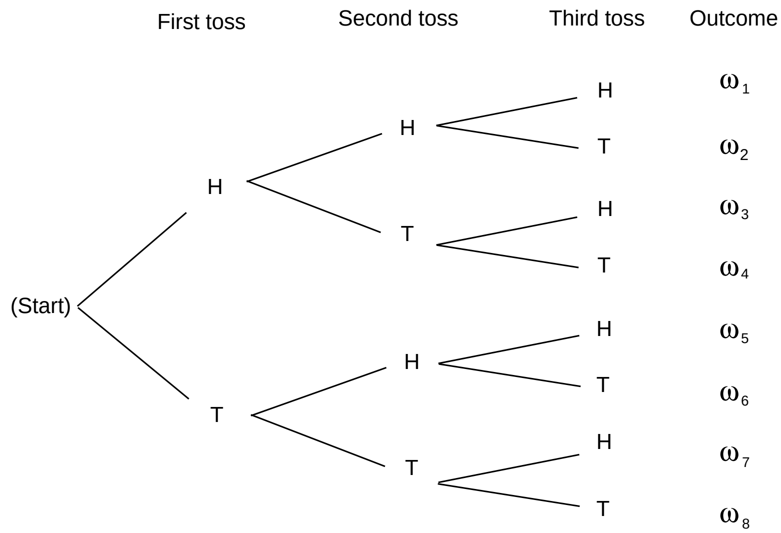


Figure 1.8: Tree diagram for three tosses of a coin.

Let A be the event "the first outcome is a head," and B the event "the second outcome is a tail." By looking at the paths in Figure 1.8, we see that

$$P(A) = P(B) = \frac{1}{2} \ .$$

Moreover, $A \cap B = \{\omega_3, \omega_4\}$, and so $P(A \cap B) = 1/4$. Using Theorem 1.4, we obtain

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4} \ . \end{aligned}$$

Since $A \cup B$ is the 6-element set,

$$A \cup B = \{\text{HHH,HHT,HTH,HTT,TTH,TTT}\} \ ,$$

we see that we obtain the same result by direct enumeration. □

In our coin tossing examples and in the die rolling example, we have assigned an equal probability to each possible outcome of the experiment. Corresponding to this method of assigning probabilities, we have the following definitions.

Uniform Distribution

Definition 1.3 The *uniform distribution* on a sample space Ω containing n elements is the function m defined by

$$m(\omega) = \frac{1}{n} \ ,$$

for every $\omega \in \Omega$. □

It is important to realize that when an experiment is analyzed to describe its possible outcomes, there is no single correct choice of sample space. For the experiment of tossing a coin twice in Example 1.2, we selected the 4-element set Ω ={HH,HT,TH,TT} as a sample space and assigned the uniform distribution function. These choices are certainly intuitively natural. On the other hand, for some purposes it may be more useful to consider the 3-element sample space $\bar{\Omega} = \{0, 1, 2\}$ in which 0 is the outcome "no heads turn up," 1 is the outcome "exactly one head turns up," and 2 is the outcome "two heads turn up." The distribution function $\bar{m}$ on $\bar{\Omega}$ defined by the equations

$$\bar{m}(0) = \frac{1}{4} , \qquad \bar{m}(1) = \frac{1}{2} , \qquad \bar{m}(2) = \frac{1}{4}$$

is the one corresponding to the uniform probability density on the original sample space Ω. Notice that it is perfectly possible to choose a different distribution function. For example, we may consider the uniform distribution function on $\bar{\Omega}$, which is the function $\bar{q}$ defined by

$$\bar{q}(0) = \bar{q}(1) = \bar{q}(2) = \frac{1}{3} .$$

Although $\bar{q}$ is a perfectly good distribution function, it is not consistent with observed data on coin tossing.

Example 1.11 Consider the experiment that consists of rolling a pair of dice. We take as the sample space Ω the set of all ordered pairs (i, j) of integers with $1 \leq i \leq 6$ and $1 \leq j \leq 6$. Thus,

$$\Omega = \{ (i, j) : 1 \leq i, j \leq 6 \} .$$

(There is at least one other "reasonable" choice for a sample space, namely the set of all unordered pairs of integers, each between 1 and 6. For a discussion of why we do not use this set, see Example 3.14.) To determine the size of Ω, we note that there are six choices for i, and for each choice of i there are six choices for j, leading to 36 different outcomes. Let us assume that the dice are not loaded. In mathematical terms, this means that we assume that each of the 36 outcomes is equally likely, or equivalently, that we adopt the uniform distribution function on Ω by setting

$$m((i, j)) = \frac{1}{36}, \qquad 1 \leq i, j \leq 6 .$$

What is the probability of getting a sum of 7 on the roll of two dice—or getting a sum of 11? The first event, denoted by E, is the subset

$$E = \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\} .$$

A sum of 11 is the subset F given by

$$F = \{(5, 6), (6, 5)\} .$$

Consequently,

$$\begin{aligned} P(E) = &\ \textstyle\sum_{\omega \in E} m(\omega) = 6 \cdot \frac{1}{36} = \frac{1}{6} , \\ P(F) = &\ \textstyle\sum_{\omega \in F} m(\omega) = 2 \cdot \frac{1}{36} = \frac{1}{18} . \end{aligned}$$

What is the probability of getting neither *snakeeyes* (double ones) nor *boxcars* (double sixes)? The event of getting either one of these two outcomes is the set

$$E = \{(1,1),(6,6)\} .$$

Hence, the probability of obtaining neither is given by

$$P(\tilde{E}) = 1 - P(E) = 1 - \frac{2}{36} = \frac{17}{18} .$$

□

In the above coin tossing and the dice rolling experiments, we have assigned an equal probability to each outcome. That is, in each example, we have chosen the uniform distribution function. These are the natural choices provided the coin is a fair one and the dice are not loaded. However, the decision as to which distribution function to select to describe an experiment is *not* a part of the basic mathematical theory of probability. The latter begins only when the sample space and the distribution function have already been defined.

Determination of Probabilities

It is important to consider ways in which probability distributions are determined in practice. One way is by *symmetry.* For the case of the toss of a coin, we do not see any physical difference between the two sides of a coin that should affect the chance of one side or the other turning up. Similarly, with an ordinary die there is no essential difference between any two sides of the die, and so by symmetry we assign the same probability for any possible outcome. In general, considerations of symmetry often suggest the uniform distribution function. Care must be used here. We should not always assume that, just because we do not know any reason to suggest that one outcome is more likely than another, it is appropriate to assign equal probabilities. For example, consider the experiment of guessing the sex of a newborn child. It has been observed that the proportion of newborn children who are boys is about .513. Thus, it is more appropriate to assign a distribution function which assigns probability .513 to the outcome *boy* and probability .487 to the outcome *girl* than to assign probability 1/2 to each outcome. This is an example where we use statistical observations to determine probabilities. Note that these probabilities may change with new studies and may vary from country to country. Genetic engineering might even allow an individual to influence this probability for a particular case.

Odds

Statistical estimates for probabilities are fine if the experiment under consideration can be repeated a number of times under similar circumstances. However, assume that, at the beginning of a football season, you want to assign a probability to the event that Dartmouth will beat Harvard. You really do not have data that relates to this year's football team. However, you can determine your own personal probability by seeing what kind of a bet you would be willing to make. For example, suppose that you are willing to make a 1 dollar bet giving 2 to 1 odds that Dartmouth will win. Then you are willing to pay 2 dollars if Dartmouth loses in return for receiving 1 dollar if Dartmouth wins. This means that you think the appropriate probability for Dartmouth winning is 2/3.

Let us look more carefully at the relation between odds and probabilities. Suppose that we make a bet at r to 1 odds that an event E occurs. This means that we think that it is r times as likely that E will occur as that E will not occur. In general, r to s odds will be taken to mean the same thing as r/s to 1, i.e., the ratio between the two numbers is the only quantity of importance when stating odds.

Now if it is r times as likely that E will occur as that E will not occur, then the probability that E occurs must be $r/(r+1)$, since we have

$$P(E) = r\,P(\tilde{E})$$

and

$$P(E) + P(\tilde{E}) = 1\ .$$

In general, the statement that the odds are r to s in favor of an event E occurring is equivalent to the statement that

$$\begin{aligned} P(E) &= \frac{r/s}{(r/s)+1} \\ &= \frac{r}{r+s}\ . \end{aligned}$$

If we let $P(E) = p$, then the above equation can easily be solved for r/s in terms of p; we obtain $r/s = p/(1-p)$. We summarize the above discussion in the following definition.

Definition 1.4 If $P(E) = p$, the *odds* in favor of the event E occurring are $r : s$ (r to s) where $r/s = p/(1-p)$. If r and s are given, then p can be found by using the equation $p = r/(r+s)$. □

Example 1.12 (Example 1.9 continued) In Example 1.9 we assigned probability 1/5 to the event that candidate C wins the race. Thus the odds in favor of C winning are $1/5 : 4/5$. These odds could equally well have been written as $1 : 4$, $2 : 8$, and so forth. A bet that C wins is fair if we receive 4 dollars if C wins and pay 1 dollar if C loses. □

Infinite Sample Spaces

If a sample space has an infinite number of points, then the way that a distribution function is defined depends upon whether or not the sample space is countable. A sample space is *countably infinite* if the elements can be counted, i.e., can be put in one-to-one correspondence with the positive integers, and *uncountably infinite* otherwise. Infinite sample spaces require new concepts in general (see Chapter 2), but countably infinite spaces do not. If

$$\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$$

is a countably infinite sample space, then a distribution function is defined exactly as in Definition 1.2, except that the sum must now be a *convergent* infinite sum. Theorem 1.1 is still true, as are its extensions Theorems 1.2 and 1.4. One thing we cannot do on a countably infinite sample space that we could do on a finite sample space is to define a *uniform* distribution function as in Definition 1.3. You are asked in Exercise 20 to explain why this is not possible.

Example 1.13 A coin is tossed until the first time that a head turns up. Let the outcome of the experiment, ω, be the first time that a head turns up. Then the possible outcomes of our experiment are

$$\Omega = \{1, 2, 3, \ldots\} \ .$$

Note that even though the coin could come up tails every time we have not allowed for this possibility. We will explain why in a moment. The probability that heads comes up on the first toss is 1/2. The probability that tails comes up on the first toss and heads on the second is 1/4. The probability that we have two tails followed by a head is 1/8, and so forth. This suggests assigning the distribution function $m(n) = 1/2^n$ for $n = 1, 2, 3, \ldots$. To see that this is a distribution function we must show that

$$\sum_\omega m(\omega) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1 \ .$$

That this is true follows from the formula for the sum of a geometric series,

$$1 + r + r^2 + r^3 + \cdots = \frac{1}{1-r} \ ,$$

or

$$r + r^2 + r^3 + r^4 + \cdots = \frac{r}{1-r} \ , \tag{1.2}$$

for $-1 < r < 1$.

Putting $r = 1/2$, we see that we have a probability of 1 that the coin eventually turns up heads. The possible outcome of tails every time has to be assigned probability 0, so we omit it from our sample space of possible outcomes.

Let E be the event that the first time a head turns up is after an even number of tosses. Then

$$E = \{2, 4, 6, 8, \ldots\} \ ,$$

and

$$P(E) = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \cdots \ .$$

Putting $r = 1/4$ in Equation 1.2 see that

$$P(E) = \frac{1/4}{1 - 1/4} = \frac{1}{3} \ .$$

Thus the probability that a head turns up for the first time after an even number of tosses is 1/3 and after an odd number of tosses is 2/3. □

Historical Remarks

An interesting question in the history of science is: Why was probability not developed until the sixteenth century? We know that in the sixteenth century problems in gambling and games of chance made people start to think about probability. But gambling and games of chance are almost as old as civilization itself. In ancient Egypt (at the time of the First Dynasty, ca. 3500 B.C.) a game now called "Hounds and Jackals" was played. In this game the movement of the hounds and jackals was based on the outcome of the roll of four-sided dice made out of animal bones called astragali. Six-sided dice made of a variety of materials date back to the sixteenth century B.C. Gambling was widespread in ancient Greece and Rome. Indeed, in the Roman Empire it was sometimes found necessary to invoke laws against gambling. Why, then, were probabilities not calculated until the sixteenth century?

Several explanations have been advanced for this late development. One is that the relevant mathematics was not developed and was not easy to develop. The ancient mathematical notation made numerical calculation complicated, and our familiar algebraic notation was not developed until the sixteenth century. However, as we shall see, many of the combinatorial ideas needed to calculate probabilities were discussed long before the sixteenth century. Since many of the chance events of those times had to do with lotteries relating to religious affairs, it has been suggested that there may have been religious barriers to the study of chance and gambling. Another suggestion is that a stronger incentive, such as the development of commerce, was necessary. However, none of these explanations seems completely satisfactory, and people still wonder why it took so long for probability to be studied seriously. An interesting discussion of this problem can be found in Hacking.[14]

The first person to calculate probabilities systematically was Gerolamo Cardano (1501–1576) in his book *Liber de Ludo Aleae.* This was translated from the Latin by Gould and appears in the book *Cardano: The Gambling Scholar* by Ore.[15] Ore provides a fascinating discussion of the life of this colorful scholar with accounts of his interests in many different fields, including medicine, astrology, and mathematics. You will also find there a detailed account of Cardano's famous battle with Tartaglia over the solution to the cubic equation.

In his book on probability Cardano dealt only with the special case that we have called the uniform distribution function. This restriction to equiprobable outcomes was to continue for a long time. In this case Cardano realized that the probability that an event occurs is the ratio of the number of favorable outcomes to the total number of outcomes.

Many of Cardano's examples dealt with rolling dice. Here he realized that the outcomes for two rolls should be taken to be the 36 ordered pairs (i, j) rather than the 21 unordered pairs. This is a subtle point that was still causing problems much later for other writers on probability. For example, in the eighteenth century the famous French mathematician d'Alembert, author of several works on probability, claimed that when a coin is tossed twice the number of heads that turn up would

[14] I. Hacking, *The Emergence of Probability* (Cambridge: Cambridge University Press, 1975).
[15] O. Ore, *Cardano: The Gambling Scholar* (Princeton: Princeton University Press, 1953).

be 0, 1, or 2, and hence we should assign equal probabilities for these three possible outcomes.[16] Cardano chose the correct sample space for his dice problems and calculated the correct probabilities for a variety of events.

Cardano's mathematical work is interspersed with a lot of advice to the potential gambler in short paragraphs, entitled, for example: "Who Should Play and When," "Why Gambling Was Condemned by Aristotle," "Do Those Who Teach Also Play Well?" and so forth. In a paragraph entitled "The Fundamental Principle of Gambling," Cardano writes:

> The most fundamental principle of all in gambling is simply equal conditions, e.g., of opponents, of bystanders, of money, of situation, of the dice box, and of the die itself. To the extent to which you depart from that equality, if it is in your opponent's favor, you are a fool, and if in your own, you are unjust.[17]

Cardano did make mistakes, and if he realized it later he did not go back and change his error. For example, for an event that is favorable in three out of four cases, Cardano assigned the correct odds $3 : 1$ that the event will occur. But then he assigned odds by squaring these numbers (i.e., $9 : 1$) for the event to happen twice in a row. Later, by considering the case where the odds are $1 : 1$, he realized that this cannot be correct and was led to the correct result that when f out of n outcomes are favorable, the odds for a favorable outcome twice in a row are $f^2 : n^2 - f^2$. Ore points out that this is equivalent to the realization that if the probability that an event happens in one experiment is p, the probability that it happens twice is p^2. Cardano proceeded to establish that for three successes the formula should be p^3 and for four successes p^4, making it clear that he understood that the probability is p^n for n successes in n independent repetitions of such an experiment. This will follow from the concept of independence that we introduce in Section 4.1.

Cardano's work was a remarkable first attempt at writing down the laws of probability, but it was not the spark that started a systematic study of the subject. This came from a famous series of letters between Pascal and Fermat. This correspondence was initiated by Pascal to consult Fermat about problems he had been given by Chevalier de Méré, a well-known writer, a prominent figure at the court of Louis XIV, and an ardent gambler.

The first problem de Méré posed was a dice problem. The story goes that he had been betting that at least one six would turn up in four rolls of a die and winning too often, so he then bet that a pair of sixes would turn up in 24 rolls of a pair of dice. The probability of a six with one die is $1/6$ and, by the product law for independent experiments, the probability of two sixes when a pair of dice is thrown is $(1/6)(1/6) = 1/36$. Ore[18] claims that a gambling rule of the time suggested that, since four repetitions was favorable for the occurrence of an event with probability $1/6$, for an event six times as unlikely, $6 \cdot 4 = 24$ repetitions would be sufficient for

[16] J. d'Alembert, "Croix ou Pile," in *L'Encyclopédie,* ed. Diderot, vol. 4 (Paris, 1754).

[17] O. Ore, op. cit., p. 189.

[18] O. Ore, "Pascal and the Invention of Probability Theory," *American Mathematics Monthly*, vol. 67 (1960), pp. 409–419.

a favorable bet. Pascal showed, by exact calculation, that 25 rolls are required for a favorable bet for a pair of sixes.

The second problem was a much harder one: it was an old problem and concerned the determination of a fair division of the stakes in a tournament when the series, for some reason, is interrupted before it is completed. This problem is now referred to as the problem of points. The problem had been a standard problem in mathematical texts; it appeared in Fra Luca Paccioli's book *summa de Arithmetica, Geometria, Proportioni et Proportionalità*, printed in Venice in 1494,[19] in the form:

> A team plays ball such that a total of 60 points are required to win the game, and each inning counts 10 points. The stakes are 10 ducats. By some incident they cannot finish the game and one side has 50 points and the other 20. One wants to know what share of the prize money belongs to each side. In this case I have found that opinions differ from one to another but all seem to me insufficient in their arguments, but I shall state the truth and give the correct way.

Reasonable solutions, such as dividing the stakes according to the ratio of games won by each player, had been proposed, but no correct solution had been found at the time of the Pascal-Fermat correspondence. The letters deal mainly with the attempts of Pascal and Fermat to solve this problem. Blaise Pascal (1623–1662) was a child prodigy, having published his treatise on conic sections at age sixteen, and having invented a calculating machine at age eighteen. At the time of the letters, his demonstration of the weight of the atmosphere had already established his position at the forefront of contemporary physicists. Pierre de Fermat (1601–1665) was a learned jurist in Toulouse, who studied mathematics in his spare time. He has been called by some the prince of amateurs and one of the greatest pure mathematicians of all times.

The letters, translated by Maxine Merrington, appear in Florence David's fascinating historical account of probability, *Games, Gods and Gambling*.[20] In a letter dated Wednesday, 29th July, 1654, Pascal writes to Fermat:

> Sir,
>
> Like you, I am equally impatient, and although I am again ill in bed, I cannot help telling you that yesterday evening I received from M. de Carcavi your letter on the problem of points, which I admire more than I can possibly say. I have not the leisure to write at length, but, in a word, you have solved the two problems of points, one with dice and the other with sets of games with perfect justness; I am entirely satisfied with it for I do not doubt that I was in the wrong, seeing the admirable agreement in which I find myself with you now...
>
> Your method is very sound and is the one which first came to my mind in this research; but because the labour of the combination is excessive, I have found a short cut and indeed another method which is much

[19] ibid., p. 414.
[20] F. N. David, *Games, Gods and Gambling* (London: G. Griffin, 1962), p. 230 ff.

Number of games B has won				
3	0	0	0	
2	8	16	32	64
1	20	32	48	64
0	32	44	56	64
	0	1	2	3
	Number of games A has won			

Figure 1.9: Pascal's table.

> quicker and neater, which I would like to tell you here in a few words: for henceforth I would like to open my heart to you, if I may, as I am so overjoyed with our agreement. I see that truth is the same in Toulouse as in Paris.
>
> Here, more or less, is what I do to show the fair value of each game, when two opponents play, for example, in three games and each person has staked 32 pistoles.
>
> Let us say that the first man had won twice and the other once; now they play another game, in which the conditions are that, if the first wins, he takes all the stakes; that is 64 pistoles; if the other wins it, then they have each won two games, and therefore, if they wish to stop playing, they must each take back their own stake, that is, 32 pistoles each.
>
> Then consider, Sir, if the first man wins, he gets 64 pistoles; if he loses he gets 32. Thus if they do not wish to risk this last game but wish to separate without playing it, the first man must say: 'I am certain to get 32 pistoles, even if I lost I still get them; but as for the other 32, perhaps I will get them, perhaps you will get them, the chances are equal. Let us then divide these 32 pistoles in half and give one half to me as well as my 32 which are mine for sure.' He will then have 48 pistoles and the other 16...

Pascal's argument produces the table illustrated in Figure 1.9 for the amount due player A at any quitting point.

Each entry in the table is the average of the numbers just above and to the right of the number. This fact, together with the known values when the tournament is completed, determines all the values in this table. If player A wins the first game, then he needs two games to win and B needs three games to win; and so, if the tounament is called off, A should receive 44 pistoles.

The letter in which Fermat presented his solution has been lost; but fortunately, Pascal describes Fermat's method in a letter dated Monday, 24th August, 1654. From Pascal's letter:[21]

> This is your procedure when there are two players: If two players, playing several games, find themselves in that position when the first man needs *two* games and second needs *three*, then to find the fair division of stakes, you say that one must know in how many games the play will be absolutely decided.
>
> It is easy to calculate that this will be in *four* games, from which you can conclude that it is necessary to see in how many ways four games can be arranged between two players, and one must see how many combinations would make the first man win and how many the second and to share out the stakes in this proportion. I would have found it difficult to understand this if I had not known it myself already; in fact you had explained it with this idea in mind.

Fermat realized that the number of ways that the game might be finished may not be equally likely. For example, if A needs two more games and B needs three to win, two possible ways that the tournament might go for A to win are WLW and LWLW. These two sequences do not have the same chance of occurring. To avoid this difficulty, Fermat extended the play, adding fictitious plays, so that all the ways that the games might go have the same length, namely four. He was shrewd enough to realize that this extension would not change the winner and that he now could simply count the number of sequences favorable to each player since he had made them all equally likely. If we list all possible ways that the extended game of four plays might go, we obtain the following 16 possible outcomes of the play:

$\underline{\text{WWWW}}$	$\underline{\text{WLWW}}$	$\underline{\text{LWWW}}$	$\underline{\text{LLWW}}$
$\underline{\text{WWWL}}$	$\underline{\text{WLWL}}$	$\underline{\text{LWWL}}$	LLWL
$\underline{\text{WWLW}}$	$\underline{\text{WLLW}}$	$\underline{\text{LWLW}}$	LLLW
$\underline{\text{WWLL}}$	WLLL	LWLL	LLLL .

Player A wins in the cases where there are at least two wins (the 11 underlined cases), and B wins in the cases where there are at least three losses (the other 5 cases). Since A wins in 11 of the 16 possible cases Fermat argued that the probability that A wins is 11/16. If the stakes are 64 pistoles, A should receive 44 pistoles in agreement with Pascal's result. Pascal and Fermat developed more systematic methods for counting the number of favorable outcomes for problems like this, and this will be one of our central problems. Such counting methods fall under the subject of *combinatorics*, which is the topic of Chapter 3.

[21] ibid., p. 239ff.

We see that these two mathematicians arrived at two very different ways to solve the problem of points. Pascal's method was to develop an algorithm and use it to calculate the fair division. This method is easy to implement on a computer and easy to generalize. Fermat's method, on the other hand, was to change the problem into an equivalent problem for which he could use counting or combinatorial methods. We will see in Chapter 3 that, in fact, Fermat used what has become known as Pascal's triangle! In our study of probability today we shall find that both the algorithmic approach and the combinatorial approach share equal billing, just as they did 300 years ago when probability got its start.

Exercises

1 Let $\Omega = \{a, b, c\}$ be a sample space. Let $m(a) = 1/2$, $m(b) = 1/3$, and $m(c) = 1/6$. Find the probabilities for all eight subsets of Ω.

2 Give a possible sample space Ω for each of the following experiments:

(a) An election decides between two candidates A and B.

(b) A two-sided coin is tossed.

(c) A student is asked for the month of the year and the day of the week on which her birthday falls.

(d) A student is chosen at random from a class of ten students.

(e) You receive a grade in this course.

3 For which of the cases in Exercise 2 would it be reasonable to assign the uniform distribution function?

4 Describe in words the events specified by the following subsets of

$$\Omega = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}$$

(see Example 1.6).

(a) $E = \{\text{HHH,HHT,HTH,HTT}\}$.

(b) $E = \{\text{HHH,TTT}\}$.

(c) $E = \{\text{HHT,HTH,THH}\}$.

(d) $E = \{\text{HHT,HTH,HTT,THH,THT,TTH,TTT}\}$.

5 What are the probabilities of the events described in Exercise 4?

6 A die is loaded in such a way that the probability of each face turning up is proportional to the number of dots on that face. (For example, a six is three times as probable as a two.) What is the probability of getting an even number in one throw?

7 Let A and B be events such that $P(A \cap B) = 1/4$, $P(\tilde{A}) = 1/3$, and $P(B) = 1/2$. What is $P(A \cup B)$?

8 A student must choose one of the subjects, art, geology, or psychology, as an elective. She is equally likely to choose art or psychology and twice as likely to choose geology. What are the respective probabilities that she chooses art, geology, and psychology?

9 A student must choose exactly two out of three electives: art, French, and mathematics. He chooses art with probability 5/8, French with probability 5/8, and art and French together with probability 1/4. What is the probability that he chooses mathematics? What is the probability that he chooses either art or French?

10 For a bill to come before the president of the United States, it must be passed by both the House of Representatives and the Senate. Assume that, of the bills presented to these two bodies, 60 percent pass the House, 80 percent pass the Senate, and 90 percent pass at least one of the two. Calculate the probability that the next bill presented to the two groups will come before the president.

11 What odds should a person give in favor of the following events?

(a) A card chosen at random from a 52-card deck is an ace.

(b) Two heads will turn up when a coin is tossed twice.

(c) Boxcars (two sixes) will turn up when two dice are rolled.

12 You offer 3 : 1 odds that your friend Smith will be elected mayor of your city. What probability are you assigning to the event that Smith wins?

13 In a horse race, the odds that Romance will win are listed as 2 : 3 and that Downhill will win are 1 : 2. What odds should be given for the event that either Romance or Downhill wins?

14 Let X be a random variable with distribution function $m_X(x)$ defined by

$$m_X(-1) = 1/5, \quad m_X(0) = 1/5, \quad m_X(1) = 2/5, \quad m_X(2) = 1/5\ .$$

(a) Let Y be the random variable defined by the equation $Y = X + 3$. Find the distribution function $m_Y(y)$ of Y.

(b) Let Z be the random variable defined by the equation $Z = X^2$. Find the distribution function $m_Z(z)$ of Z.

***15** John and Mary are taking a mathematics course. The course has only three grades: A, B, and C. The probability that John gets a B is .3. The probability that Mary gets a B is .4. The probability that neither gets an A but at least one gets a B is .1. What is the probability that at least one gets a B but neither gets a C?

16 In a fierce battle, not less than 70 percent of the soldiers lost one eye, not less than 75 percent lost one ear, not less than 80 percent lost one hand, and not less than 85 percent lost one leg. What is the minimal possible percentage of those who simultaneously lost one ear, one eye, one hand, and one leg?[22]

***17** Assume that the probability of a "success" on a single experiment with n outcomes is $1/n$. Let m be the number of experiments necessary to make it a favorable bet that at least one success will occur (see Exercise 1.1.5).

(a) Show that the probability that, in m trials, there are no successes is $(1-1/n)^m$.

(b) (de Moivre) Show that if $m = n \log 2$ then

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^m = \frac{1}{2}\ .$$

Hint:

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = e^{-1}\ .$$

Hence for large n we should choose m to be about $n \log 2$.

(c) Would DeMoivre have been led to the correct answer for de Méré's two bets if he had used his approximation?

18 (a) For events $A_1, \ldots, A_n$, prove that

$$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n)\ .$$

(b) For events A and B, prove that

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

19 If A, B, and C are any three events, show that

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

20 Explain why it is not possible to define a uniform distribution function (see Definition 1.3) on a countably infinite sample space. *Hint*: Assume $m(\omega) = a$ for all ω, where $0 \leq a \leq 1$. Does $m(\omega)$ have all the properties of a distribution function?

21 In Example 1.13 find the probability that the coin turns up heads for the first time on the tenth, eleventh, or twelfth toss.

22 A die is rolled until the first time that a six turns up. We shall see that the probability that this occurs on the nth roll is $(5/6)^{n-1} \cdot (1/6)$. Using this fact, describe the appropriate infinite sample space and distribution function for the experiment of rolling a die until a six turns up for the first time. Verify that for your distribution function $\sum_\omega m(\omega) = 1$.

[22]See Knot X, in Lewis Carroll, *Mathematical Recreations,* vol. 2 (Dover, 1958).

23 Let Ω be the sample space

$$\Omega = \{0, 1, 2, \ldots\} \ ,$$

and define a distribution function by

$$m(j) = (1 - r)^j r \ ,$$

for some fixed r, $0 < r < 1$, and for $j = 0, 1, 2, \ldots$. Show that this is a distribution function for Ω.

24 Our calendar has a 400-year cycle. B. H. Brown noticed that the number of times the thirteenth of the month falls on each of the days of the week in the 4800 months of a cycle is as follows:

Sunday 687

Monday 685

Tuesday 685

Wednesday 687

Thursday 684

Friday 688

Saturday 684

From this he deduced that the thirteenth was more likely to fall on Friday than on any other day. Explain what he meant by this.

25 Tversky and Kahneman[23] asked a group of subjects to carry out the following task. They are told that:

> Linda is 31, single, outspoken, and very bright. She majored in philosophy in college. As a student, she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations.

The subjects are then asked to rank the likelihood of various alternatives, such as:

(1) Linda is active in the feminist movement.
(2) Linda is a bank teller.
(3) Linda is a bank teller and active in the feminist movement.

Tversky and Kahneman found that between 85 and 90 percent of the subjects rated alternative (1) most likely, but alternative (3) more likely than alternative (2). Is it? They call this phenomenon the *conjunction fallacy,* and note that it appears to be unaffected by prior training in probability or statistics. Is this phenomenon a fallacy? If so, why? Can you give a possible explanation for the subjects' choices?

[23] K. McKean, "Decisions, Decisions," pp. 22–31.

26 Two cards are drawn successively from a deck of 52 cards. Find the probability that the second card is higher in rank than the first card. *Hint*: Show that $1 = P(\text{higher}) + P(\text{lower}) + P(\text{same})$ and use the fact that $P(\text{higher}) = P(\text{lower})$.

27 A *life table* is a table that lists for a given number of births the estimated number of people who will live to a given age. In Appendix C we give a life table based upon 100,000 births for ages from 0 to 85, both for women and for men. Show how from this table you can estimate the probability $m(x)$ that a person born in 1981 would live to age x. Write a program to plot $m(x)$ both for men and for women, and comment on the differences that you see in the two cases.

***28** Here is an attempt to get around the fact that we cannot choose a "random integer."

(a) What, intuitively, is the probability that a "randomly chosen" positive integer is a multiple of 3?

(b) Let $P_3(N)$ be the probability that an integer, chosen at random between 1 and N, is a multiple of 3 (since the sample space is finite, this is a legitimate probability). Show that the limit

$$P_3 = \lim_{N\to\infty} P_3(N)$$

exists and equals 1/3. This formalizes the intuition in (a), and gives us a way to assign "probabilities" to certain events that are infinite subsets of the positive integers.

(c) If A is any set of positive integers, let $A(N)$ mean the number of elements of A which are less than or equal to N. Then define the "probability" of A as

$$P(A) = \lim_{N\to\infty} A(N)/N\ ,$$

provided this limit exists. Show that this definition would assign probability 0 to any finite set and probability 1 to the set of all positive integers. Thus, the probability of the set of all integers is not the sum of the probabilities of the individual integers in this set. This means that the definition of probability given here is not a completely satisfactory definition.

(d) Let A be the set of all positive integers with an odd number of digits. Show that $P(A)$ does not exist. This shows that under the above definition of probability, not all sets have probabilities.

29 (from Sholander[24]) In a standard clover-leaf interchange, there are four ramps for making right-hand turns, and inside these four ramps, there are four more ramps for making left-hand turns. Your car approaches the interchange from the south. A mechanism has been installed so that at each point where there exists a choice of directions, the car turns to the right with fixed probability r.

[24] M. Sholander, Problem #1034, *Mathematics Magazine,* vol. 52, no. 3 (May 1979), p. 183.

(a) If $r = 1/2$, what is your chance of emerging from the interchange going west?

(b) Find the value of r that maximizes your chance of a westward departure from the interchange.

30 (from Benkoski[25]) Consider a "pure" cloverleaf interchange in which there are no ramps for right-hand turns, but only the two intersecting straight highways with cloverleaves for left-hand turns. (Thus, to turn right in such an interchange, one must make three left-hand turns.) As in the preceding problem, your car approaches the interchange from the south. What is the value of r that maximizes your chances of an eastward departure from the interchange?

31 (from vos Savant[26]) A reader of Marilyn vos Savant's column wrote in with the following question:

> My dad heard this story on the radio. At Duke University, two students had received A's in chemistry all semester. But on the night before the final exam, they were partying in another state and didn't get back to Duke until it was over. Their excuse to the professor was that they had a flat tire, and they asked if they could take a make-up test. The professor agreed, wrote out a test and sent the two to separate rooms to take it. The first question (on one side of the paper) was worth 5 points, and they answered it easily. Then they flipped the paper over and found the second question, worth 95 points: 'Which tire was it?' What was the probability that both students would say the same thing? My dad and I think it's 1 in 16. Is that right?"

(a) Is the answer 1/16?

(b) The following question was asked of a class of students. "I was driving to school today, and one of my tires went flat. Which tire do you think it was?" The responses were as follows: right front, 58%, left front, 11%, right rear, 18%, left rear, 13%. Suppose that this distribution holds in the general population, and assume that the two test-takers are randomly chosen from the general population. What is the probability that they will give the same answer to the second question?

[25] S. Benkoski, Comment on Problem #1034, *Mathematics Magazine,* vol. 52, no. 3 (May 1979), pp. 183-184.

[26] M. vos Savant, *Parade Magazine*, 3 March 1996, p. 14.

Chapter 2

Continuous Probability Densities

2.1 Simulation of Continuous Probabilities

In this section we shall show how we can use computer simulations for experiments that have a whole continuum of possible outcomes.

Probabilities

Example 2.1 We begin by constructing a spinner, which consists of a circle of *unit circumference* and a pointer as shown in Figure 2.1. We pick a point on the circle and label it 0, and then label every other point on the circle with the distance, say x, from 0 to that point, measured counterclockwise. The experiment consists of spinning the pointer and recording the label of the point at the tip of the pointer. We let the random variable X denote the value of this outcome. The sample space is clearly the interval $[0, 1)$. We would like to construct a probability model in which each outcome is equally likely to occur.

If we proceed as we did in Chapter 1 for experiments with a finite number of possible outcomes, then we must assign the probability 0 to each outcome, since otherwise, the sum of the probabilities, over all of the possible outcomes, would not equal 1. (In fact, summing an uncountable number of real numbers is a tricky business; in particular, in order for such a sum to have any meaning, at most countably many of the summands can be different than 0.) However, if all of the assigned probabilities are 0, then the sum is 0, not 1, as it should be.

In the next section, we will show how to construct a probability model in this situation. At present, we will assume that such a model can be constructed. We will also assume that in this model, if E is an arc of the circle, and E is of length p, then the model will assign the probability p to E. This means that if the pointer is spun, the probability that it ends up pointing to a point in E equals p, which is certainly a reasonable thing to expect.

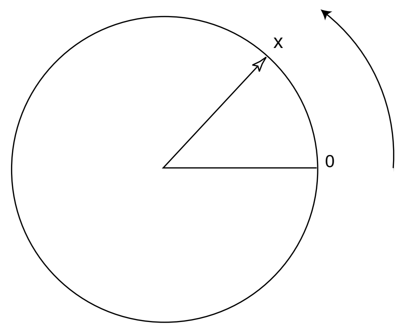


Figure 2.1: A spinner.

To simulate this experiment on a computer is an easy matter. Many computer software packages have a function which returns a random real number in the interval $[0, 1]$. Actually, the returned value is always a rational number, and the values are determined by an algorithm, so a sequence of such values is not truly random. Nevertheless, the sequences produced by such algorithms behave much like theoretically random sequences, so we can use such sequences in the simulation of experiments. On occasion, we will need to refer to such a function. We will call this function *rnd*. $\square$

Monte Carlo Procedure and Areas

It is sometimes desirable to estimate quantities whose exact values are difficult or impossible to calculate exactly. In some of these cases, a procedure involving chance, called a *Monte Carlo procedure*, can be used to provide such an estimate.

Example 2.2 In this example we show how simulation can be used to estimate areas of plane figures. Suppose that we program our computer to provide a pair (x, y) or numbers, each chosen independently at random from the interval $[0, 1]$. Then we can interpret this pair (x, y) as the coordinates of a point chosen *at random* from the unit square. Events are subsets of the unit square. Our experience with Example 2.1 suggests that the point is equally likely to fall in subsets of equal area. Since the total area of the square is 1, the probability of the point falling in a specific subset E of the unit square should be equal to its area. Thus, we can estimate the area of any subset of the unit square by estimating the probability that a point chosen at random from this square falls in the subset.

We can use this method to estimate the area of the region E under the curve $y = x^2$ in the unit square (see Figure 2.2). We choose a large number of points (x, y) at random and record what fraction of them fall in the region $E = \{ (x, y) : y \le x^2 \}$.

The program **MonteCarlo** will carry out this experiment for us. Running this program for 10,000 experiments gives an estimate of .325 (see Figure 2.3).

From these experiments we would estimate the area to be about 1/3. Of course,

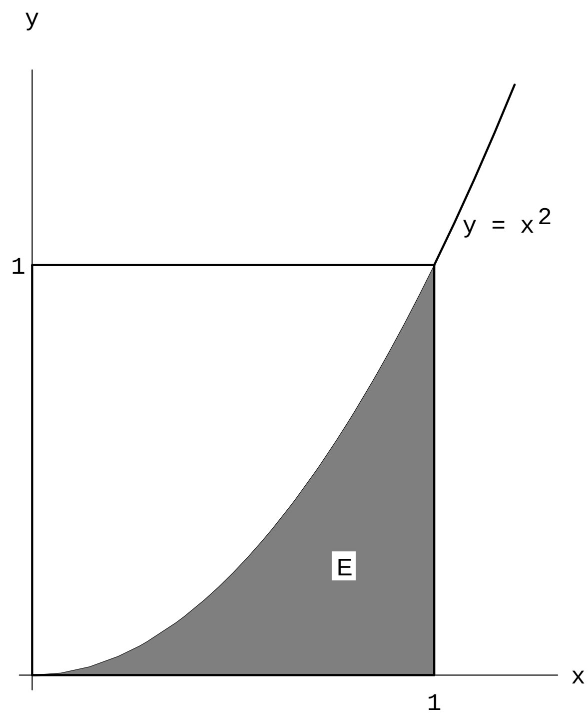


Figure 2.2: Area under $y = x^2$.

for this simple region we can find the exact area by calculus. In fact,

$$\text{Area of } E = \int_0^1 x^2\,dx = \frac{1}{3}\ .$$

We have remarked in Chapter 1 that, when we simulate an experiment of this type n times to estimate a probability, we can expect the answer to be in error by at most $1/\sqrt{n}$ at least 95 percent of the time. For 10,000 experiments we can expect an accuracy of 0.01, and our simulation did achieve this accuracy.

This same argument works for any region E of the unit square. For example, suppose E is the circle with center $(1/2, 1/2)$ and radius $1/2$. Then the probability that our random point (x, y) lies inside the circle is equal to the area of the circle, that is,

$$P(E) = \pi\Bigl(\frac{1}{2}\Bigr)^2 = \frac{\pi}{4}\ .$$

If we did not know the value of π, we could estimate the value by performing this experiment a large number of times! □

The above example is not the only way of estimating the value of π by a chance experiment. Here is another way, discovered by Buffon.[1]

[1] G. L. Buffon, in "Essai d'Arithmétique Morale," *Oeuvres Complètes de Buffon avec Supplements,* tome iv, ed. Duménil (Paris, 1836).

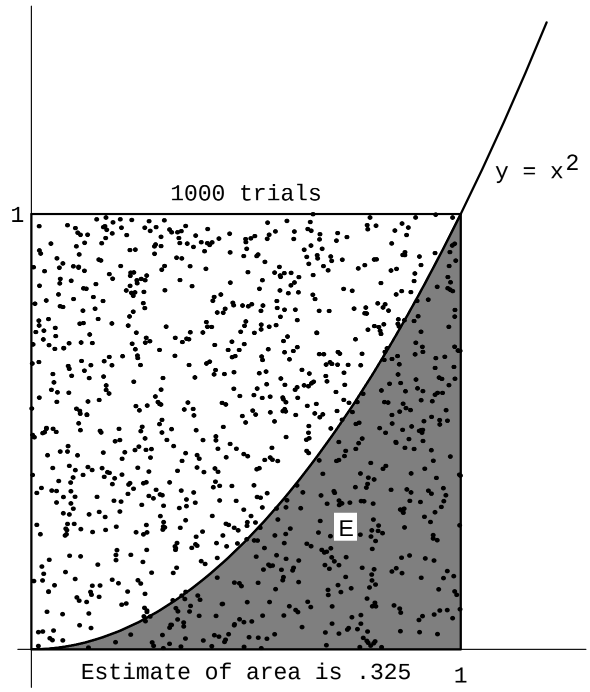


Figure 2.3: Computing the area by simulation.

Buffon's Needle

Example 2.3 Suppose that we take a card table and draw across the top surface a set of parallel lines a unit distance apart. We then drop a common needle of unit length at random on this surface and observe whether or not the needle lies across one of the lines. We can describe the possible outcomes of this experiment by coordinates as follows: Let d be the distance from the center of the needle to the nearest line. Next, let L be the line determined by the needle, and define θ as the acute angle that the line L makes with the set of parallel lines. (The reader should certainly be wary of this description of the sample space. We are attempting to coordinatize a set of line segments. To see why one must be careful in the choice of coordinates, see Example 2.6.) Using this description, we have $0 \leq d \leq 1/2$, and $0 \leq \theta \leq \pi/2$. Moreover, we see that the needle lies across the nearest line if and only if the hypotenuse of the triangle (see Figure 2.4) is less than half the length of the needle, that is,

$$\frac{d}{\sin \theta} < \frac{1}{2} .$$

Now we assume that when the needle drops, the pair (θ, d) is chosen at random from the rectangle $0 \leq \theta \leq \pi/2$, $0 \leq d \leq 1/2$. We observe whether the needle lies across the nearest line (i.e., whether $d \leq (1/2) \sin \theta$). The probability of this event E is the fraction of the area of the rectangle which lies inside E (see Figure 2.5).

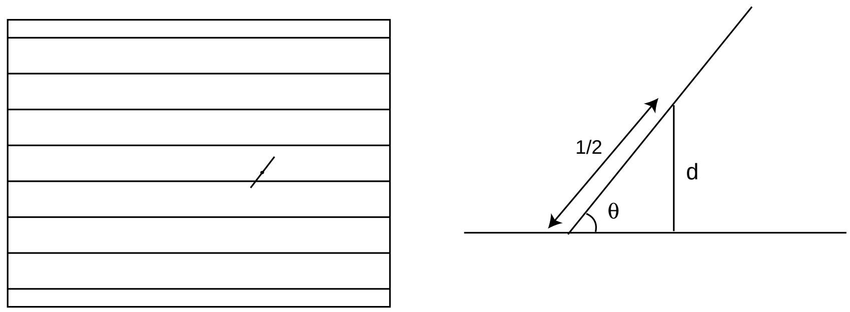


Figure 2.4: Buffon's experiment.

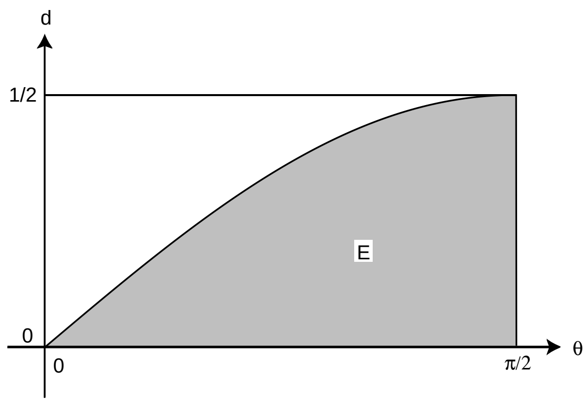


Figure 2.5: Set E of pairs (θ, d) with $d < \frac{1}{2}\sin\theta$.

Now the area of the rectangle is $\pi/4$, while the area of E is

$$\text{Area} = \int_0^{\pi/2} \frac{1}{2}\sin\theta\, d\theta = \frac{1}{2}\ .$$

Hence, we get

$$P(E) = \frac{1/2}{\pi/4} = \frac{2}{\pi}\ .$$

The program **BuffonsNeedle** simulates this experiment. In Figure 2.6, we show the position of every 100th needle in a run of the program in which 10,000 needles were "dropped." Our final estimate for π is 3.139. While this was within 0.003 of the true value for π we had no right to expect such accuracy. The reason for this is that our simulation estimates $P(E)$. While we can expect this estimate to be in error by at most 0.001, a small error in $P(E)$ gets magnified when we use this to compute $\pi = 2/P(E)$. Perlman and Wichura, in their article "Sharpening Buffon's

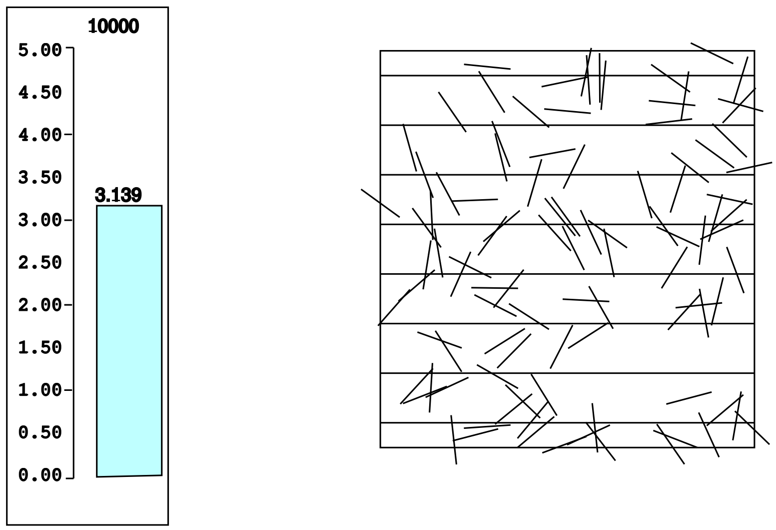


Figure 2.6: Simulation of Buffon's needle experiment.

Needle,"[2] show that we can expect to have an error of not more than $5/\sqrt{n}$ about 95 percent of the time. Here n is the number of needles dropped. Thus for 10,000 needles we should expect an error of no more than 0.05, and that was the case here. We see that a large number of experiments is necessary to get a decent estimate for π. □

In each of our examples so far, events of the same size are equally likely. Here is an example where they are not. We will see many other such examples later.

Example 2.4 Suppose that we choose two random real numbers in $[0, 1]$ and add them together. Let X be the sum. How is X distributed?

To help understand the answer to this question, we can use the program **Areabargraph**. This program produces a bar graph with the property that on each interval, the *area*, rather than the height, of the bar is equal to the fraction of outcomes that fell in the corresponding interval. We have carried out this experiment 1000 times; the data is shown in Figure 2.7. It appears that the function defined by

$$f(x) = \begin{cases} x, & \text{if } 0 \le x \le 1, \\ 2 - x, & \text{if } 1 < x \le 2 \end{cases}$$

fits the data very well. (It is shown in the figure.) In the next section, we will see that this function is the "right" function. By this we mean that if a and b are any two real numbers between 0 and 2, with $a \le b$, then we can use this function to calculate the probability that $a \le X \le b$. To understand how this calculation might be performed, we again consider Figure 2.7. Because of the way the bars were constructed, the sum of the areas of the bars corresponding to the interval

[2]M. D. Perlman and M. J. Wichura, "Sharpening Buffon's Needle," *The American Statistician,* vol. 29, no. 4 (1975), pp. 157–163.

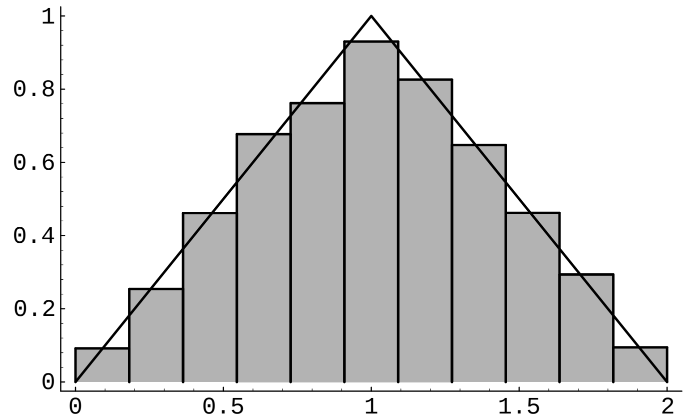


Figure 2.7: Sum of two random numbers.

$[a, b]$ approximates the probability that $a \leq X \leq b$. But the sum of the areas of these bars also approximates the integral

$$\int_a^b f(x)\, dx\ .$$

This suggests that for an experiment with a continuum of possible outcomes, if we find a function with the above property, then we will be able to use it to calculate probabilities. In the next section, we will show how to determine the function $f(x)$. $\square$

Example 2.5 Suppose that we choose 100 random numbers in $[0, 1]$, and let X represent their sum. How is X distributed? We have carried out this experiment 10000 times; the results are shown in Figure 2.8. It is not so clear what function fits the bars in this case. It turns out that the type of function which does the job is called a *normal density* function. This type of function is sometimes referred to as a "bell-shaped" curve. It is among the most important functions in the subject of probability, and will be formally defined in Section 5.2 of Chapter 4.3. $\square$

Our last example explores the fundamental question of how probabilities are assigned.

Bertrand's Paradox

Example 2.6 A chord of a circle is a line segment both of whose endpoints lie on the circle. Suppose that a chord is drawn *at random* in a unit circle. What is the probability that its length exceeds $\sqrt{3}$?

Our answer will depend on what we mean by *random,* which will depend, in turn, on what we choose for coordinates. The sample space Ω is the set of all possible chords in the circle. To find coordinates for these chords, we first introduce a

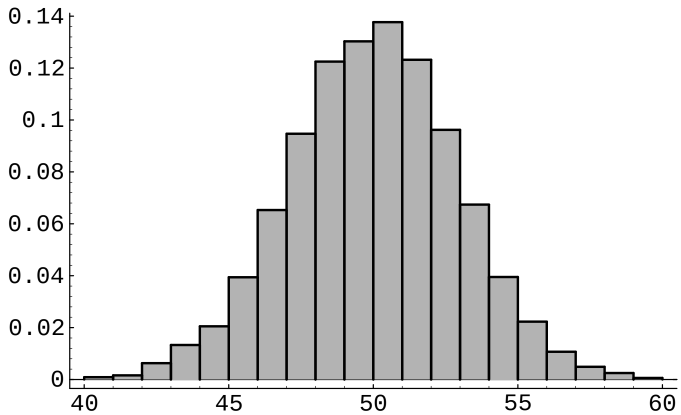


Figure 2.8: Sum of 100 random numbers.

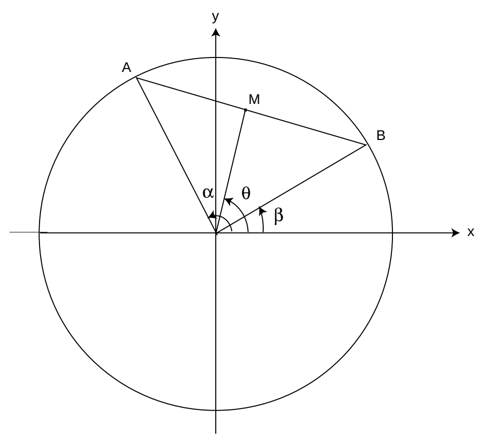


Figure 2.9: Random chord.

rectangular coordinate system with origin at the center of the circle (see Figure 2.9). We note that a chord of a circle is perpendicular to the radial line containing the midpoint of the chord. We can describe each chord by giving:

1. The rectangular coordinates (x, y) of the midpoint M, or

2. The polar coordinates (r, θ) of the midpoint M, or

3. The polar coordinates $(1, \alpha)$ and $(1, \beta)$ of the endpoints A and B.

In each case we shall interpret *at random* to mean: choose these coordinates at random.

We can easily estimate this probability by computer simulation. In programming this simulation, it is convenient to include certain simplifications, which we describe in turn:

1. To simulate this case, we choose values for x and y from $[-1, 1]$ at random. Then we check whether $x^2 + y^2 \leq 1$. If not, the point $M = (x, y)$ lies outside the circle and cannot be the midpoint of any chord, and we ignore it. Otherwise, M lies inside the circle and is the midpoint of a unique chord, whose length L is given by the formula:
$$L = 2\sqrt{1 - (x^2 + y^2)}\ .$$

2. To simulate this case, we take account of the fact that any rotation of the circle does not change the length of the chord, so we might as well assume in advance that the chord is horizontal. Then we choose r from $[-1, 1]$ at random, and compute the length of the resulting chord with midpoint $(r, \pi/2)$ by the formula:
$$L = 2\sqrt{1 - r^2}\ .$$

3. To simulate this case, we assume that one endpoint, say B, lies at $(1, 0)$ (i.e., that $\beta = 0$). Then we choose a value for α from $[0, 2\pi]$ at random and compute the length of the resulting chord, using the Law of Cosines, by the formula:
$$L = \sqrt{2 - 2\cos\alpha}\ .$$

The program **BertrandsParadox** carries out this simulation. Running this program produces the results shown in Figure 2.10. In the first circle in this figure, a smaller circle has been drawn. Those chords which intersect this smaller circle have length at least $\sqrt{3}$. In the second circle in the figure, the vertical line intersects all chords of length at least $\sqrt{3}$. In the third circle, again the vertical line intersects all chords of length at least $\sqrt{3}$.

In each case we run the experiment a large number of times and record the fraction of these lengths that exceed $\sqrt{3}$. We have printed the results of every 100th trial up to 10,000 trials.

It is interesting to observe that these fractions are *not* the same in the three cases; they depend on our choice of coordinates. This phenomenon was first observed by Bertrand, and is now known as *Bertrand's paradox.*[3] It is actually not a paradox at all; it is merely a reflection of the fact that different choices of coordinates will lead to different assignments of probabilities. Which assignment is "correct" depends on what application or interpretation of the model one has in mind.

One can imagine a real experiment involving throwing long straws at a circle drawn on a card table. A "correct" assignment of coordinates should not depend on where the circle lies on the card table, or where the card table sits in the room. Jaynes[4] has shown that the only assignment which meets this requirement is (2). In this sense, the assignment (2) is the natural, or "correct" one (see Exercise 11).

We can easily see in each case what the true probabilities are if we note that $\sqrt{3}$ is the length of the side of an inscribed equilateral triangle. Hence, a chord has

[3] J. Bertrand, *Calcul des Probabilités* (Paris: Gauthier-Villars, 1889).

[4] E. T. Jaynes, "The Well-Posed Problem," in *Papers on Probability, Statistics and Statistical Physics,* R. D. Rosencrantz, ed. (Dordrecht: D. Reidel, 1983), pp. 133–148.

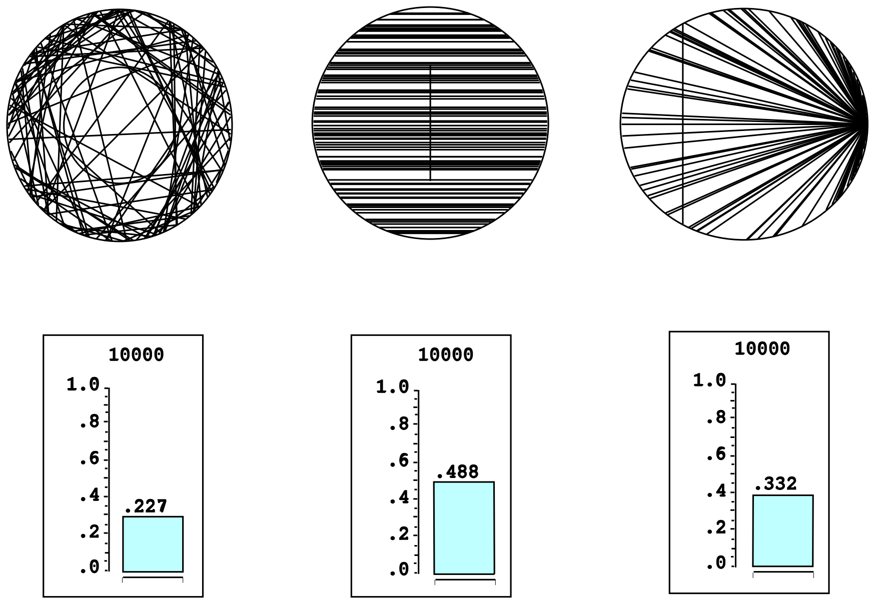


Figure 2.10: Bertrand's paradox.

length $L > \sqrt{3}$ if its midpoint has distance $d < 1/2$ from the origin (see Figure 2.9). The following calculations determine the probability that $L > \sqrt{3}$ in each of the three cases.

1. $L > \sqrt{3}$ if(x, y) lies inside a circle of radius 1/2, which occurs with probability
$$p = \frac{\pi(1/2)^2}{\pi(1)^2} = \frac{1}{4}\ .$$

2. $L > \sqrt{3}$ if $|r| < 1/2$, which occurs with probability
$$\frac{1/2 - (-1/2)}{1 - (-1)} = \frac{1}{2}\ .$$

3. $L > \sqrt{3}$ if $2\pi/3 < \alpha < 4\pi/3$, which occurs with probability
$$\frac{4\pi/3 - 2\pi/3}{2\pi - 0} = \frac{1}{3}\ .$$

We see that our simulations agree quite well with these theoretical values. □

Historical Remarks

G. L. Buffon (1707–1788) was a natural scientist in the eighteenth century who applied probability to a number of his investigations. His work is found in his monumental 44-volume *Histoire Naturelle* and its supplements.[5] For example, he

[5]G. L. Buffon, *Histoire Naturelle, Generali et Particular avec le Descriptión du Cabinet du Roy,* 44 vols. (Paris: L'Imprimerie Royale, 1749–1803).

Experimenter	Length of needle	Number of casts	Number of crossings	Estimate for π
Wolf, 1850	.8	5000	2532	3.1596
Smith, 1855	.6	3204	1218.5	3.1553
De Morgan, c.1860	1.0	600	382.5	3.137
Fox, 1864	.75	1030	489	3.1595
Lazzerini, 1901	.83	3408	1808	3.1415929
Reina, 1925	.5419	2520	869	3.1795

Table 2.1: Buffon needle experiments to estimate π.

presented a number of mortality tables and used them to compute, for each age group, the expected remaining lifetime. From his table he observed: the expected remaining lifetime of an infant of one year is 33 years, while that of a man of 21 years is also approximately 33 years. Thus, a father who is not yet 21 can hope to live longer than his one year old son, but if the father is 40, the odds are already 3 to 2 that his son will outlive him.[6]

Buffon wanted to show that not all probability calculations rely only on algebra, but that some rely on geometrical calculations. One such problem was his famous "needle problem" as discussed in this chapter.[7] In his original formulation, Buffon describes a game in which two gamblers drop a loaf of French bread on a wide-board floor and bet on whether or not the loaf falls across a crack in the floor. Buffon asked: what length L should the bread loaf be, relative to the width W of the floorboards, so that the game is fair. He found the correct answer ($L = (\pi/4)W$) using essentially the methods described in this chapter. He also considered the case of a checkerboard floor, but gave the wrong answer in this case. The correct answer was given later by Laplace.

The literature contains descriptions of a number of experiments that were actually carried out to estimate π by this method of dropping needles. N. T. Gridgeman[8] discusses the experiments shown in Table 2.1. (The halves for the number of crossing comes from a compromise when it could not be decided if a crossing had actually occurred.) He observes, as we have, that 10,000 casts could do no more than establish the first decimal place of π with reasonable confidence. Gridgeman points out that, although none of the experiments used even 10,000 casts, they are surprisingly good, and in some cases, too good. The fact that the number of casts is not always a round number would suggest that the authors might have resorted to clever stopping to get a good answer. Gridgeman comments that Lazzerini's estimate turned out to agree with a well-known approximation to π, $355/113 = 3.1415929$, discovered by the fifth-century Chinese mathematician, Tsu Ch'ungchih. Gridgeman says that he did not have Lazzerini's original report, and while waiting for it (knowing

[6] G. L. Buffon, "Essai d'Arithmétique Morale," p. 301.
[7] ibid., pp. 277–278.
[8] N. T. Gridgeman, "Geometric Probability and the Number π" *Scripta Mathematika,* vol. 25, no. 3, (1960), pp. 183–195.

only the needle crossed a line 1808 times in 3408 casts) deduced that the length of the needle must have been 5/6. He calculated this from Buffon's formula, assuming $\pi = 355/113$:

$$L = \frac{\pi P(E)}{2} = \frac{1}{2}\left(\frac{355}{113}\right)\left(\frac{1808}{3408}\right) = \frac{5}{6} = .8333\ .$$

Even with careful planning one would have to be extremely lucky to be able to stop so cleverly.

The second author likes to trace his interest in probability theory to the Chicago World's Fair of 1933 where he observed a mechanical device dropping needles and displaying the ever-changing estimates for the value of π. (The first author likes to trace his interest in probability theory to the second author.)

Exercises

***1** In the spinner problem (see Example 2.1) divide the unit circumference into three arcs of length 1/2, 1/3, and 1/6. Write a program to simulate the spinner experiment 1000 times and print out what fraction of the outcomes fall in each of the three arcs. Now plot a bar graph whose bars have width 1/2, 1/3, and 1/6, and areas equal to the corresponding fractions as determined by your simulation. Show that the heights of the bars are all nearly the same.

2 Do the same as in Exercise 1, but divide the unit circumference into five arcs of length 1/3, 1/4, 1/5, 1/6, and 1/20.

3 Alter the program **MonteCarlo** to estimate the area of the circle of radius 1/2 with center at $(1/2, 1/2)$ inside the unit square by choosing 1000 points at random. Compare your results with the true value of $\pi/4$. Use your results to estimate the value of π. How accurate is your estimate?

4 Alter the program **MonteCarlo** to estimate the area under the graph of $y = \sin \pi x$ inside the unit square by choosing 10,000 points at random. Now calculate the true value of this area and use your results to estimate the value of π. How accurate is your estimate?

5 Alter the program **MonteCarlo** to estimate the area under the graph of $y = 1/(x + 1)$ in the unit square in the same way as in Exercise 4. Calculate the true value of this area and use your simulation results to estimate the value of $\log 2$. How accurate is your estimate?

6 To simulate the Buffon's needle problem we choose independently the distance d and the angle θ at random, with $0 \le d \le 1/2$ and $0 \le \theta \le \pi/2$, and check whether $d \le (1/2)\sin\theta$. Doing this a large number of times, we estimate π as $2/a$, where a is the fraction of the times that $d \le (1/2)\sin\theta$. Write a program to estimate π by this method. Run your program several times for each of 100, 1000, and 10,000 experiments. Does the accuracy of the experimental approximation for π improve as the number of experiments increases?

7 For Buffon's needle problem, Laplace[9] considered a grid with *horizontal* and *vertical* lines one unit apart. He showed that the probability that a needle of length $L \leq 1$ crosses at least one line is

$$p = \frac{4L - L^2}{\pi} \ .$$

To simulate this experiment we choose at random an angle θ between 0 and $\pi/2$ and independently two numbers d_1 and d_2 between 0 and $L/2$. (The two numbers represent the distance from the center of the needle to the nearest horizontal and vertical line.) The needle crosses a line if either $d_1 \leq (L/2)\sin\theta$ or $d_2 \leq (L/2)\cos\theta$. We do this a large number of times and estimate π as

$$\bar{\pi} = \frac{4L - L^2}{a} \ ,$$

where a is the proportion of times that the needle crosses at least one line. Write a program to estimate π by this method, run your program for 100, 1000, and 10,000 experiments, and compare your results with Buffon's method described in Exercise 6. (Take $L = 1$.)

8 A long needle of length L much bigger than 1 is dropped on a grid with horizontal and vertical lines one unit apart. We will see (in Exercise 6.3.28) that the average number a of lines crossed is approximately

$$a = \frac{4L}{\pi} \ .$$

To estimate π by simulation, pick an angle θ at random between 0 and $\pi/2$ and compute $L\sin\theta + L\cos\theta$. This may be used for the number of lines crossed. Repeat this many times and estimate π by

$$\bar{\pi} = \frac{4L}{a} \ ,$$

where a is the average number of lines crossed per experiment. Write a program to simulate this experiment and run your program for the number of experiments equal to 100, 1000, and 10,000. Compare your results with the methods of Laplace or Buffon for the same number of experiments. (Use $L = 100$.)

The following exercises involve experiments in which not all outcomes are equally likely. We shall consider such experiments in detail in the next section, but we invite you to explore a few simple cases here.

9 A large number of waiting time problems have an *exponential distribution* of outcomes. We shall see (in Section 5.2) that such outcomes are simulated by computing $(-1/\lambda)\log(\text{rnd})$, where $\lambda > 0$. For waiting times produced in this way, the average waiting time is $1/\lambda$. For example, the times spent waiting for

[9] P. S. Laplace, *Théorie Analytique des Probabilités* (Paris: Courcier, 1812).

a car to pass on a highway, or the times between emissions of particles from a radioactive source, are simulated by a sequence of random numbers, each of which is chosen by computing $(-1/\lambda)\log(\text{rnd})$, where $1/\lambda$ is the average time between cars or emissions. Write a program to simulate the times between cars when the average time between cars is 30 seconds. Have your program compute an area bar graph for these times by breaking the time interval from 0 to 120 into 24 subintervals. On the same pair of axes, plot the function $f(x) = (1/30)e^{-(1/30)x}$. Does the function fit the bar graph well?

10 In Exercise 9, the distribution came "out of a hat." In this problem, we will again consider an experiment whose outcomes are not equally likely. We will determine a function $f(x)$ which can be used to determine the probability of certain events. Let T be the right triangle in the plane with vertices at the points $(0,0)$, $(1,0)$, and $(0,1)$. The experiment consists of picking a point at random in the interior of T, and recording only the x-coordinate of the point. Thus, the sample space is the set $[0,1]$, but the outcomes do not seem to be equally likely. We can simulate this experiment by asking a computer to return two random real numbers in $[0,1]$, and recording the first of these two numbers if their sum is less than 1. Write this program and run it for 10,000 trials. Then make a bar graph of the result, breaking the interval $[0,1]$ into 10 intervals. Compare the bar graph with the function $f(x) = 2 - 2x$. Now show that there is a constant c such that the height of T at the x-coordinate value x is c times $f(x)$ for every x in $[0,1]$. Finally, show that

$$\int_0^1 f(x)\,dx = 1 .$$

How might one use the function $f(x)$ to determine the probability that the outcome is between .2 and .5?

11 Here is another way to pick a chord *at random* on the circle of unit radius. Imagine that we have a card table whose sides are of length 100. We place coordinate axes on the table in such a way that each side of the table is parallel to one of the axes, and so that the center of the table is the origin. We now place a circle of unit radius on the table so that the center of the circle is the origin. Now pick out a point (x_0, y_0) at random in the square, and an angle θ at random in the interval $(-\pi/2, \pi/2)$. Let $m = \tan\theta$. Then the equation of the line passing through (x_0, y_0) with slope m is

$$y = y_0 + m(x - x_0) ,$$

and the distance of this line from the center of the circle (i.e., the origin) is

$$d = \left|\frac{y_0 - mx_0}{\sqrt{m^2+1}}\right| .$$

We can use this distance formula to check whether the line intersects the circle (i.e., whether $d < 1$). If so, we consider the resulting chord a *random* chord.

This describes an experiment of dropping a long straw at random on a table on which a circle is drawn.

Write a program to simulate this experiment 10000 times and estimate the probability that the length of the chord is greater than $\sqrt{3}$. How does your estimate compare with the results of Example 2.6?

2.2 Continuous Density Functions

In the previous section we have seen how to simulate experiments with a whole continuum of possible outcomes and have gained some experience in thinking about such experiments. Now we turn to the general problem of assigning probabilities to the outcomes and events in such experiments. We shall restrict our attention here to those experiments whose sample space can be taken as a suitably chosen subset of the line, the plane, or some other Euclidean space. We begin with some simple examples.

Spinners

Example 2.7 The spinner experiment described in Example 2.1 has the interval $[0, 1)$ as the set of possible outcomes. We would like to construct a probability model in which each outcome is equally likely to occur. We saw that in such a model, it is necessary to assign the probability 0 to each outcome. This does not at all mean that the probability of *every* event must be zero. On the contrary, if we let the random variable X denote the outcome, then the probability

$$P(\,0 \le X \le 1)$$

that the head of the spinner comes to rest *somewhere* in the circle, should be equal to 1. Also, the probability that it comes to rest in the upper half of the circle should be the same as for the lower half, so that

$$P\biggl(0 \le X < \frac{1}{2}\biggr) = P\biggl(\frac{1}{2} \le X < 1\biggr) = \frac{1}{2}\ .$$

More generally, in our model, we would like the equation

$$P(c \le X < d) = d - c$$

to be true for every choice of c and d.

If we let $E = [c, d]$, then we can write the above formula in the form

$$P(E) = \int_E f(x)\, dx\ ,$$

where $f(x)$ is the constant function with value 1. This should remind the reader of the corresponding formula in the discrete case for the probability of an event:

$$P(E) = \sum_{\omega \in E} m(\omega)\ .$$

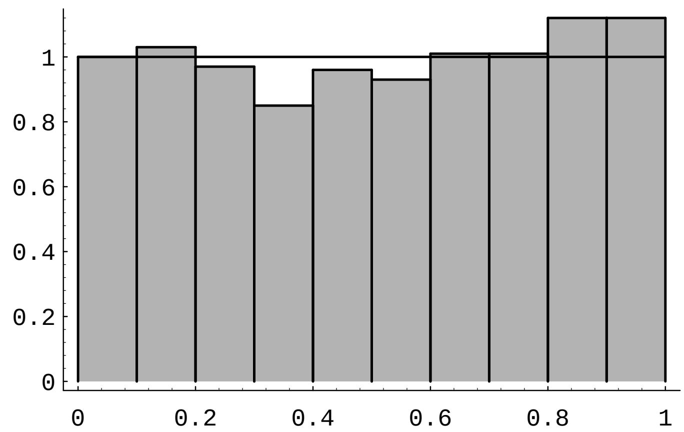


Figure 2.11: Spinner experiment.

The difference is that in the continuous case, the quantity being integrated, $f(x)$, is not the probability of the outcome x. (However, if one uses infinitesimals, one can consider $f(x)\,dx$ as the probability of the outcome x.)

In the continuous case, we will use the following convention. If the set of outcomes is a set of real numbers, then the individual outcomes will be referred to by small Roman letters such as x. If the set of outcomes is a subset of R^2, then the individual outcomes will be denoted by (x, y). In either case, it may be more convenient to refer to an individual outcome by using ω, as in Chapter 1.

Figure 2.11 shows the results of 1000 spins of the spinner. The function $f(x)$ is also shown in the figure. The reader will note that the area under $f(x)$ and above a given interval is approximately equal to the fraction of outcomes that fell in that interval. The function $f(x)$ is called the *density function* of the random variable X. The fact that the area under $f(x)$ and above an interval corresponds to a probability is the defining property of density functions. A precise definition of density functions will be given shortly. $\square$

Darts

Example 2.8 A game of darts involves throwing a dart at a circular target of *unit radius.* Suppose we throw a dart once so that it hits the target, and we observe where it lands.

To describe the possible outcomes of this experiment, it is natural to take as our sample space the set Ω of all the points in the target. It is convenient to describe these points by their rectangular coordinates, relative to a coordinate system with origin at the center of the target, so that each pair (x, y) of coordinates with $x^2+y^2 \le 1$ describes a possible outcome of the experiment. Then $\Omega = \{\,(x, y) : x^2 + y^2 \le 1\,\}$ is a subset of the Euclidean plane, and the event $E = \{\,(x, y) : y > 0\,\}$, for example, corresponds to the statement that the dart lands in the upper half of the target, and so forth. Unless there is reason to believe otherwise (and with experts at the game there may well be!), it is natural to assume that the coordinates are chosen *at random.* (When doing this with a computer, each coordinate is chosen uniformly from the interval $[-1, 1]$. If the resulting point does not lie inside the unit circle, the point is not counted.) Then the arguments used in the preceding example show that the probability of any elementary event, consisting of a single outcome, must be zero, and suggest that the probability of the event that the dart lands in any subset E of the target should be determined by what fraction of the target area lies in E. Thus,

$$P(E) = \frac{\text{area of } E}{\text{area of target}} = \frac{\text{area of } E}{\pi} \ .$$

This can be written in the form

$$P(E) = \int_E f(x)\, dx \ ,$$

where $f(x)$ is the constant function with value $1/\pi$. In particular, if $E = \{\, (x, y) : x^2 + y^2 \le a^2 \,\}$ is the event that the dart lands within distance $a < 1$ of the center of the target, then

$$P(E) = \frac{\pi a^2}{\pi} = a^2 \ .$$

For example, the probability that the dart lies within a distance 1/2 of the center is 1/4. $\square$

Example 2.9 In the dart game considered above, suppose that, instead of observing where the dart lands, we observe how far it lands from the center of the target.

In this case, we take as our sample space the set Ω of all circles with centers at the center of the target. It is convenient to describe these circles by their radii, so that each circle is identified by its radius r, $0 \le r \le 1$. In this way, we may regard Ω as the subset $[0, 1]$ of the real line.

What probabilities should we assign to the events E of Ω? If

$$E = \{\, r : 0 \le r \le a \,\} \ ,$$

then E occurs if the dart lands within a distance a of the center, that is, within the circle of radius a, and we saw in the previous example that under our assumptions the probability of this event is given by

$$P([0, a]) = a^2 \ .$$

More generally, if

$$E = \{\, r : a \le r \le b \,\} \ ,$$

then by our basic assumptions,

$$\begin{aligned} P(E) = P([a, b]) &= P([0, b]) - P([0, a]) \\ &= b^2 - a^2 \\ &= (b - a)(b + a) \\ &= 2(b - a)\frac{(b + a)}{2} \ . \end{aligned}$$

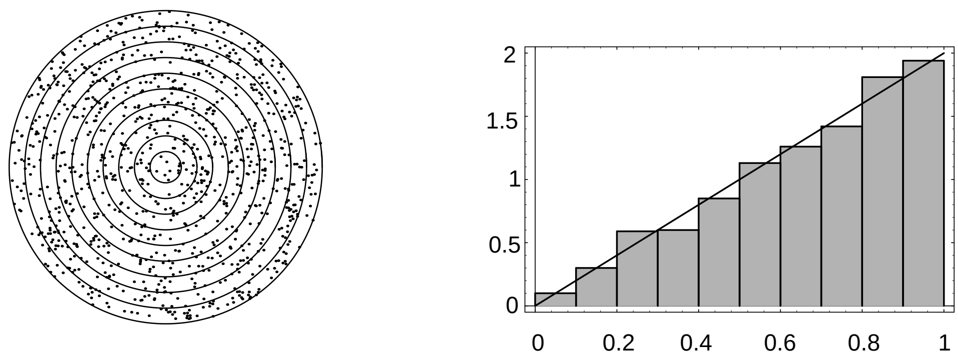


Figure 2.12: Distribution of dart distances in 400 throws.

Thus, $P(E)$ =2(length of E)(midpoint of E). Here we see that the probability assigned to the interval E depends not only on its length but also on its midpoint (i.e., not only on how long it is, but also on where it is). Roughly speaking, in this experiment, events of the form $E = [a, b]$ are more likely if they are near the rim of the target and less likely if they are near the center. (A common experience for beginners! The conclusion might well be different if the beginner is replaced by an expert.)

Again we can simulate this by computer. We divide the target area into ten concentric regions of equal thickness.

The computer program **Darts** throws n darts and records what fraction of the total falls in each of these concentric regions. The program **Areabargraph** then plots a bar graph with the *area* of the ith bar equal to the fraction of the total falling in the ith region. Running the program for 1000 darts resulted in the bar graph of Figure 2.12.

Note that here the heights of the bars are not all equal, but grow approximately linearly with r. In fact, the linear function $y = 2r$ appears to fit our bar graph quite well. This suggests that the probability that the dart falls within a distance a of the center should be given by the *area* under the graph of the function $y = 2r$ between 0 and a. This area is a^2, which agrees with the probability we have assigned above to this event. $\square$

Sample Space Coordinates

These examples suggest that for continuous experiments of this sort we should assign probabilities for the outcomes to fall in a given interval by means of the area under a suitable function.

More generally, we suppose that suitable coordinates can be introduced into the sample space Ω, so that we can regard Ω as a subset of $\mathbf{R}^n$. We call such a sample space a *continuous sample space.* We let X be a random variable which represents the outcome of the experiment. Such a random variable is called a *continuous random variable.* We then define a density function for X as follows.

Density Functions of Continuous Random Variables

Definition 2.1 Let X be a continuous real-valued random variable. A *density function* for X is a real-valued function f which satisfies

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

for all $a,\ b \in \mathbf{R}$. □

We note that it is *not* the case that all continuous real-valued random variables possess density functions. However, in this book, we will only consider continuous random variables for which density functions exist.

In terms of the density $f(x)$, if E is a subset of $\mathbf{R}$, then

$$P(X \in E) = \int_E f(x)\,dx\ .$$

The notation here assumes that E is a subset of $\mathbf{R}$ for which $\int_E f(x)\,dx$ makes sense.

Example 2.10 (Example 2.7 continued) In the spinner experiment, we choose for our set of outcomes the interval $0 \le x < 1$, and for our density function

$$f(x) = \begin{cases} 1, & \text{if } 0 \le x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

If E is the event that the head of the spinner falls in the upper half of the circle, then $E = \{\, x : 0 \le x \le 1/2 \,\}$, and so

$$P(E) = \int_0^{1/2} 1\,dx = \frac{1}{2}\ .$$

More generally, if E is the event that the head falls in the interval $[a, b]$, then

$$P(E) = \int_a^b 1\,dx = b - a\ .$$

□

Example 2.11 (Example 2.8 continued) In the first dart game experiment, we choose for our sample space a disc of unit radius in the plane and for our density function the function

$$f(x, y) = \begin{cases} 1/\pi, & \text{if } x^2 + y^2 \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

The probability that the dart lands inside the subset E is then given by

$$\begin{aligned} P(E) &= \int\int_E \frac{1}{\pi}\,dx\,dy \\ &= \frac{1}{\pi} \cdot (\text{area of } E)\ . \end{aligned}$$

□

In these two examples, the density function is constant and does not depend on the particular outcome. It is often the case that experiments in which the coordinates are chosen *at random* can be described by *constant* density functions, and, as in Section 1.2, we call such density functions *uniform* or *equiprobable.* Not all experiments are of this type, however.

Example 2.12 (Example 2.9 continued) In the second dart game experiment, we choose for our sample space the unit interval on the real line and for our density the function

$$f(r) = \begin{cases} 2r, & \text{if } 0 < r < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then the probability that the dart lands at distance r, $a \leq r \leq b$, from the center of the target is given by

$$\begin{aligned} P([a,b]) &= \int_a^b 2r\,dr \\ &= b^2 - a^2 \ . \end{aligned}$$

Here again, since the density is small when r is near 0 and large when r is near 1, we see that in this experiment the dart is more likely to land near the rim of the target than near the center. In terms of the bar graph of Example 2.9, the heights of the bars approximate the density function, while the areas of the bars approximate the probabilities of the subintervals (see Figure 2.12). $\square$

We see in this example that, unlike the case of discrete sample spaces, the value $f(x)$ of the density function for the outcome x is *not* the probability of x occurring (we have seen that this probability is always 0) and in general $f(x)$ is *not a probability at all.* In this example, if we take $\lambda = 2$ then $f(3/4) = 3/2$, which being bigger than 1, cannot be a probability.

Nevertheless, the density function f does contain all the probability information about the experiment, since the probabilities of all events can be derived from it. In particular, the probability that the outcome of the experiment falls in an interval $[a, b]$ is given by

$$P([a,b]) = \int_a^b f(x)\,dx \ ,$$

that is, by the *area* under the graph of the density function in the interval $[a, b]$. Thus, there is a close connection here between probabilities and areas. We have been guided by this close connection in making up our bar graphs; each bar is chosen so that its *area,* and not its height, represents the relative frequency of occurrence, and hence estimates the probability of the outcome falling in the associated interval.

In the language of the calculus, we can say that the probability of occurrence of an event of the form $[x, x + dx]$, where dx is small, is approximately given by

$$P([x, x + dx]) \approx f(x)dx \ ,$$

that is, by the area of the rectangle under the graph of f. Note that as $dx \to 0$, this probability $\to 0$, so that the probability $P(\{x\})$ of a single point is again 0, as in Example 2.7.

A glance at the graph of a density function tells us immediately which events of an experiment are more likely. Roughly speaking, we can say that where the density is large the events are more likely, and where it is small the events are less likely. In Example 2.4 the density function is largest at 1. Thus, given the two intervals $[0, a]$ and $[1, 1+a]$, where a is a small positive real number, we see that X is more likely to take on a value in the second interval than in the first.

Cumulative Distribution Functions of Continuous Random Variables

We have seen that density functions are useful when considering continuous random variables. There is another kind of function, closely related to these density functions, which is also of great importance. These functions are called *cumulative distribution* functions.

Definition 2.2 Let X be a continuous real-valued random variable. Then the cumulative distribution function of X is defined by the equation

$$F_X(x) = P(X \leq x)\ .$$

□

If X is a continuous real-valued random variable which possesses a density function, then it also has a cumulative distribution function, and the following theorem shows that the two functions are related in a very nice way.

Theorem 2.1 Let X be a continuous real-valued random variable with density function $f(x)$. Then the function defined by

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

is the cumulative distribution function of X. Furthermore, we have

$$\frac{d}{dx} F(x) = f(x)\ .$$

Proof. By definition,

$$F(x) = P(X \leq x)\ .$$

Let $E = (-\infty, x]$. Then

$$P(X \leq x) = P(X \in E)\ ,$$

which equals

$$\int_{-\infty}^{x} f(t)\, dt\ .$$

Applying the Fundamental Theorem of Calculus to the first equation in the statement of the theorem yields the second statement. □

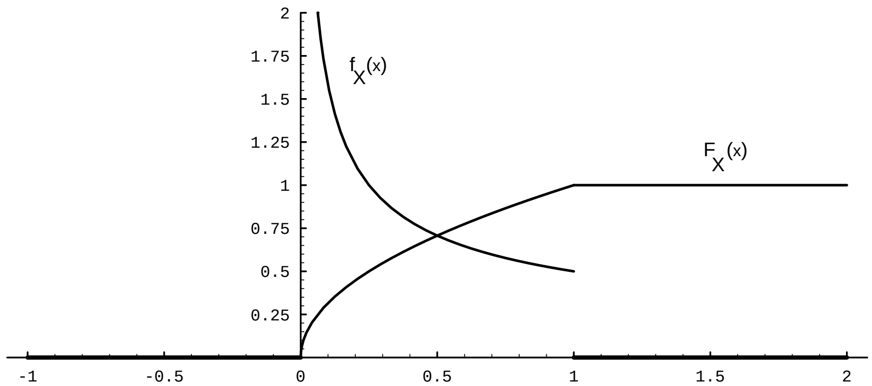


Figure 2.13: Distribution and density for $X = U^2$.

In many experiments, the density function of the relevant random variable is easy to write down. However, it is quite often the case that the cumulative distribution function is easier to obtain than the density function. (Of course, once we have the cumulative distribution function, the density function can easily be obtained by differentiation, as the above theorem shows.) We now give some examples which exhibit this phenomenon.

Example 2.13 A real number is chosen at random from $[0, 1]$ with uniform probability, and then this number is squared. Let X represent the result. What is the cumulative distribution function of X? What is the density of X?

We begin by letting U represent the chosen real number. Then $X = U^2$. If $0 \leq x \leq 1$, then we have

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(U^2 \leq x) \\ &= P(U \leq \sqrt{x}) \\ &= \sqrt{x} \ . \end{aligned}$$

It is clear that X always takes on a value between 0 and 1, so the cumulative distribution function of X is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ \sqrt{x}, & \text{if } 0 \leq x \leq 1, \\ 1, & \text{if } x \geq 1. \end{cases}$$

From this we easily calculate that the density function of X is

$$f_X(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ 1/(2\sqrt{x}), & \text{if } 0 \leq x \leq 1, \\ 0, & \text{if } x > 1. \end{cases}$$

Note that $F_X(x)$ is continuous, but $f_X(x)$ is not. (See Figure 2.13.) □

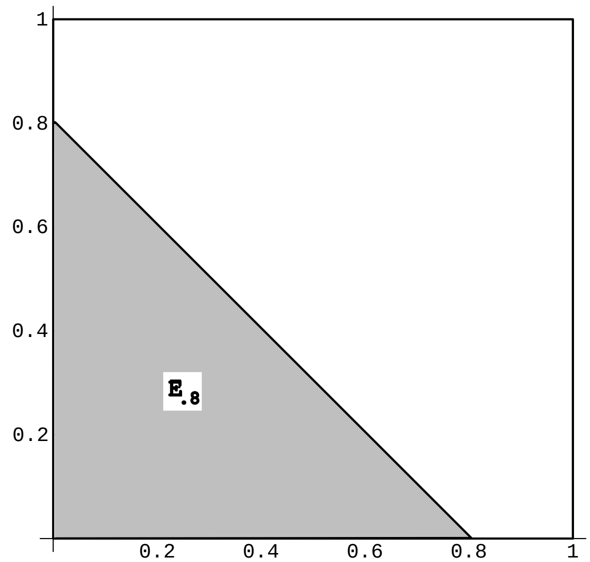


Figure 2.14: Calculation of distribution function for Example 2.14.

When referring to a continuous random variable X (say with a uniform density function), it is customary to say that "X is uniformly *distributed* on the interval $[a, b]$." It is also customary to refer to the cumulative distribution function of X as the distribution function of X. Thus, the word "distribution" is being used in several different ways in the subject of probability. (Recall that it also has a meaning when discussing discrete random variables.) When referring to the cumulative distribution function of a continuous random variable X, we will always use the word "cumulative" as a modifier, unless the use of another modifier, such as "normal" or "exponential," makes it clear. Since the phrase "uniformly densitied on the interval $[a, b]$" is not acceptable English, we will have to say "uniformly distributed" instead.

Example 2.14 In Example 2.4, we considered a random variable, defined to be the sum of two random real numbers chosen uniformly from $[0, 1]$. Let the random variables X and Y denote the two chosen real numbers. Define $Z = X + Y$. We will now derive expressions for the cumulative distribution function and the density function of Z.

Here we take for our sample space Ω the unit square in $\mathbf{R}^2$ with uniform density. A point $\omega \in \Omega$ then consists of a pair (x, y) of numbers chosen at random. Then $0 \le Z \le 2$. Let E_z denote the event that $Z \le z$. In Figure 2.14, we show the set $E_{.8}$. The event E_z, for any z between 0 and 1, looks very similar to the shaded set in the figure. For $1 < z \le 2$, the set E_z looks like the unit square with a triangle removed from the upper right-hand corner. We can now calculate the probability distribution F_Z of Z; it is given by

$$
\begin{aligned}
F_Z(z) &= P(Z \le z) \\
&= \text{Area of } E_z
\end{aligned}
$$

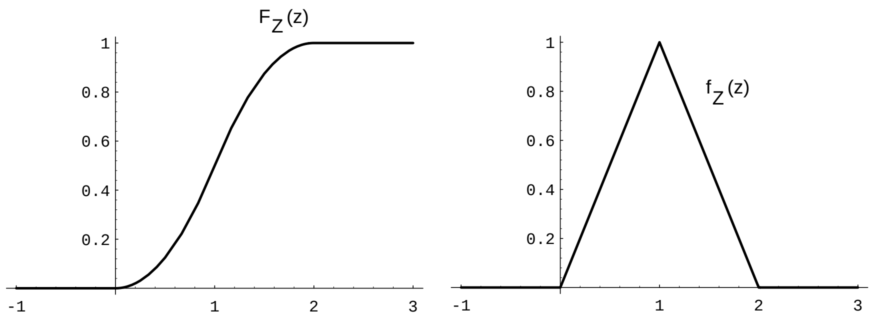


Figure 2.15: Distribution and density functions for Example 2.14.

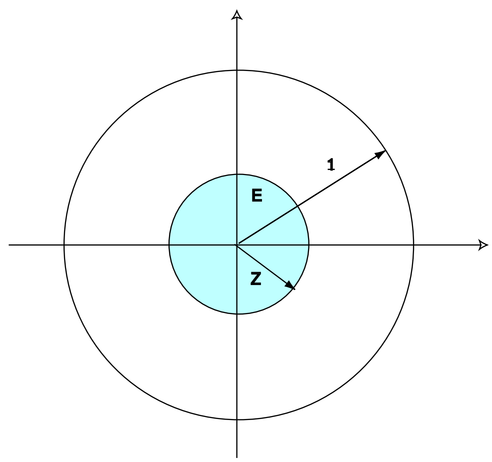


Figure 2.16: Calculation of F_z for Example 2.15.

$$= \begin{cases} 0, & \text{if } z < 0, \\ (1/2)z^2, & \text{if } 0 \le z \le 1, \\ 1 - (1/2)(2-z)^2, & \text{if } 1 \le z \le 2, \\ 1, & \text{if } 2 < z. \end{cases}$$

The density function is obtained by differentiating this function:

$$f_Z(z) = \begin{cases} 0, & \text{if } z < 0, \\ z, & \text{if } 0 \le z \le 1, \\ 2-z, & \text{if } 1 \le z \le 2, \\ 0, & \text{if } 2 < z. \end{cases}$$

The reader is referred to Figure 2.15 for the graphs of these functions. □

Example 2.15 In the dart game described in Example 2.8, what is the distribution of the distance of the dart from the center of the target? What is its density?

Here, as before, our sample space Ω is the unit disk in $\mathbf{R}^2$, with coordinates (X, Y). Let $Z = \sqrt{X^2 + Y^2}$ represent the distance from the center of the target. Let

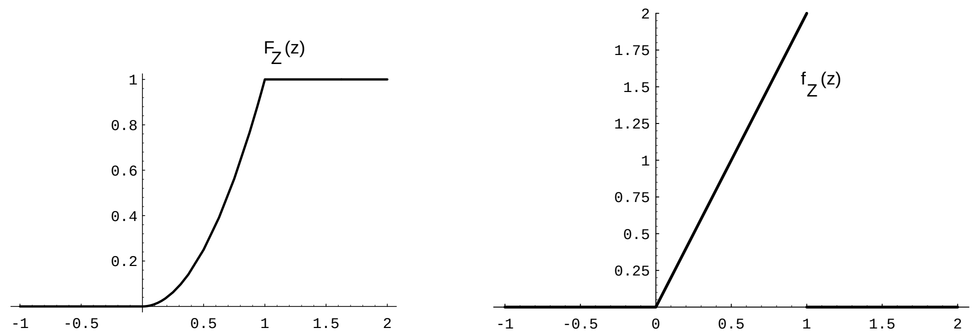


Figure 2.17: Distribution and density for $Z = \sqrt{X^2 + Y^2}$.

E be the event $\{Z \leq z\}$. Then the distribution function F_Z of Z (see Figure 2.16) is given by

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= \frac{\text{Area of } E}{\text{Area of target}} \ . \end{aligned}$$

Thus, we easily compute that

$$F_Z(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ z^2, & \text{if } 0 \leq z \leq 1, \\ 1, & \text{if } z > 1. \end{cases}$$

The density $f_Z(z)$ is given again by the derivative of $F_Z(z)$:

$$f_Z(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ 2z, & \text{if } 0 \leq z \leq 1, \\ 0, & \text{if } z > 1. \end{cases}$$

The reader is referred to Figure 2.17 for the graphs of these functions.

We can verify this result by simulation, as follows: We choose values for X and Y at random from $[0, 1]$ with uniform distribution, calculate $Z = \sqrt{X^2 + Y^2}$, check whether $0 \leq Z \leq 1$, and present the results in a bar graph (see Figure 2.18). □

Example 2.16 Suppose Mr. and Mrs. Lockhorn agree to meet at the Hanover Inn between 5:00 and 6:00 P.M. on Tuesday. Suppose each arrives at a time between 5:00 and 6:00 chosen at random with uniform probability. What is the distribution function for the length of time that the first to arrive has to wait for the other? What is the density function?

Here again we can take the unit square to represent the sample space, and (X, Y) as the arrival times (after 5:00 P.M.) for the Lockhorns. Let $Z = |X - Y|$. Then we have $F_X(x) = x$ and $F_Y(y) = y$. Moreover (see Figure 2.19),

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(|X - Y| \leq z) \\ &= \text{Area of } E \ . \end{aligned}$$

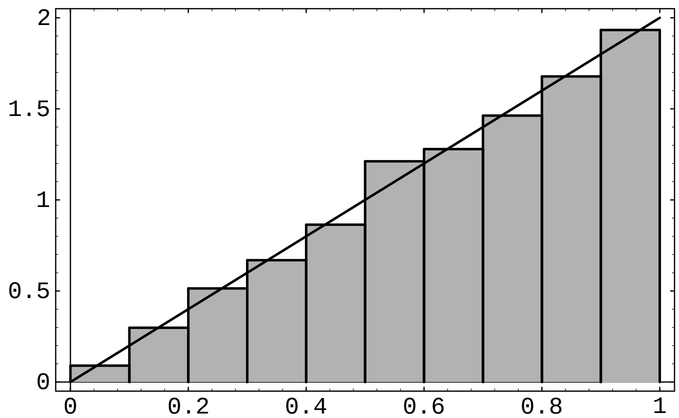


Figure 2.18: Simulation results for Example 2.15.

Thus, we have

$$F_Z(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ 1 - (1-z)^2, & \text{if } 0 \leq z \leq 1, \\ 1, & \text{if } z > 1. \end{cases}$$

The density $f_Z(z)$ is again obtained by differentiation:

$$f_Z(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ 2(1-z), & \text{if } 0 \leq z \leq 1, \\ 0, & \text{if } z > 1. \end{cases}$$

□

Example 2.17 There are many occasions where we observe a sequence of occurrences which occur at "random" times. For example, we might be observing emissions of a radioactive isotope, or cars passing a milepost on a highway, or light bulbs burning out. In such cases, we might define a random variable X to denote the time between successive occurrences. Clearly, X is a continuous random variable whose range consists of the non-negative real numbers. It is often the case that we can model X by using the *exponential density*. This density is given by the formula

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{if } t \geq 0, \\ 0, & \text{if } t < 0. \end{cases}$$

The number λ is a non-negative real number, and represents the reciprocal of the average value of X. (This will be shown in Chapter 6.) Thus, if the average time between occurrences is 30 minutes, then $\lambda = 1/30$. A graph of this density function with $\lambda = 1/30$ is shown in Figure 2.20. One can see from the figure that even though the average value is 30, occasionally much larger values are taken on by X.

Suppose that we have bought a computer that contains a Warp 9 hard drive. The salesperson says that the average time between breakdowns of this type of hard drive is 30 months. It is often assumed that the length of time between breakdowns

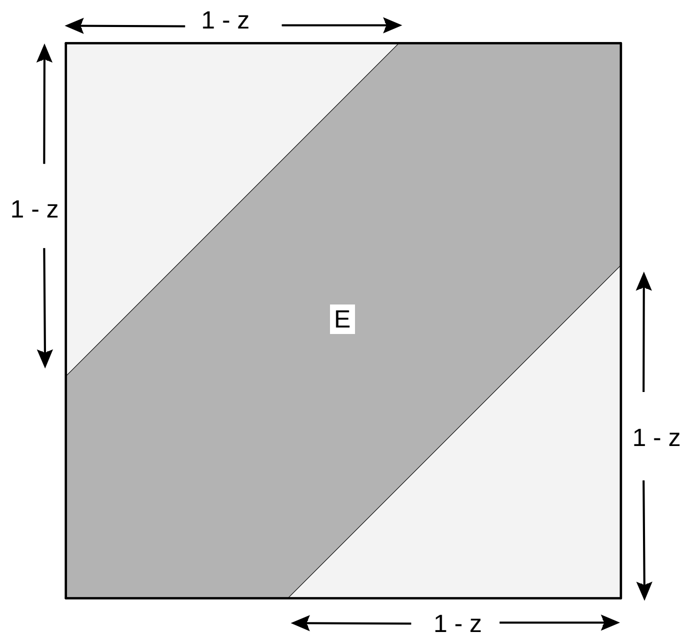


Figure 2.19: Calculation of F_Z.

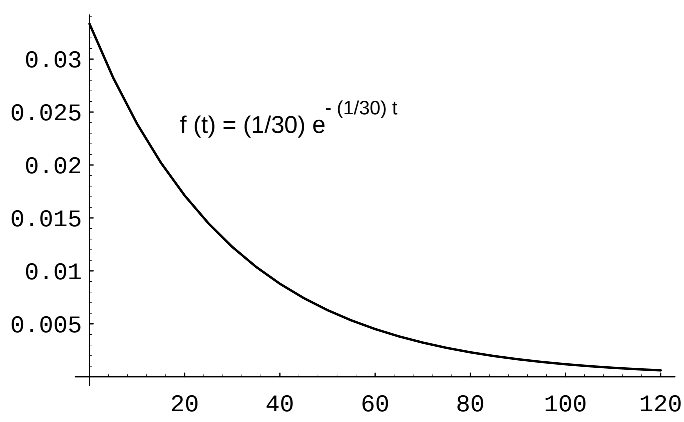


Figure 2.20: Exponential density with $\lambda = 1/30$.

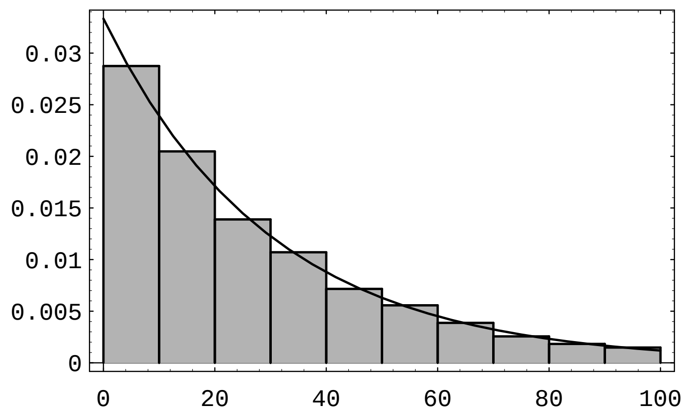


Figure 2.21: Residual lifespan of a hard drive.

is distributed according to the exponential density. We will assume that this model applies here, with $\lambda = 1/30$.

Now suppose that we have been operating our computer for 15 months. We assume that the original hard drive is still running. We ask how long we should expect the hard drive to continue to run. One could reasonably expect that the hard drive will run, on the average, another 15 months. (One might also guess that it will run more than 15 months, since the fact that it has already run for 15 months implies that we don't have a lemon.) The time which we have to wait is a new random variable, which we will call Y. Obviously, $Y = X - 15$. We can write a computer program to produce a sequence of simulated Y-values. To do this, we first produce a sequence of X's, and discard those values which are less than or equal to 15 (these values correspond to the cases where the hard drive has quit running before 15 months). To simulate a value of X, we compute the value of the expression

$$\left(-\frac{1}{\lambda}\right)\log(rnd)\ ,$$

where *rnd* represents a random real number between 0 and 1. (That this expression has the exponential density will be shown in Chapter 4.3.) Figure 2.21 shows an area bar graph of 10,000 simulated Y-values.

The average value of Y in this simulation is 29.74, which is closer to the original average life span of 30 months than to the value of 15 months which was guessed above. Also, the distribution of Y is seen to be close to the distribution of X. It is in fact the case that X and Y have the same distribution. This property is called the *memoryless property*, because the amount of time that we have to wait for an occurrence does not depend on how long we have already waited. The only continuous density function with this property is the exponential density. $\square$

Assignment of Probabilities

A fundamental question in practice is: How shall we choose the probability density function in describing any given experiment? The answer depends to a great extent on the amount and kind of information available to us about the experiment. In some cases, we can see that the outcomes are equally likely. In some cases, we can see that the experiment resembles another already described by a known density. In some cases, we can run the experiment a large number of times and make a reasonable guess at the density on the basis of the observed distribution of outcomes, as we did in Chapter 1. In general, the problem of choosing the right density function for a given experiment is a central problem for the experimenter and is not always easy to solve (see Example 2.6). We shall not examine this question in detail here but instead shall assume that the right density is already known for each of the experiments under study.

The introduction of suitable coordinates to describe a continuous sample space, and a suitable density to describe its probabilities, is not always so obvious, as our final example shows.

Infinite Tree

Example 2.18 Consider an experiment in which a fair coin is tossed repeatedly, without stopping. We have seen in Example 1.6 that, for a coin tossed n times, the natural sample space is a binary tree with n stages. On this evidence we expect that for a coin tossed repeatedly, the natural sample space is a binary tree with an infinite number of stages, as indicated in Figure 2.22.

It is surprising to learn that, although the n-stage tree is obviously a finite sample space, the unlimited tree can be described as a continuous sample space. To see how this comes about, let us agree that a typical outcome of the unlimited coin tossing experiment can be described by a sequence of the form $\omega = \{\text{H H T H T T H}\ldots\}$. If we write 1 for H and 0 for T, then $\omega = \{1\ 1\ 0\ 1\ 0\ 0\ 1\ldots\}$. In this way, each outcome is described by a sequence of 0's and 1's.

Now suppose we think of this sequence of 0's and 1's as the binary expansion of some real number $x = .1101001\cdots$ lying between 0 and 1. (A *binary expansion* is like a decimal expansion but based on 2 instead of 10.) Then each outcome is described by a value of x, and in this way x becomes a coordinate for the sample space, taking on all real values between 0 and 1. (We note that it is possible for two different sequences to correspond to the same real number; for example, the sequences $\{\text{T H H H H H}\ldots\}$ and $\{\text{H T T T T T}\ldots\}$ both correspond to the real number $1/2$. We will not concern ourselves with this apparent problem here.)

What probabilities should be assigned to the events of this sample space? Consider, for example, the event E consisting of all outcomes for which the first toss comes up heads and the second tails. Every such outcome has the form $.10{*}{*}{*}{*}\cdots$, where $*$ can be either 0 or 1. Now if x is our real-valued coordinate, then the value of x for every such outcome must lie between $1/2 = .10000\cdots$ and $3/4 = .11000\cdots$, and moreover, every value of x between 1/2 and 3/4 has a binary expansion of the

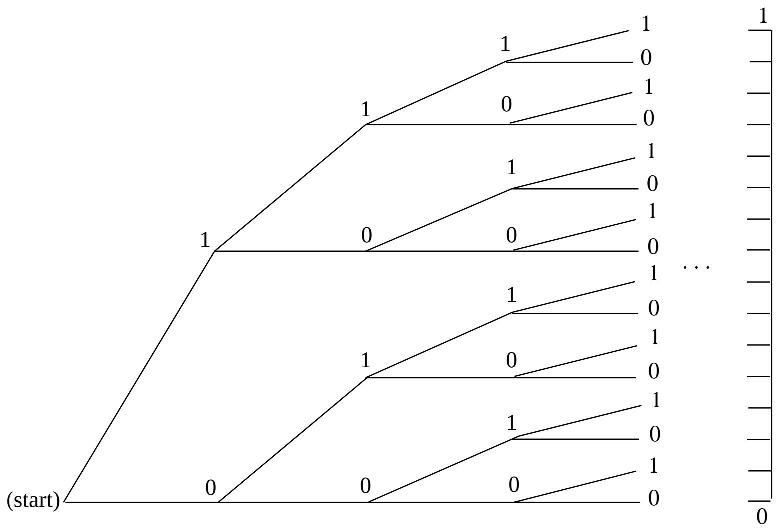


Figure 2.22: Tree for infinite number of tosses of a coin.

form $.10****\cdots$. This means that $\omega \in E$ if and only if $1/2 \leq x < 3/4$, and in this way we see that we can describe E by the interval $[1/2, 3/4)$. More generally, every event consisting of outcomes for which the results of the first n tosses are prescribed is described by a binary interval of the form $[k/2^n, (k+1)/2^n)$.

We have already seen in Section 1.2 that in the experiment involving n tosses, the probability of any one outcome must be exactly $1/2^n$. It follows that in the unlimited toss experiment, the probability of any event consisting of outcomes for which the results of the first n tosses are prescribed must also be $1/2^n$. But $1/2^n$ is exactly the length of the interval of x-values describing E! Thus we see that, just as with the spinner experiment, the probability of an event E is determined by what fraction of the unit interval lies in E.

Consider again the statement: The probability is 1/2 that a fair coin will turn up heads when tossed. We have suggested that one interpretation of this statement is that if we toss the coin indefinitely the proportion of heads will approach 1/2. That is, in our correspondence with binary sequences we expect to get a binary sequence with the proportion of 1's tending to 1/2. The event E of binary sequences for which this is true is a proper subset of the set of all possible binary sequences. It does not contain, for example, the sequence 011011011... (i.e., (011) repeated again and again). The event E is actually a very complicated subset of the binary sequences, but its probability can be determined as a limit of probabilities for events with a finite number of outcomes whose probabilities are given by finite tree measures. When the probability of E is computed in this way, its value is found to be 1. This remarkable result is known as the *Strong Law of Large Numbers* (or *Law of Averages*) and is one justification for our frequency concept of probability. We shall prove a weak form of this theorem in Chapter 8. $\square$

Exercises

1 Suppose you choose *at random* a real number X from the interval $[2, 10]$.

(a) Find the density function $f(x)$ and the probability of an event E for this experiment, where E is a subinterval $[a, b]$ of $[2, 10]$.

(b) From (a), find the probability that $X > 5$, that $5 < X < 7$, and that $X^2 - 12X + 35 > 0$.

2 Suppose you choose a real number X from the interval $[2, 10]$ with a density function of the form

$$f(x) = Cx ,$$

where C is a constant.

(a) Find C.

(b) Find $P(E)$, where $E = [a, b]$ is a subinterval of $[2, 10]$.

(c) Find $P(X > 5)$, $P(X < 7)$, and $P(X^2 - 12X + 35 > 0)$.

3 Same as Exercise 2, but suppose

$$f(x) = \frac{C}{x} .$$

4 Suppose you throw a dart at a circular target of radius 10 inches. Assuming that you hit the target and that the coordinates of the outcomes are chosen at random, find the probability that the dart falls

(a) within 2 inches of the center.

(b) within 2 inches of the rim.

(c) within the first quadrant of the target.

(d) within the first quadrant and within 2 inches of the rim.

5 Suppose you are watching a radioactive source that emits particles at a rate described by the exponential density

$$f(t) = \lambda e^{-\lambda t} ,$$

where $\lambda = 1$, so that the probability $P(0, T)$ that a particle will appear in the next T seconds is $P([0, T]) = \int_0^T \lambda e^{-\lambda t}\, dt$. Find the probability that a particle (not necessarily the first) will appear

(a) within the next second.

(b) within the next 3 seconds.

(c) between 3 and 4 seconds from now.

(d) after 4 seconds from now.

6 Assume that a new light bulb will burn out after t hours, where t is chosen from $[0, \infty)$ with an exponential density

$$f(t) = \lambda e^{-\lambda t} \ .$$

In this context, λ is often called the *failure rate* of the bulb.

(a) Assume that $\lambda = 0.01$, and find the probability that the bulb will *not* burn out before T hours. This probability is often called the *reliability* of the bulb.

(b) For what T is the reliability of the bulb $= 1/2$?

7 Choose a number B *at random* from the interval $[0, 1]$ with uniform density. Find the probability that

(a) $1/3 < B < 2/3$.

(b) $|B - 1/2| \leq 1/4$.

(c) $B < 1/4$ or $1 - B < 1/4$.

(d) $3B^2 < B$.

8 Choose independently two numbers B and C *at random* from the interval $[0, 1]$ with uniform density. Note that the point (B, C) is then chosen *at random* in the unit square. Find the probability that

(a) $B + C < 1/2$.

(b) $BC < 1/2$.

(c) $|B - C| < 1/2$.

(d) $\max\{B, C\} < 1/2$.

(e) $\min\{B, C\} < 1/2$.

(f) $B < 1/2$ and $1 - C < 1/2$.

(g) conditions (c) and (f) both hold.

(h) $B^2 + C^2 \leq 1/2$.

(i) $(B - 1/2)^2 + (C - 1/2)^2 < 1/4$.

9 Suppose that we have a sequence of occurrences. We assume that the time X between occurrences is exponentially distributed with $\lambda = 1/10$, so on the average, there is one occurrence every 10 minutes (see Example 2.17). You come upon this system at time 100, and wait until the next occurrence. Make a conjecture concerning how long, on the average, you will have to wait. Write a program to see if your conjecture is right.

10 As in Exercise 9, assume that we have a sequence of occurrences, but now assume that the time X between occurrences is uniformly distributed between 5 and 15. As before, you come upon this system at time 100, and wait until the next occurrence. Make a conjecture concerning how long, on the average, you will have to wait. Write a program to see if your conjecture is right.

11 For examples such as those in Exercises 9 and 10, it might seem that at least you should not have to wait on average *more* than 10 minutes if the average time between occurrences is 10 minutes. Alas, even this is not true. To see why, consider the following assumption about the times between occurrences. Assume that the time between occurrences is 3 minutes with probability .9 and 73 minutes with probability .1. Show by simulation that the average time between occurrences is 10 minutes, but that if you come upon this system at time 100, your average waiting time is more than 10 minutes.

12 Take a stick of unit length and break it into three pieces, choosing the break points at random. (The break points are assumed to be chosen simultaneously.) What is the probability that the three pieces can be used to form a triangle? *Hint*: The sum of the lengths of any two pieces must exceed the length of the third, so each piece must have length $< 1/2$. Now use Exercise 8(g).

13 Take a stick of unit length and break it into two pieces, choosing the break point at random. Now break the longer of the two pieces at a random point. What is the probability that the three pieces can be used to form a triangle?

14 Choose independently two numbers B and C *at random* from the interval $[-1, 1]$ with uniform distribution, and consider the quadratic equation

$$x^2 + Bx + C = 0 .$$

Find the probability that the roots of this equation

(a) are both real.

(b) are both positive.

Hints: (a) requires $0 \leq B^2 - 4C$, (b) requires $0 \leq B^2 - 4C$, $B \leq 0$, $0 \leq C$.

15 At the Tunbridge World's Fair, a coin toss game works as follows. Quarters are tossed onto a checkerboard. The management keeps all the quarters, but for each quarter landing entirely within one square of the checkerboard the management pays a dollar. Assume that the edge of each square is twice the diameter of a quarter, and that the outcomes are described by coordinates chosen *at random.* Is this a fair game?

16 Three points are chosen *at random* on a circle of *unit circumference.* What is the probability that the triangle defined by these points as vertices has three acute angles? *Hint*: One of the angles is obtuse if and only if all three points lie in the same semicircle. Take the circumference as the interval $[0, 1]$. Take one point at 0 and the others at B and C.

17 Write a program to choose a random number X in the interval $[2, 10]$ 1000 times and record what fraction of the outcomes satisfy $X > 5$, what fraction satisfy $5 < X < 7$, and what fraction satisfy $x^2 - 12x + 35 > 0$. How do these results compare with Exercise 1?

18 Write a program to choose a point (X, Y) *at random* in a square of side 20 inches, doing this 10,000 times, and recording what fraction of the outcomes fall within 19 inches of the center; of these, what fraction fall between 8 and 10 inches of the center; and, of these, what fraction fall within the first quadrant of the square. How do these results compare with those of Exercise 4?

19 Write a program to simulate the problem describe in Exercise 7 (see Exercise 17). How do the simulation results compare with the results of Exercise 7?

20 Write a program to simulate the problem described in Exercise 12.

21 Write a program to simulate the problem described in Exercise 16.

22 Write a program to carry out the following experiment. A coin is tossed 100 times and the number of heads that turn up is recorded. This experiment is then repeated 1000 times. Have your program plot a bar graph for the proportion of the 1000 experiments in which the number of heads is n, for each n in the interval $[35, 65]$. Does the bar graph look as though it can be fit with a normal curve?

23 Write a program that picks a random number between 0 and 1 and computes the negative of its logarithm. Repeat this process a large number of times and plot a bar graph to give the number of times that the outcome falls in each interval of length 0.1 in $[0, 10]$. On this bar graph plot a graph of the density $f(x) = e^{-x}$. How well does this density fit your graph?