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.
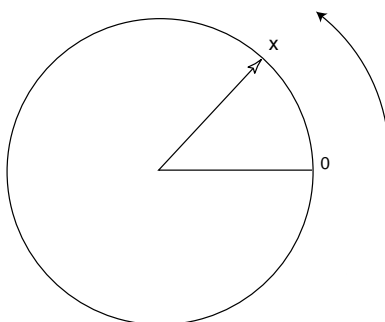
Figure 2.1: A spinner.

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*.   □

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 $E = \{ (x, y) : y \leq x^2 \}$.

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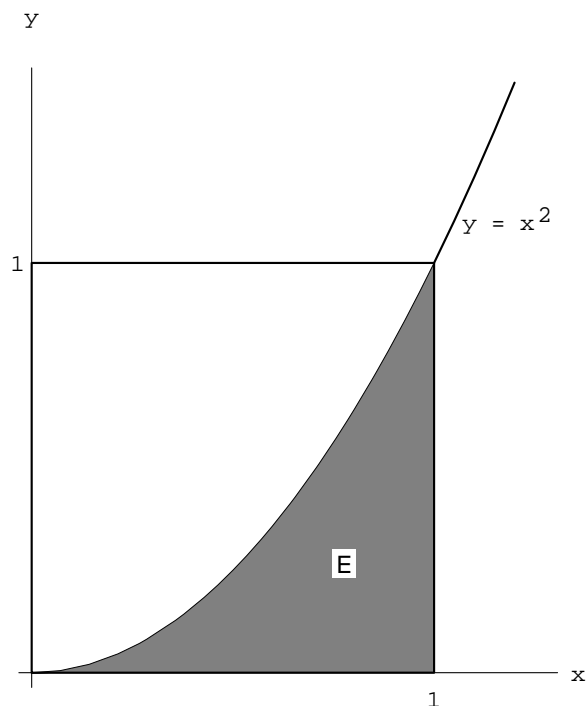
Figure 2.2: Area under $y = x^2$.

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,

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

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

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).
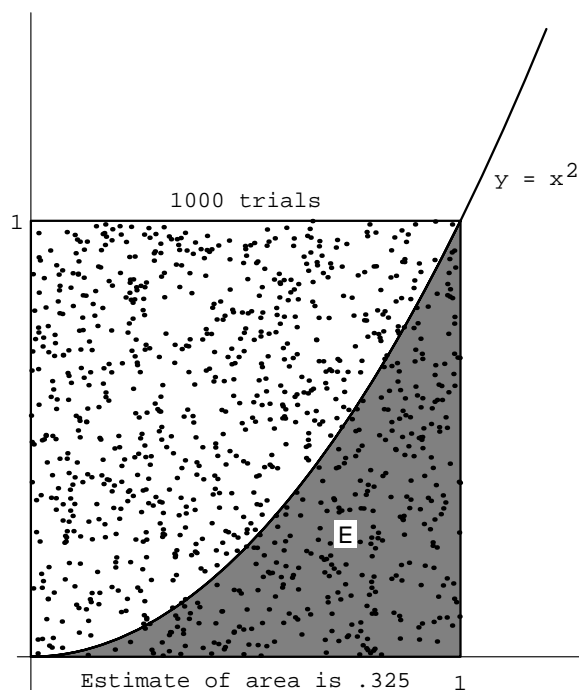
Figure 2.3: Computing the area by simulation.

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).
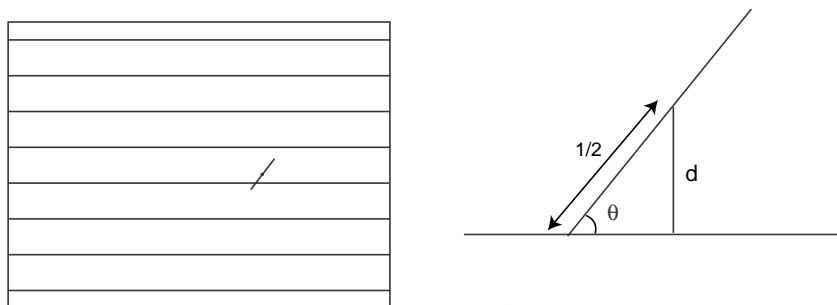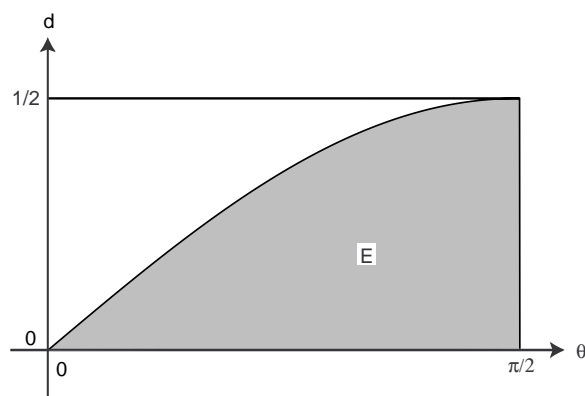
Figure 2.4: Buffon's experiment.

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
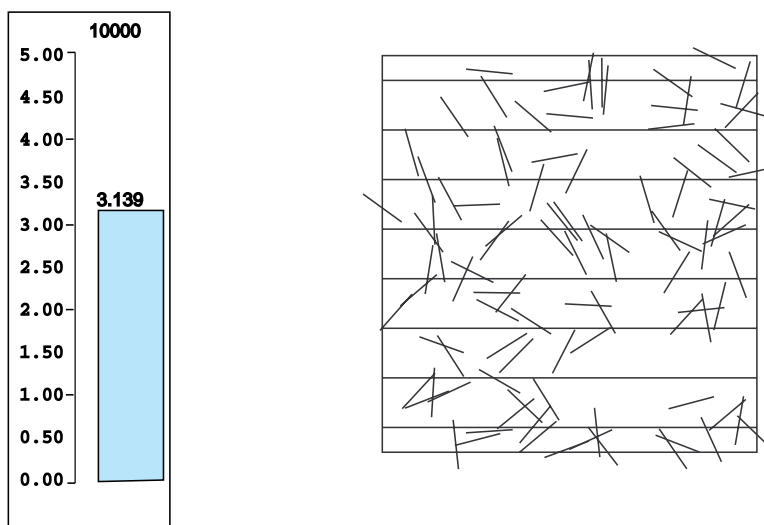
Figure 2.6: Simulation of Buffon's needle experiment.

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. $\square$

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

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

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 $a \leq b$, then we can use this function to calculate the probability that $a \leq X \leq b$. 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.
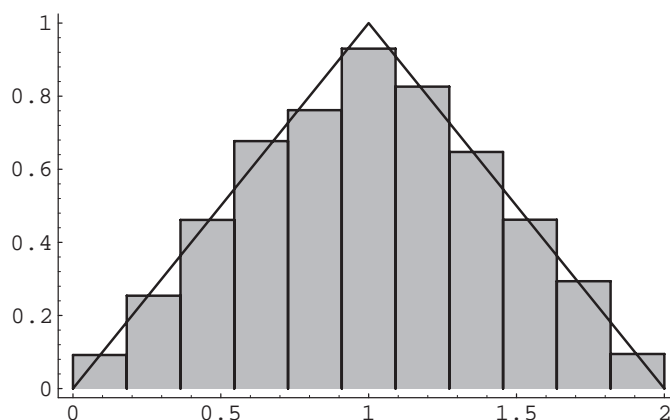
Figure 2.7: Sum of two random numbers.

$[a, b]$ 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 $f(x)$. □

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. □

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
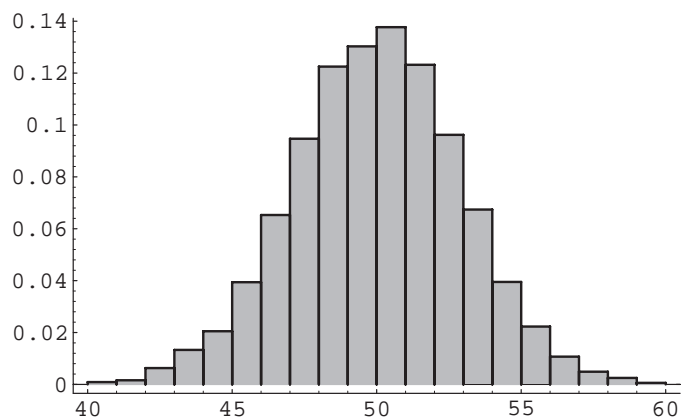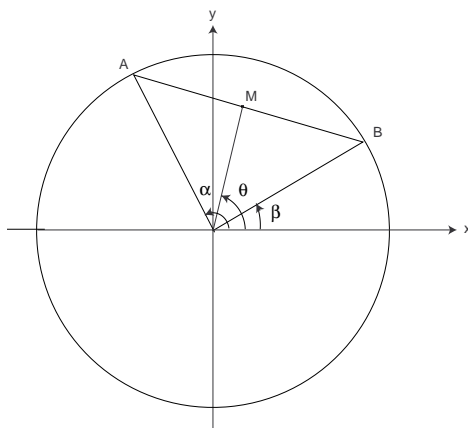
Figure 2.8: Sum of 100 random numbers.

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.
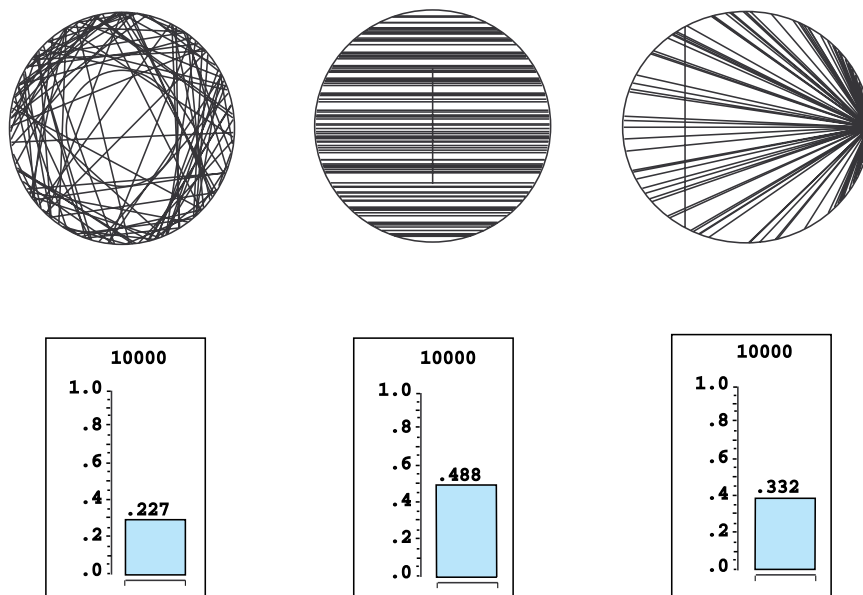
Figure 2.10: Bertrand's paradox.

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

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

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

$$P\left(0 \leq X < \frac{1}{2}\right) = P\left(\frac{1}{2} \leq X < 1\right) = \frac{1}{2} \; .$$

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

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

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:

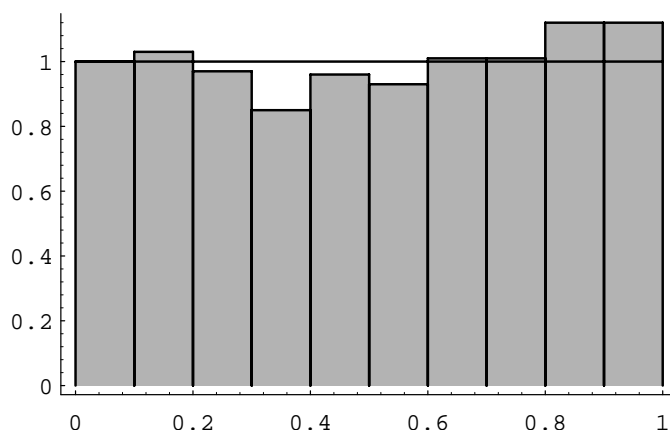$$P(E) = \sum_{\omega \in E} m(\omega) \; .$$

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.   □

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 $x^2 + y^2 \leq 1$ describes a possible outcome of the experiment. Then $\Omega = \{\,(x, y) : x^2 + y^2 \leq 1\,\}$ 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 $1/4$.   □

Example 2.9 In the dart game considered above, suppose that, instead of observ-
ing 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,

$$
\begin{aligned}
P(E) = P([a, b]) \ &= \ P([0, b]) - P([0, a]) \\
&= \ b^2 - a^2 \\
&= \ (b - a)(b + a) \\
&= \ 2(b - a)\frac{(b + a)}{2} \ .
\end{aligned}
$$
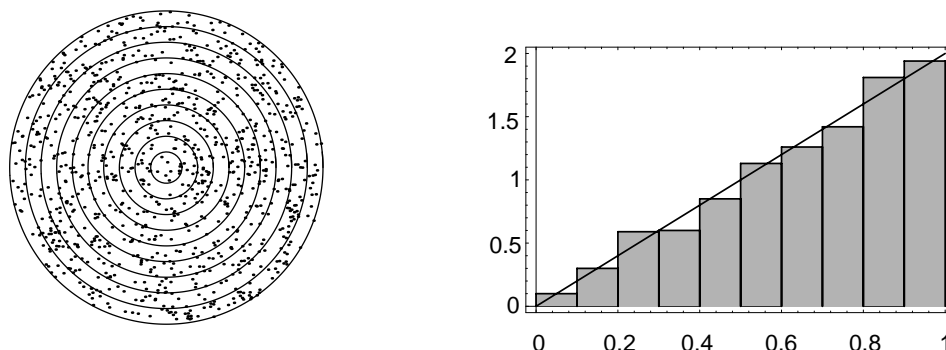
Figure 2.12: Distribution of dart distances in 400 throws.

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.   □

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

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

for all $a,\ b \in \mathbf{R}$. $\qquad\square$

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 $0 \leq x < 1$, and for our density function

$$f(x) = \left\{ \begin{array}{ll} 1, & \text{if } 0 \leq x < 1, \\ 0, & \text{otherwise.} \end{array} \right.$$

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

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

$\qquad\square$

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

$$f(x, y) = \left\{ \begin{array}{ll} 1/\pi, & \text{if } x^2 + y^2 \leq 1, \\ 0, & \text{otherwise.} \end{array} \right.$$

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

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

$\qquad\square$

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). □

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

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

$\square$

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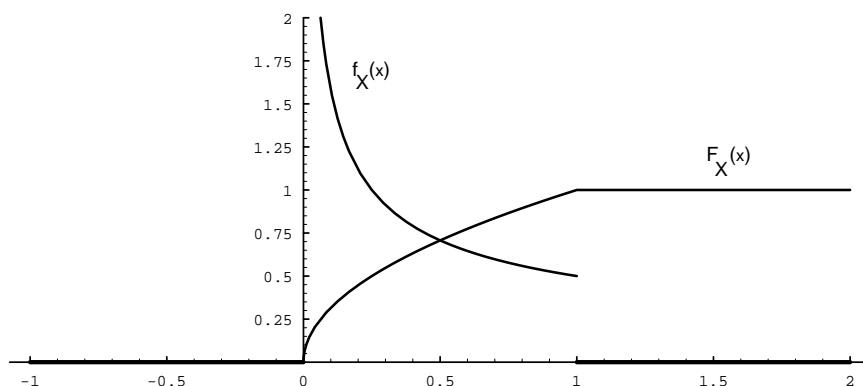$$F(x) = P(X \leq x) \ .$$

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. $\square$

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

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

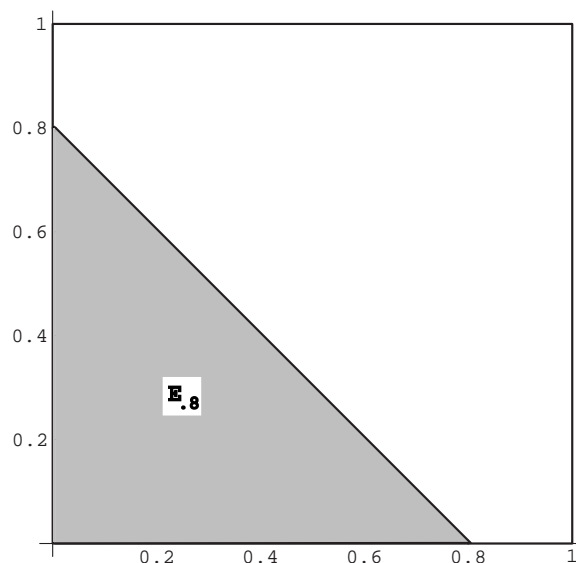Note that $F_X(x)$ is continuous, but $f_X(x)$ is not. (See Figure 2.13.)   □

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

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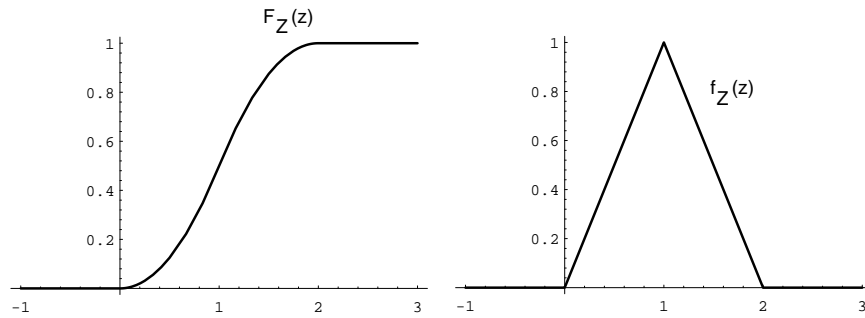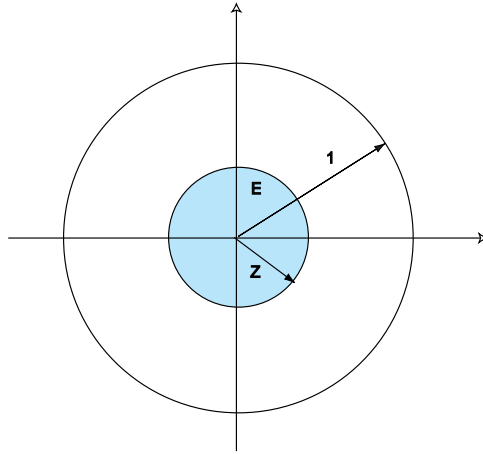&= \text{Area of } E_z
\end{aligned}
$$

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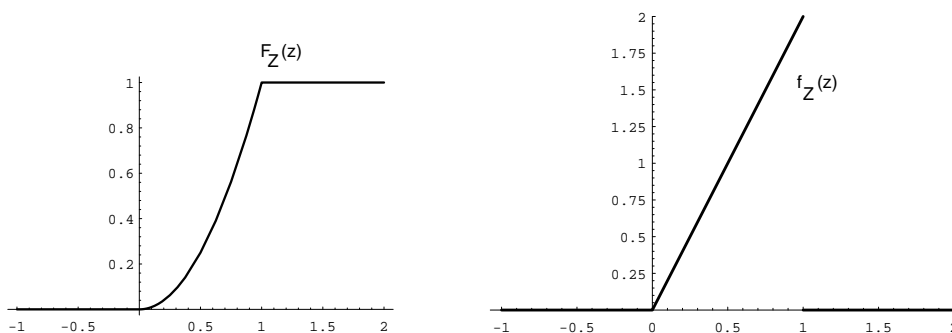
Figure 2.17: Distribution and density for $Z = \sqrt{X^2 + Y^2}$.

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),

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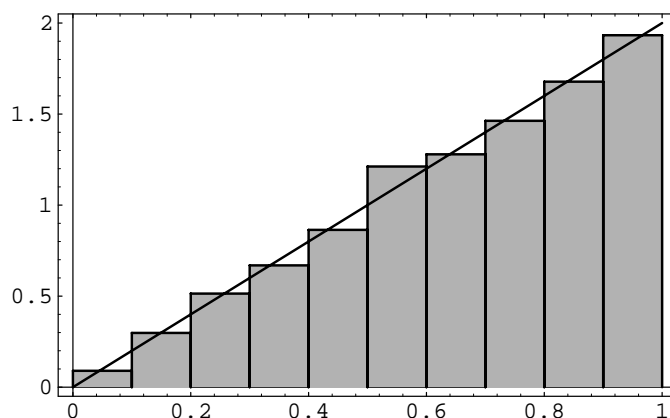&= \text{Area of } E .
\end{aligned}
$$

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:

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

$\square$

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
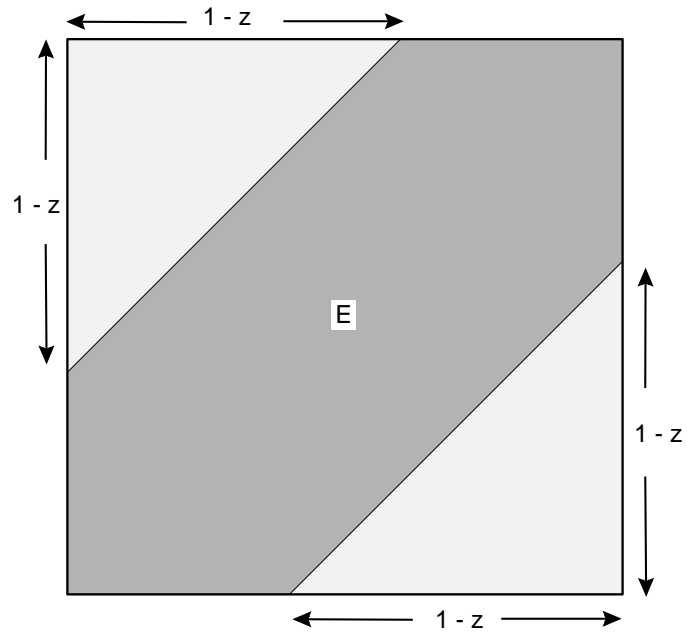
Figure 2.19: Calculation of F_Z.

$$f(t) = (1/30)\, e^{-(1/30)\, t}$$
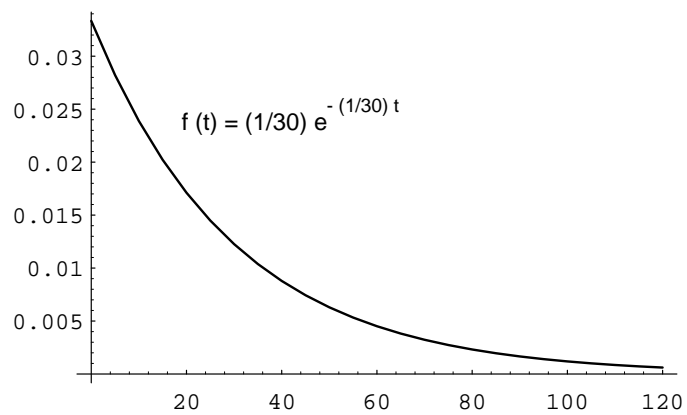
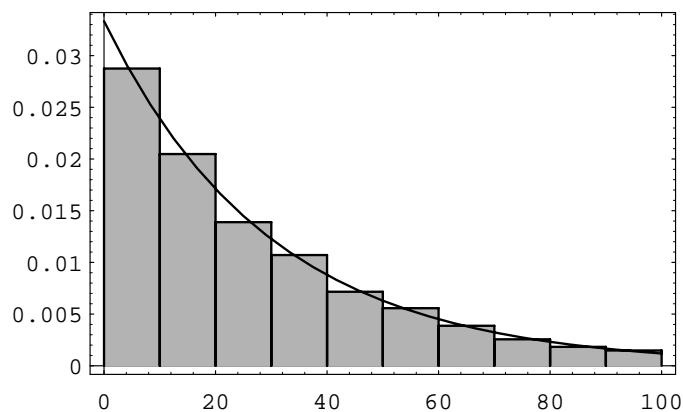Figure 2.20: Exponential density with $\lambda = 1/30$.

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.   □

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 $\omega = \{$H H T H T T H...$\}$. If we write 1 for H and 0 for T, then $\omega = \{1\ 1\ 0\ 1\ 0\ 0\ 1...\}$. 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 $\{$T H H H H H...$\}$ and $\{$H T T T T T...$\}$ 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
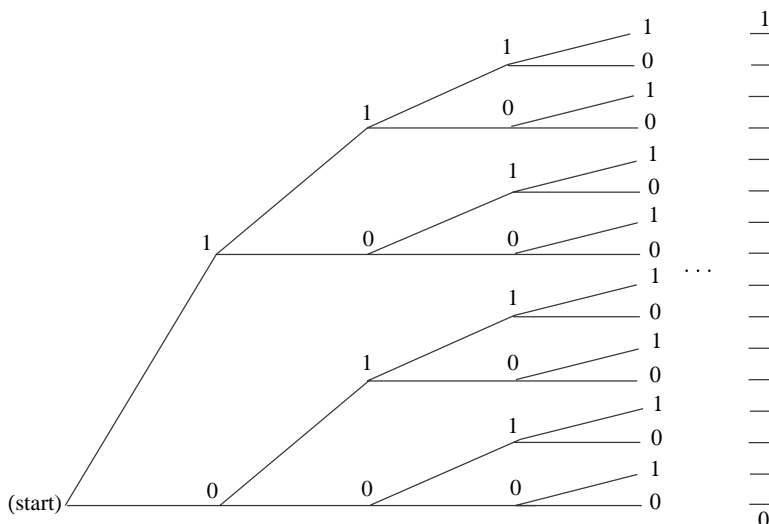
Figure 2.22: Tree for infinite number of tosses of a coin.

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 $011011011\ldots$ (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.   □

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 $0 \le B^2 - 4C$, (b) requires $0 \le B^2 - 4C$, $B \le 0$, $0 \le C$.

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?