

Chapter 2

Densidades de Probabilidad Continuas

2.1 Simulación de Probabilidades Continuas

En esta sección describiremos el uso de simulaciones por computadora de experimentos que tienen un conjunto posible de resultados con estructura de un continuo.

Probabilidades

Ejemplo 2.1 Para comenzar construimos una ruleta, que consiste en un círculo de *circunferencia unidad* y un puntero, como se ilustra en la Figura 2.1. Escogemos un punto de la circunferencia y lo etiquetamos 0, y luego etiquetamos cada uno de los puntos de la circunferencia según su distancia x con respecto al punto 0, midiendo en sentido contrario a las agujas del reloj. El experimento consiste en registrar la etiqueta del punto sobre la ruleta donde el puntero se detiene cada vez que lo hacemos girar. Utilizamos la variable aleatoria X para denotar el valor de este resultado. El espacio muestral es claramente el intervalo $[0, 1)$. Deseamos construir un modelo de probabilidad donde todos los resultados son igualmente probables.

Si procediéramos como se hizo en el Chapter 1 para experimentos que tienen un número finito de resultados posibles, deberíamos asignar probabilidad 0 a cada resultado, ya que de otro modo, la suma de probabilidades sobre todos los resultados posibles no sería igual 1. (De hecho, sumar un número incontable de números reales es un negocio engañoso; en particular, para que dicha suma tenga algún significado, a lo sumo un número contable de sumandos deben ser diferentes de 0.) Sin embargo, si todas las probabilidades asignadas son 0, entonces la suma es 0, no 1, como debería ser.

En la siguiente sección, explicaremos cómo contruir un modelo de probabilidad en este situación. Por ahora, asumiremos que dicho modelo puede ser construido. También asumiremos que en este modelo, si E es un arco de círculo y es de longitud p , entonces el modelo asignará la probabilidad p a E . Esto significa que si hacemos

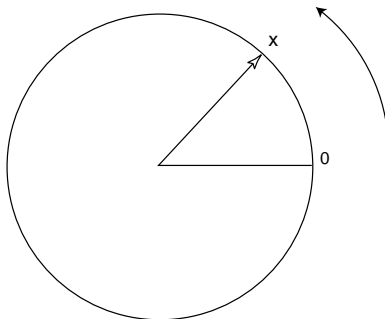


Figure 2.1: Una ruleta.

girar el puntero, la probabilidad que éste se detenga apuntando a un punto en E es igual a p , lo cual es ciertamente razonable de esperar.

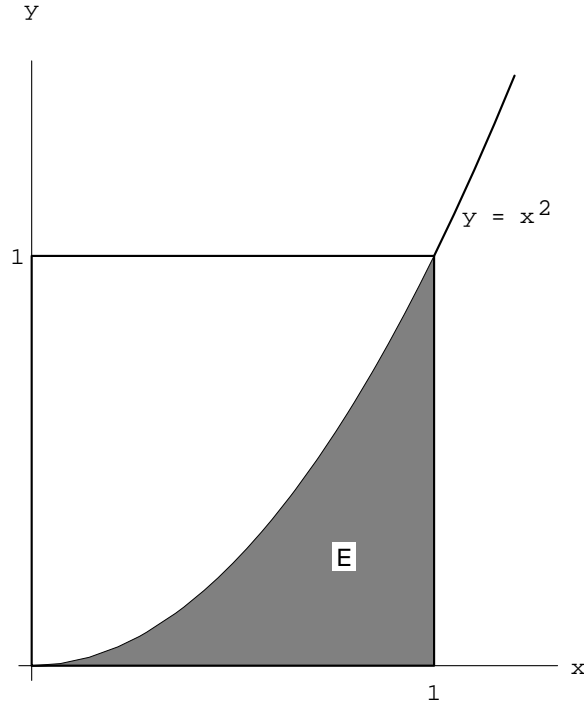
Simular este experimento en una computadora es un asunto fácil. Varios paquetes de software tienen una función que retorna un número aleatorio real del intervalo $[0, 1]$. Actualmente, los valores entregados son siempre racionales, resultados de un algoritmo, de manera que la secuencia de valores no es ciertamente aleatoria. Sin embargo, las secuencias de números producidos por dichos algoritmos tiene un comportamiento teóricamente semejante a las secuencias aleatorias, por eso utilizamos dichas secuencias en la simulación de experimentos. En ocasiones, necesitaremos referirnos a dicha función. Llamaremos a esta función *rnd*. \square

Método Monte Carlo y Áreas

Es a veces deseable estimar cantidades cuyo valor exacto es dificultoso o imposible de calcular exactamente. En algunos de estos casos, un procedimiento que involucra el azar, llamado el *Método Monte Carlo*, puede ser usado para obtener tal estimación.

Ejemplo 2.2 En este ejemplo mostraremos como la simulación puede usarse para estimar áreas de figuras planas. Suponga que programamos nuestra computadora para obtener un par de números (x, y) , cada uno escogido independientemente al azar del intervalo $[0, 1]$. Entonces podemos interpretar este par de coordenadas (x, y) como un punto escogido *al azar* del cuadrado unidad. Los sucesos son subconjuntos del cuadrado unidad. Nuestra experiencia con el Ejemplo 2.1 sugiere que dicho punto caiga en los subconjuntos de igual área es igualmente probable. Como el área total del cuadrado es 1, la probabilidad del punto de caer en un subconjunto específico E del cuadrado unidad sería igual a su área. Así, podemos estimar el área de cualquier subconjunto del cuadrado unidad estimando la probabilidad de que un punto elegido al azar de este cuadrado caiga en el subconjunto.

Podemos usar este método para estimar el área de la región E bajo la curva $y = x^2$ en el cuadrado unidad (vea Figura 2.2). Escogemos un gran número de puntos (x, y) al azar y registramos qué fracción de los mismo caen en la región $E = \{ (x, y) : y \leq x^2 \}$.

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

El programa **MonteCarlo** realizará este experimento por nosotros. Corriendo este programa para 10,000 experimentos da un estimado de .325 (vea Figura 2.3).

De estos experimentos podemos estimar que el área es cercana a $1/3$. Por supuesto, para esta región sencilla podemos encontrar el área exacta mediante el análisis matemático. De hecho,

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

Hemos resaltado en el Chapter 1 que, cuando simulamos un experimento de este tipo n veces para estimar una probabilidad, podemos esperar que la respuesta tenga error a lo sumo de $1/\sqrt{n}$ al menos 95 por ciento de las veces. Para 10,000 experimentos podemos contar con una exactitud de 0.01, y nuestra simulación alcanzó esta exactitud.

El mismo argumento funciona para cualquier región E del cuadrado unidad. Por ejemplo, suponga que E es el círculo con centro $(1/2, 1/2)$ y radio $1/2$. Entonces la probabilidad de que nuestro punto (x, y) al azar caiga dentro del círculo es igual al área del círculo, que es,

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

Si no conocemos el valor de π , podemos estimar el valor realizando este experimento una gran cantidad de veces! \square

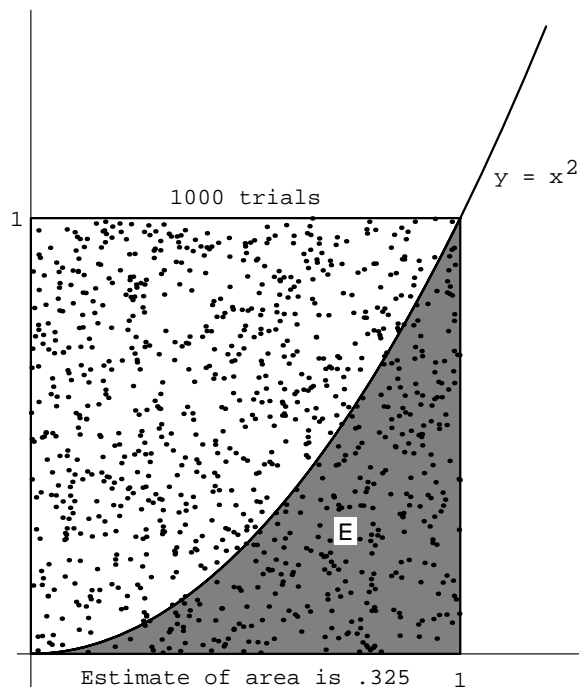


Figure 2.3: Calculando el área por simulación.

El ejemplo anterior no es la única manera de estimar el valor de π mediante un experimento de azar. Aquí está otra manera, descubierta por Buffon.¹

Las agujas de Buffon

Ejemplo 2.3 Suponga que tomamos una tabla plana y dibujamos a través de la superficie superior un conjunto de líneas paralelas a distancia unidad entre sí. Luego arrojamos una aguja común de longitud unidad al azar sobre esta superficie y observamos cuando la aguja cae o no cruzando una de las líneas. Podemos describir los posibles resultados de este experimento mediante coordenadas como sigue: Sea d la distancia desde el centro de la aguja a la línea más cercana. Luego, sea L la línea determinada por la aguja y defino θ como el ángulo agudo que forma la línea L con el conjunto de líneas paralelas. (El lector debe ciertamente ser cuidadoso de esta descripción del espacio muestral. Estamos intentando dar coordenadas a un conjunto de segmentos de línea. Para ver por qué uno debe ser cuidadoso en la elección de coordenadas, vea el Ejemplo 2.6.) Usando esta descripción, tenemos que $0 \leq d \leq 1/2$, y $0 \leq \theta \leq \pi/2$. Más aún, vemos que la aguja cae cruzando la línea más cercana si y solo si la hipotenusa del triángulo (ver Figura 2.4) es menor

¹G. L. Buffon, en “Essai d’Arithmétique Morale,” *Oeuvres Complètes de Buffon avec Suppléments*, tome iv, ed. Duménil (Paris, 1836).

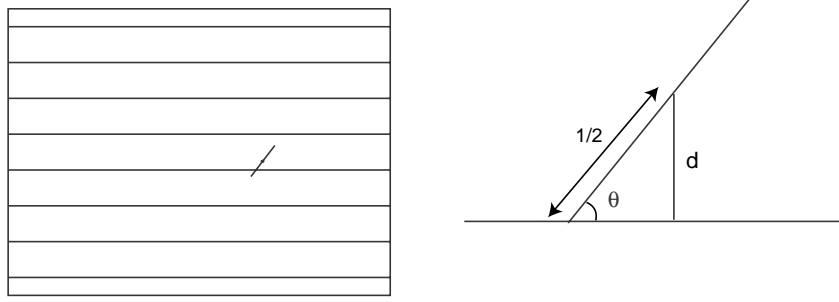


Figure 2.4: El experimento de Buffon.

que la mitad de la longitud de la aguja, esto es,

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

Ahora asumiremos que cuando la aguja cae, el par (θ, d) es escogido aleatoriamente del rectángulo $0 \leq \theta \leq \pi/2$, $0 \leq d \leq 1/2$. Observamos si la aguja cae cruzando la línea más cercana (esto es decir, si $d \leq (1/2) \sin \theta$). La probabilidad de este suceso E es la fracción del área del rectángulo que caiga dentro de E (ver Figura 2.5). Ahora el área del rectángulo es $\pi/4$, mientras que el área de E es

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

Por lo tanto, tenemos

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

El programa **BuffonsNeedle** simula este experimento. En la Figura 2.6, mostramos la posición de la 100va aguja en una ejecución del programa en el cual “arrojamos” 10,000 agujas. Nuestra estimación final de π es 3.139. Mientras que este estuvo dentro de 0.003 del valor real de π no teníamos derecho de esperar tal exactitud. La razón para esto es que nuestra simulación estima $P(E)$. Mientras que podemos esperar que esta estimación tenga a lo sumo un error de 0.001, un pequeño error en $P(E)$ resulta magnificado cuando lo usamos para calcular $\pi = 2/P(E)$. Perlman y Wichura, en su artículo “Sharpening Buffon’s Needle,”² muestran que podemos esperar tener un error de no más de $5/\sqrt{n}$ sobre el 95 por ciento de las veces. Aquí n es el número de agujas arrojadas. Así para 10,000 agujas debemos esperar un error de no más de 0.05, y este era aquí el caso. Vemos que es necesario una gran cantidad de experimentos para obtener una estimación decente para π . \square

En cada uno de nuestros ejemplos hasta ahora, sucesos de igual tamaño son igualmente probables. Aquí hay un ejemplo donde no lo son. Veremos muchos otros de estos ejemplos luego.

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

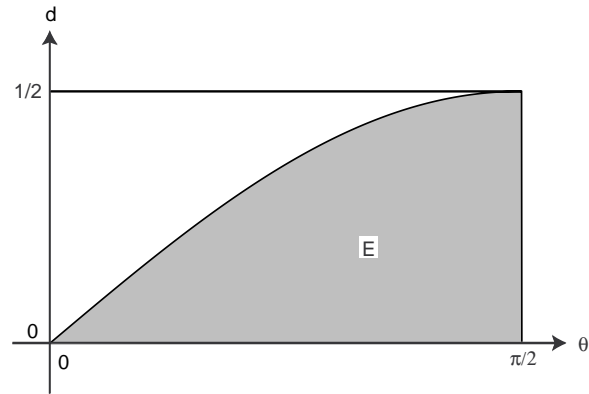


Figure 2.5: Conjunto E de pares (θ, d) con $d < \frac{1}{2} \sin \theta$.

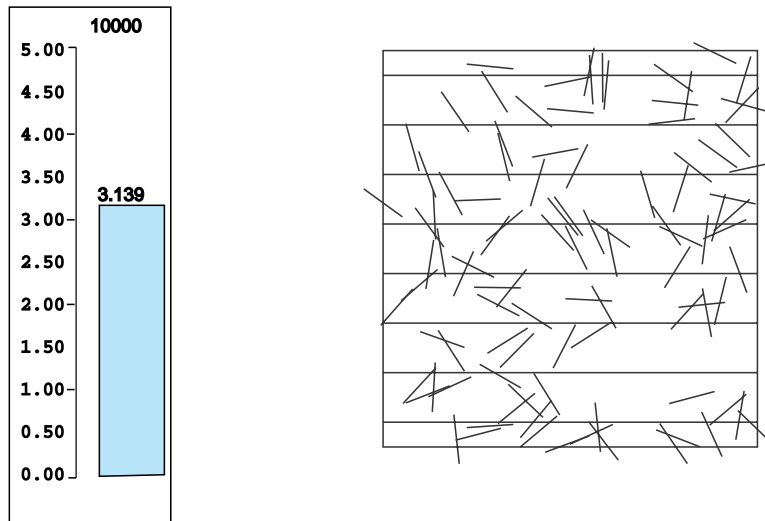


Figure 2.6: Simulación del experimento de las agujas de Buffon.

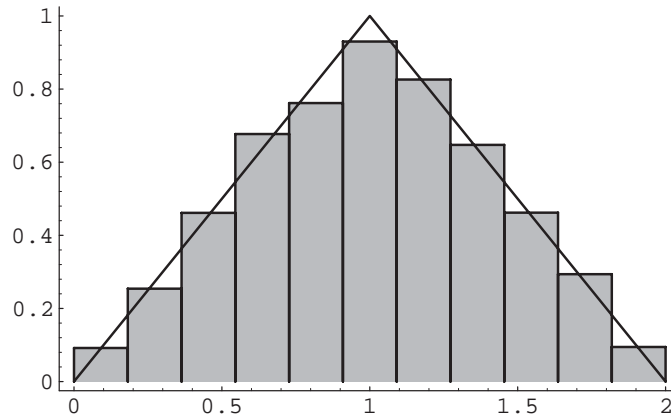


Figure 2.7: Suma de dos números aleatorios.

Ejemplo 2.4 Suponga que elegimos dos números reales aleatorios en $[0, 1]$ y los sumamos. Sea X la suma. Cómo está distribuida X ?

Para ayudar a entender la respuesta a esta pregunta, usaremos el programa **Areabargraph**. Este programa produce un gráfico de barra con la propiedad que en cada intervalo, el *área*, en vez de la altura de la barra es igual a la fracción de resultados que caen en el correspondiente intervalo. LLevamos este experimento 1000 veces y los datos son mostrados en la Figura 2.7. Parece que la función definida por

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

ajusta los datos muy bien. (Se demuestra en la figura.) En la siguiente sección, veremos que esta función es la función “correcta”. Con esto significamos que si a y b son dos números reales cualquier entre 0 y 2, con $a \leq b$, entonces podemos utilizar esta función para calcular la probabilidad que $a \leq X \leq b$. Para entender cómo este cálculo pudo ser realizado, consideremos de nuevo la Figura 2.7. Debido a la manera en que las barras fueron construidas, la suma de las áreas de las barras que corresponden al intervalo $[a, b]$ aproximan la probabilidad que $a \leq X \leq b$. Pero la suma de las áreas de estas barras también aproximan la integral

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

Esto sugiere que para un experimento con un conjunto continuo de posibles resultados, si encontramos una función con las propiedad anterior, entonces podremos utilizarla para calcular probabilidades. En la siguiente sección, mostraremos cómo determinar la función. $f(x)$. □

Ejemplo 2.5 Suponga que tomamos 100 números aleatorios en $[0, 1]$ y que X represente su suma. Cómo está distribuida X ? Hemos realizado este experimento

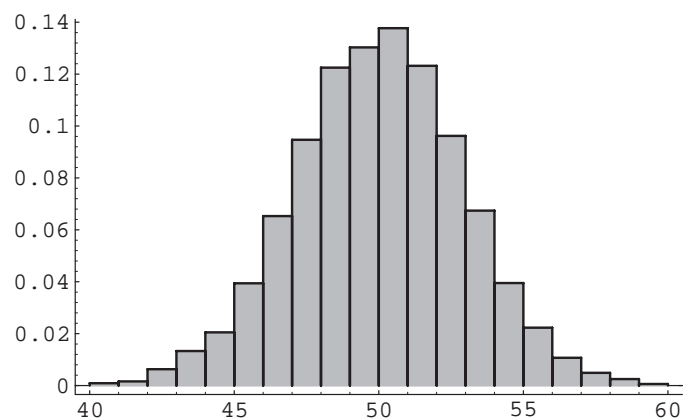


Figure 2.8: Suma de 100 números aleatorios.

10000 veces y los resultados son mostrados en la Figura 2.8. No está muy claro que función cubre las barras en este caso. Resulta que el tipo de función que corresponde se llama función de *densidad normal*. Este tipo de función es referida algunas veces como curva tipo “campana”. Está entre las funciones más importantes en el campo de la probabilidad, y será formalmente definida en la Sección 5.2 del Capítulo 4.3.

□

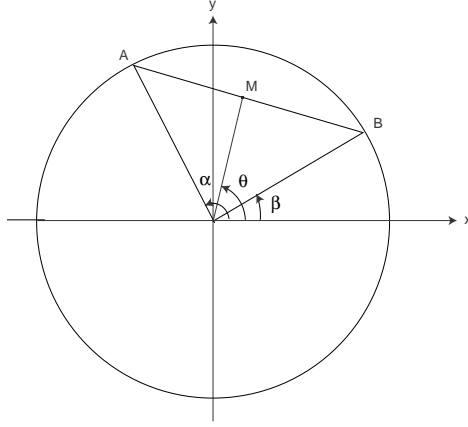


Figure 2.9: Random chord.

HASTA ACA

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

Bertrand's Paradox

Ejemplo 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 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. Jaynes⁴ 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 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.

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

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

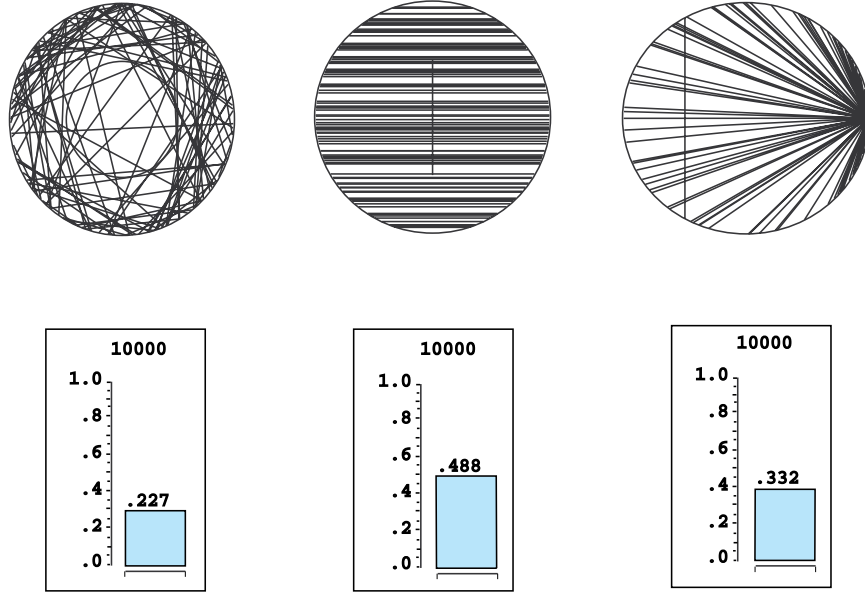


Figure 2.10: Bertrand's paradox.

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 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

⁵G. L. Buffon, *Histoire Naturelle, Generali et Particular avec le Description 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 .

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. Gridgeman⁸ 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 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

⁶G. L. Buffon, “Essai d’Arithmétique Morale,” p. 301.

⁷ibid., pp. 277–278.

⁸N. T. Gridgeman, “Geometric Probability and the Number π ” *Scripta Mathematica*, vol. 25, no. 3, (1960), pp. 183–195.

$\pi = 355/113$:

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

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.)

Ejercicios

- *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 $0 \leq d \leq 1/2$ and $0 \leq \theta \leq \pi/2$, and check whether $d \leq (1/2) \sin \theta$. Doing this a large number of times, we estimate π as $2/a$, where a is the fraction of the times that $d \leq (1/2) \sin \theta$. 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, Laplace⁹ considered a grid with *horizontal* and

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

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 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 Funciones continuas de densidad

En la sección anterior hemos visto como simular experimentos en todo el continuo de posibles resultados y hemos ganado alguna experiencia en pensar acerca de los mismos. Volvamos al problema general de asignar probabilidades a los resultados y sucesos en dichos experimentos. Restringiremos nuestra atención a aquellos experimentos cuyo espacio muestral puede tomarse como un subconjunto adecuadamente escogido de la recta, el plano, o cualquier otro espacio euclidiano. Comenzaremos con algunos ejemplos simples.

Spinners

Ejemplo 2.7 El experimento del spinner descrito en el Ejemplo 2.1 tiene como conjunto de posibles resultados el intervalo $[0, 1)$. Nos gustaría construir un modelo de probabilidad en que cada resultado puede ocurrir igualmente. Vemos que en tal modelo, es necesario asignar probabilidad 0 a cada resultado. Esto no significa que la probabilidad de *cada* suceso deba ser cero. Al contrario, si permitimos que la variable aleatoria X denote el resultado, entonces la probabilidad

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

que la cabeza del spinner descansa en *algún lugar* en el círculo debe ser igual a 1. También, la probabilidad de que descansa en la mitad superior del círculo debe ser la misma que la mitad inferior, por lo tanto

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

Más generalmente, en nuestro modelo, tomaremos como verdad la ecuación

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

para cada elección de c y d .

Sea $E = [c, d]$, entonces podemos escribir la fórmula de arriba en la forma

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

donde $f(x)$ es la función constante con valor 1. Esto debe recordar al lector a la correspondiente fórmula del caso discreto para la probabilidad de un suceso:

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

La diferencia es que en el caso del continuo, la cantidad integrada, $f(x)$, no es la probabilidad de x . (Sin embargo, si uno usa infinitesimales, se puede considerar $f(x) dx$ como la probabilidad de x .)

En el caso del continuo, usaremos la siguiente convención. Si el conjunto de resultados es un conjunto de números reales, entonces los resultados individuales

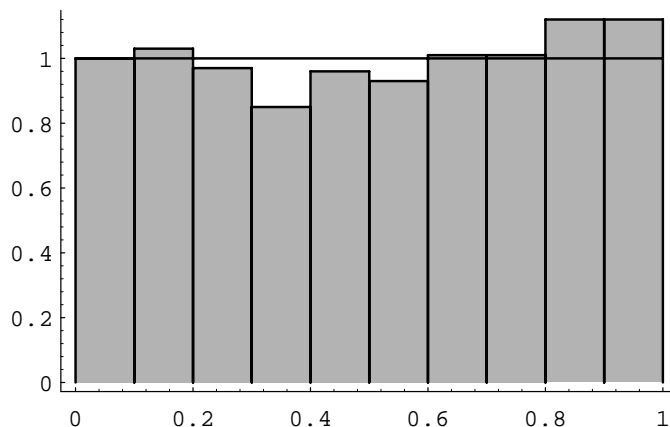


Figure 2.11: Experimento del Spinner.

serán referidos por una pequeña letra romana como x . Si el conjunto de resultados es un subconjunto de R^2 , entonces los resultados individuales serán denotados por (x, y) . En cualquier caso, puede ser más conveniente referirse a un resultado individual usando ω , como en el Capítulo 1.

La figura 2.11 muestra el resultados de 1000 giros del spinner. La función $f(x)$ es mostrada en la figura también. El lector notará que el área bajo $f(x)$ y el intervalo dado anteriormente son aproximadamente igual a la fracción de resultados que cayeron en ese intervalo. La función $f(x)$ es llamada *función densidad* de la variable aleatoria X . El hecho que el área bajo $f(x)$ y el intervalo anterior correspondan a la probabilidad es la propiedad que define a las funciones de densidad. Una definición precisa de las funciones de densidad será dada en breve. \square

Dardos

Ejemplo 2.8 Un juego de dardos consiste en tirar un dardo a un blanco circular de *radio unidad*. Suponga que tiramos el dardo una vez para que pegue en el blanco, y observamos donde aterriza. Para describir los resultados posibles del experimento, es natural tomar como nuestro espacio muestral, el conjunto Ω de todos los puntos del blanco. Es conveniente describir estos puntos por sus coordenadas rectangulares, relativo a un sistema de coordenadas con origen en el centro del blanco, para que cada par (x, y) de coordenadas, con $x^2 + y^2 \leq 1$, describa un resultado posible del experimento. Entonces $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$ es un subconjunto del plano Euclidean, y el suceso $E = \{(x, y) : y > 0\}$, por ejemplo, corresponde a la declaración que el dardo aterrice en la mitad superior del blanco, y así en adelante. A menos que haya una razón para creer otra cosa (y con los expertos del juego puede ser!), es natural asumir que las coordenadas son elegidas *al azar*. (Cuando hacemos esto con una computadora, cada coordenada es uniformemente elegida del intervalo $[-1, 1]$. Si el punto resultante no cae dentro del círculo unidad, el punto no es contado.) Entonces los argumentos usados en el ejemplo precedente muestran que

la probabilidad de cualquier suceso elemental, consistiendo de un solo resultado, debe ser cero, y sugiere que la probabilidad del suceso, que el dardo aterrice en cualquier subconjunto E del blanco, debe ser determinada por la fracción del área del blanco que cae en E . Así,

$$P(E) = \frac{\text{área de } E}{\text{área del blanco}} = \frac{\text{área de } E}{\pi} .$$

Esto puede ser escrito en la forma

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

donde $f(x)$ es la función constante con valor $1/\pi$. En particular, si $E = \{ (x, y) : x^2 + y^2 \leq a^2 \}$ es el suceso que el dardo aterrice dentro de la distancia $a < 1$ del centro del blanco, entonces

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

Por ejemplo, la probabilidad que el dardo caiga dentro de la distancia $1/2$ del centro es $1/4$. \square

Ejemplo 2.9 En el juego de dardos considerado arriba, supongamos que, en vez de observar donde aterrizó el dardo, observamos cuán lejos aterrizó desde el centro del blanco.

En este caso, tomamos como nuestro espacio muestral el conjunto Ω de todos los círculos centrados en el centro del blanco. Es conveniente describir a estos por sus radios, para que cada círculo se identifique por su radio r , $0 \leq r \leq 1$. De esta manera, podemos considerar a Ω como el subconjunto $[0, 1]$ de la recta real.

Que probabilidades debemos asignar a los sucesos E de Ω ? Si

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

entonces ocurre E si el dardo aterrizó dentro de la distancia a desde el centro, es decir, dentro del círculo de radio a , y vimos en el ejemplo anterior que, bajo nuestras suposiciones, la probabilidad de este suceso está dada por

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

Más generalmente, si

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

entonces por nuestra suposición básica,

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

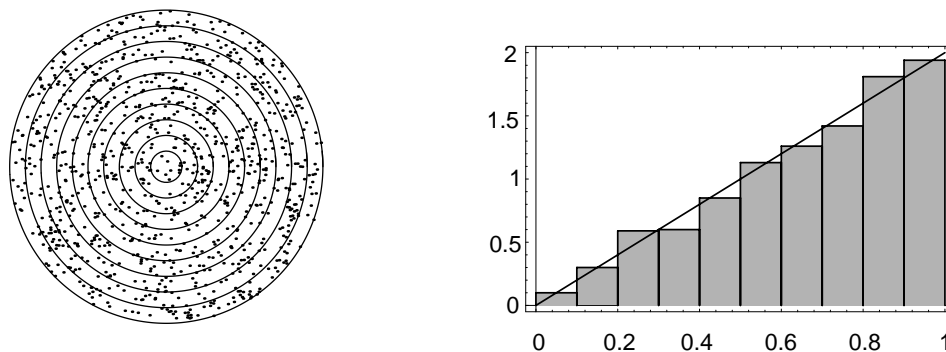


Figure 2.12: Distribución de las distancias de los dardos en 400 tiros.

Así, $P(E) = 2(\text{longitud de } E)(\text{punto medio de } E)$. Aquí vemos que la probabilidad asignada al intervalo E depende no solo de su longitud sino también de su punto medio (es decir, no solo cuán largo es, sino también en donde está). Aproximadamente hablando, en este experimento, sucesos de la forma $E = [a, b]$ son más probables si están cercanos al borde del blanco y menos probable si están cerca del centro. (Un experiencia común para los principiantes! La conclusión podría ser bien diferente si el principiante es reemplazado por un experto.)

Nuevamente podemos simularlo por computadora. Dividimos el area del blanco en diez regiones concéntricas de igual espesor.

El programa de computadora **Darts** lanza n dardos y registra que fracción del total cae en cada una de estas regiones concéntricas. El programa **Areabargraph** entonces dibuja un gráfico de barras donde el *área* de la i -ésima barra es igual a la fracción del total caído en la i -ésima región. Ejecutando el programa para 1000 dardos produce el gráfico de barra de la Figura 2.12.

Notar aquí que las alturas de las barras no son todas iguales, sino que crecen aproximadamente lineales con r . De hecho, la función lineal $y = 2r$ parece encajar bastante bien en nuestro gráfico de barra. Esto sugiere que la probabilidad de que el dardo caiga dentro de la distancia a desde el centro estará dada por el *área* debajo de la gráfica de la función $y = 2r$ entre 0 y a . Esta área es a^2 , que concuerda con la probabilidad que asignamos anteriormente a este suceso. \square

Espacio Muestrales de Coordenadas

Estos ejemplos sugieren que para los experimentos continuos de esta clase debemos asignar las probabilidades para los resultados que caen dentro de un intervalo dado por medio del área bajo una función conveniente.

Más generalmente, suponemos que coordenadas convenientes pueden introducirse en el espacio muestral Ω , para que podamos considerar a Ω como un subconjunto de \mathbf{R}^n . Llamamos a tal espacio muestral *espacio muestral continuo*. Sea X la variable aleatoria que representa el resultado del experimento. Tal variable aleatoria es llamada *variable aleatoria continua*. Entonces podemos definir la función densidad

como sigue.

Funciones Densidad de Variables Aleatorias Continuas

Definición 2.1 Sea X una variable aleatoria continua a valor real. Una *función densidad* para X es una función a valor real f que satisface

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

para todo $a, b \in \mathbf{R}$. □

Notamos que *no* es el caso que todas las variables aleatorias a valor real poseen función densidad. Sin embargo, en este libro, consideraremos solo variables aleatorias continuas cuya función densidad existe.

En términos de la densidad $f(x)$, si E es un subconjunto de \mathbf{R} , entonces

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

Aquí la notación asume que E es un subconjunto de \mathbf{R} por lo cual $\int_E f(x) dx$ tiene sentido.

Ejemplo 2.10 (Ejemplo 2.7 continuado) En el experimento spinner, escogimos para nuestro conjunto de resultados el intervalo $0 \leq x < 1$, y para nuestra función densidad

$$f(x) = \begin{cases} 1, & \text{si } 0 \leq x < 1, \\ 0, & \text{otro caso.} \end{cases}$$

Si E es el suceso que la cabeza del spinner caiga en la mitad superior del círculo, entonces $E = \{x : 0 \leq x \leq 1/2\}$, y por lo tanto

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

Más generalmente, si E es el suceso que la cabeza caiga en el intervalo $[a, b]$, entonces

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

□

Ejemplo 2.11 (Ejemplo 2.8 continuado) En el primer experimento con juego de dardos, escogimos como nuestro espacio muestral el disco de radio unidad en el plano y para nuestra función densidad la función

$$f(x, y) = \begin{cases} 1/\pi, & \text{si } x^2 + y^2 \leq 1, \\ 0, & \text{otro caso.} \end{cases}$$

La probabilidad que el dardo aterrice dentro del subconjunto E está dada por

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

□

En estos dos ejemplos, la función densidad es constante y no depende del resultado particular. Es a menudo el caso, de experimentos en que las coordenadas son escogidas *al azar*, pueden ser descriptos por funciones de densidad *constantes*, y, como en la Sección 1.2, llamamos a tales funciones de densidad *uniformes* o *equiprobables*. Sin embargo, no todos los experimentos son de este tipo.

Ejemplo 2.12 (Ejemplo 2.9 continuado) En el segundo experimento del juego de dardos, escogimos para nuestro espacio muestral el intervalo unidad sobre la recta real y para nuestra función densidad

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

Entonces la probabilidad que el dardo aterrice a distancia r , $a \leq r \leq b$, del centro del blanco está dada por

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

Aquí nuevamente, dado que la densidad es pequeña cuando r es cercana a 0 y grande cuando r es cercana a 1, vemos que en este experimento que es más probable que el dardo aterrice cerca del borde del blanco que cerca del centro. En términos del gráfico de barras del Ejemplo 2.9, las alturas de las barras aproximan la función de densidad, mientras que el área de las barras aproximan las probabilidades de los subintervalos (vea Figura 2.12). □

Vemos en este ejemplo que, a diferencia del caso de espacio muestrales discretos, el valor $f(x)$ de la función densidad para el resultado x *no* es la probabilidad de ocurrir x (hemos visto que dicha probabilidad es siempre 0) y en general $f(x)$ *no es en absoluto una probabilidad*. En este ejemplo, si tomamos $\lambda = 2$ entonces $f(3/4) = 3/2$, siendo más grande que 1, no puede ser una probabilidad.

No obstante, la función densidad f contiene toda la información de probabilidad acerca del experimento, dado que las probabilidades de todos los sucesos pueden ser derivados de esta. En particular, la probabilidad que el resultado de un experimento caiga en un intervalo $[a, b]$ está dado por

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

es decir, por el *área* debajo del gráfico de la función de densidad en el intervalo $[a, b]$. Así, hay una conexión íntima aquí entre probabilidades y áreas. Nos hemos guiado por esta conexión íntima al construir nuestros gráficos de barra; cada barra es escogida para que su *área*, y no su altura, represente la frecuencia relativa de ocurrencia, y de aquí calcular la probabilidad del resultado que cae en el intervalo asociado.

En el lenguaje del análisis diferencial, decimos que la probabilidad de ocurrir del suceso de la forma $[x, x + dx]$, donde dx es pequeño, es aproximadamente dado por

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

es decir, por el área del rectángulo debajo de la gráfica de f . Notar que como $dx \rightarrow 0$, esta probabilidad $\rightarrow 0$, por lo que la probabilidad $P(\{x\})$ de un solo punto es de nuevo 0, como en el Ejemplo 2.7.

Una mirada al gráfico de la función densidad nos dice inmediatamente que sucesos de un experimento son mas probables. Hablando aproximadamente, podemos decir que donde la densidad es grande los sucesos son más probables, y donde es pequeña los sucesos son menos probables. En el Ejemplo 2.4 la función densidad es más grande cerca de 1. Así, dado los dos intervalos $[0, a]$ and $[1, 1 + a]$, donde a es un número real positivo pequeño, vemos que X es m'as probable que tome valores en el segundo intervalo que en el primero.

Funciones de Distribución Acumulada de Variables Aleatorias Continuas

Hemos visto que las funciones de densidad son útiles al considerar variables aleatorias continuas. Hay otro tipo de función, estrechamente ligado a estas funciones de densidad, que también es de gran importancia. Estas funciones son llamadas funciones de distribución acumulada.

Definición 2.2 Sea X una variable aleatoria continua a valor real. Entonces la función de distribución acumulada de X es definida por la ecuación

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

□

Si X es una variable aleatoria continua a valor real que posee una función de densidad, entonces también tiene una función de distribución acumulada, y el siguiente teorema muestra que las dos funciones están relacionadas de una manera cercana.

Teorema 2.1 Sea X una variable aleatoria continua a valor real con una función de densidad $f(x)$. Entonces la función definida por

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

es la función de distribución acumulada de X . Además, tenemos

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

Dem. Por definición,

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

Sea $E = (-\infty, x]$. Entonces

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

que es igual a

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

Aplicando el Teorema Fundamental del Cálculo a la primera ecuación en la declaración del teorema sigue la segunda. \square

En muchos experimentos, la función de densidad de la variable aleatoria pertinente es fácil de encontrar. Sin embargo, es muy común el caso en que la función de distribución acumulada es más fácil de obtener que la función de densidad. (Por supuesto, una vez que tenemos la función de distribución acumulada, puede fácilmente obtenerse la función de densidad por derivación, como muestra el teorema anterior.) Daremos algunos ejemplos que exhiben este fenómeno ahora.

Ejemplo 2.13 Un número real es escogido al azar de $[0, 1]$ con probabilidad uniforme, y luego es elevado al cuadrado. Sea X el resultado. Cuál es la función de distribución acumulada de X ? Cuál es la densidad de X ?

Comenzemos permitiendo que U represente el número real escogido. Entonces $X = U^2$. Si $0 \leq x \leq 1$, entonces tenemos

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

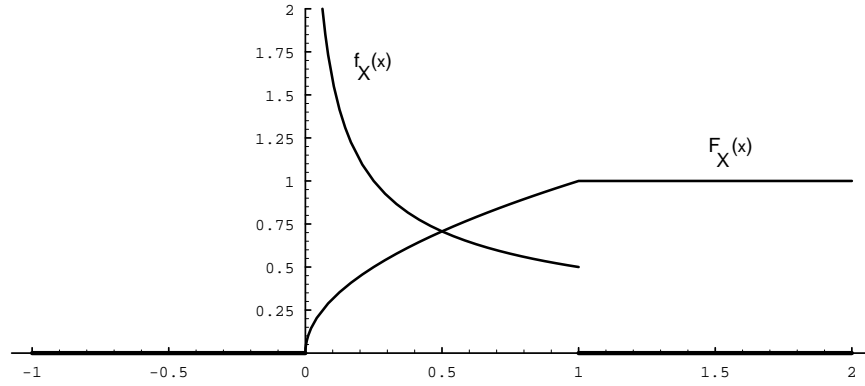
Es claro que X siempre toma un valor entre 0 y 1, por lo tanto la función de distribución acumulada de X está dada por

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

Desde aquí podemos calcular que la función de densidad de X es

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

Notar que $F_X(x)$ es continua pero $f_X(x)$ no. (Mirar la Figura 2.13.) \square

Figure 2.13: Distribución y densidad para $X = U^2$.

Cuando nos referimos a una variable aleatoria continua X (digamos, con una función densidad uniforme), es costumbre decir que “ X está uniformemente *distribuida* en el intervalo $[a, b]$.” Es también costumbre referirse a la función de distribución acumulada de X como la función de distribución de X . Así que la palabra “distribución” está siendo usada de varias maneras directas en el tema de probabilidad. (Recordar que también tiene significado al discutir variables aleatorias discretas.) Al referirse a la función de distribución acumulada de una variable aleatoria continua X , dejamos claro, que siempre usaremos la palabra “acumulada” como modificador, al menos que usemos otro modificador, como “normal” o “exponencial”. Dado que la frase “uniformemente densificada en el intervalo $[a, b]$ ” no es aceptable en Castellano, siempre diremos en cambio “uniformemente distribuida”.

Ejemplo 2.14 En el Ejemplo 2.4, consideramos una variable aleatoria, definida como la suma de dos números aleatorios reales escogidos uniformemente de $[0, 1]$. Sean las variables aleatorias X e Y los dos números reales escogidos. Definamos $Z = X + Y$. Ahora, derivaremos las expresiones de la función de distribución acumulada y la función de densidad de Z .

Aquí tomamos para nuestro espacio muestral Ω el cuadrado unidad en \mathbf{R}^2 con densidad uniforme. Entonces un punto $\omega \in \Omega$ consiste en un par (x, y) de números elegidos al azar. Entonces $0 \leq Z \leq 2$. Sea E_z el suceso que $Z \leq z$. En la Figura 2.14, mostramos el conjunto $E_{.8}$. El suceso E_z , para cualquier z entre 0 y 1, se ve muy similar al conjunto sombreado en la figura. Para $1 < z \leq 2$, el conjunto E_z se ve como el cuadrado unidad con un triángulo removido de la esquina superior derecha. Ahora podemos calcular la distribución de probabilidad F_Z de Z ; esta está dada por

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

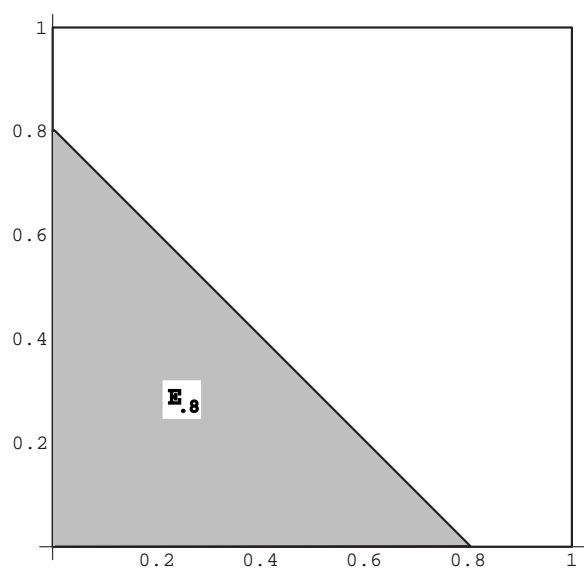


Figure 2.14: Cálculo de la función de distribución para el Ejemplo 2.14.

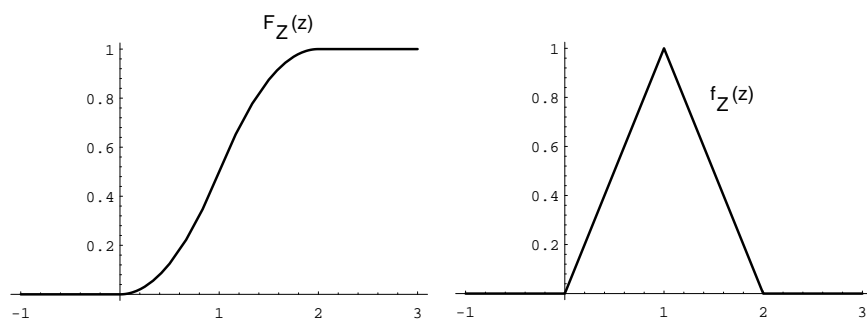
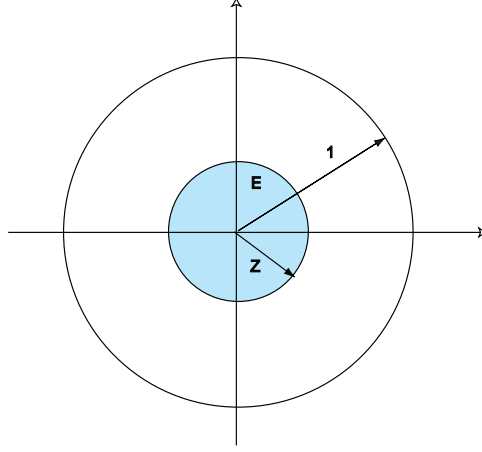


Figure 2.15: Funciones de distribución y densidad para el Ejemplo 2.14.

Figure 2.16: Cálculo de F_z para el Ejemplo 2.15.

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

La función de densidad se obtiene derivando la función distribución:

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

El lector es remitido a la Figura 2.15 para los gráficos de estas funciones. \square

Ejemplo 2.15 En el juego de dardos descrito en el Ejemplo 2.8, cuál es la distribución de la distancia del dardo al centro del blanco? Cuál es su densidad?

Aquí, como antes, nuestro espacio muestral Ω es el disco unidad en \mathbf{R}^2 , con coordenadas (X, Y) . Sea $Z = \sqrt{X^2 + Y^2}$ que representa la distancia desde el centro del blanco. Sea E el suceso $\{Z \leq z\}$. Entonces la función distribución F_Z de Z (vea Figura 2.16) está dada por

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

Por lo tanto, facilmente calculamos que

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

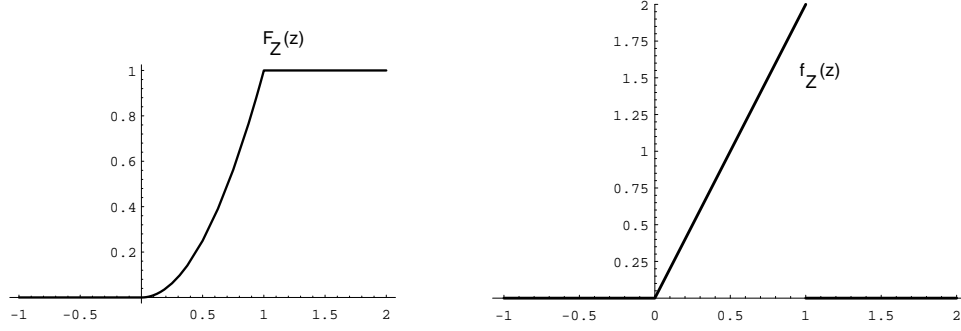
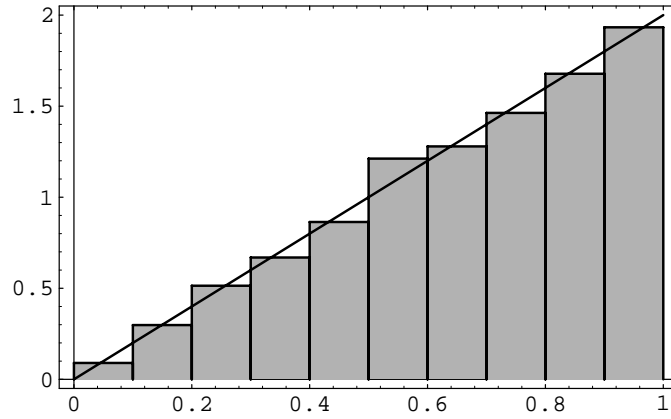
Figure 2.17: Distribución y densidad para $Z = \sqrt{X^2 + Y^2}$.

Figure 2.18: Resultados de la simulación para el Ejemplo 2.15.

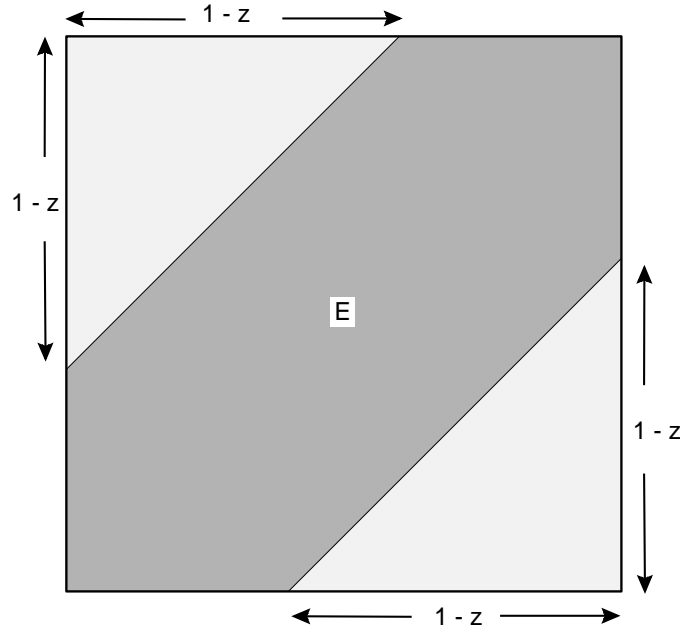
La densidad $f_Z(z)$ está nuevamente dada por la derivada de $F_Z(z)$:

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

El lector es remitido a la Figura 2.17 para el gráfico de estas funciones.

Podemos verificar este resultados por simulación, como sigue: Escogemos valores para X e Y al azar del intervalo $[0, 1]$ con distribución uniforme, calculamos $Z = \sqrt{X^2 + Y^2}$, chequeamos si $0 \leq Z \leq 1$, y presentamos los resultados en un gráfico de barras (vea Figura 2.18). \square

Ejemplo 2.16 Supongamos que el Sr. y la Sra. Lockhorn acuerdan encontrarse en el Hanover Inn entre las 5:00 y 6:00 P.M. del Jueves. Supongamos que cada uno llega a una hora entre 5:00 y 6:00 escogida al azar con probabilidad uniforme. Cuál es la función distribución de longitud de tiempo que el primero en llegar tiene que esperar al otro? Cuál es la función densidad?

Figure 2.19: Cálculo de F_Z .

Aquí nuevamente podemos tomar el cuadrado unidad para representar el espacio muestral, y (X, Y) como los tiempos de llegada (después de las 5:00 P.M.) para los Lockhorns. Sea $Z = |X - Y|$. Entonces tenemos $F_X(x) = x$ and $F_Y(y) = y$. Más aún (vea Figura 2.19),

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

Así, tenemos

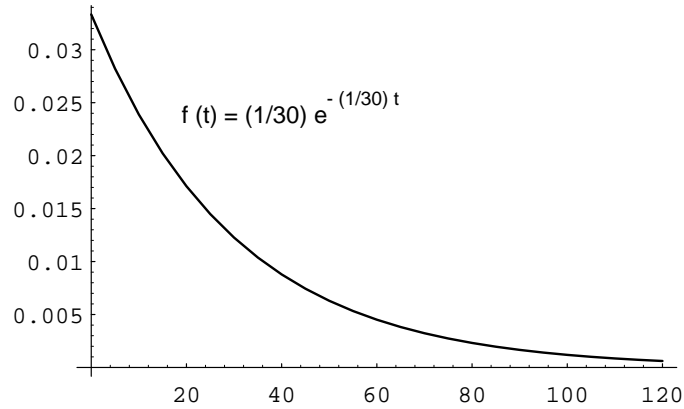
$$F_Z(z) = \begin{cases} 0, & \text{si } z \leq 0, \\ 1 - (1 - z)^2, & \text{si } 0 \leq z \leq 1, \\ 1, & \text{si } z > 1. \end{cases}$$

La densidad $f_Z(z)$ es obtenida nuevamente por diferenciación:

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

□

Ejemplo 2.17 Hay muchas ocasiones dónde observamos una sucesión de incidentes que ocurren en tiempos “aleatorios”. Por ejemplo, podríamos estar observando emisiones de un isótopo radiactivo, o automóviles que pasan postes en una carretera,

Figure 2.20: Densidad exponencial con $\lambda = 1/30$.

o bombillas encendiéndose. En estos casos, podríamos definir una variable aleatoria X para denotar el tiempo entre las ocurrencias sucesivas. Claramente, X es una variable aleatoria continua cuyo rango consiste en números reales no negativos. Es a menudo el caso, que podemos modelar X usando la *densidad exponencial*. Esta densidad está dada por la fórmula

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

El número λ es un número real no negativo, y representa el recíproco del valor medio de X . (Esto se mostrará en el Capítulo 6.) Así, si el tiempo medio entre las ocurrencias es de 30 minutos, entonces, $\lambda = 1/30$. Un gráfico de esta función densidad con $\lambda = 1/30$ es mostrado en la Figura 2.20. Uno puede ver en la figura que aunque el valor medio es 30, de vez en cuando X toma valores mucho más grandes.

Suponga que hemos comprado una computadora que contiene un disco rígido Warp 9. El vendedor dice que el tiempo medio entre las averías de este tipo de disco rígido es 30 meses. Es supuesto a menudo que la longitud de tiempo entre las averías está distribuido según la densidad exponencial. Asumiremos que este modelo se aplica aquí, con $\lambda = 1/30$.

Ahora supone que hemos estado usando nuestra computadora durante 15 meses. Asumimos que el disco rígido original todavía está andando. Preguntamos cuánto tiempo debemos esperar que el disco rígido siga funcionando. Uno podría esperar razonablemente que el disco rígido corriera, en promedio, otros 15 meses. (Uno también podría suponer que andará más de 15 meses, dado el hecho que ya ha funcionado durante 15 meses, implica que no es un fiasco.) El tiempo que tenemos que esperar es una nueva variable aleatoria que llamaremos Y . Obviamente, $Y = X - 15$. Podemos escribir un programa de computadora para producir una sucesión de valores Y simulados. Para hacer esto, producimos una sucesión de X primero, y desecharnos aquellos valores que son menores o igual a 15 (estos valores corresponden a los casos dónde el disco rígido ha dejado de correr antes de 15 meses). Para simular

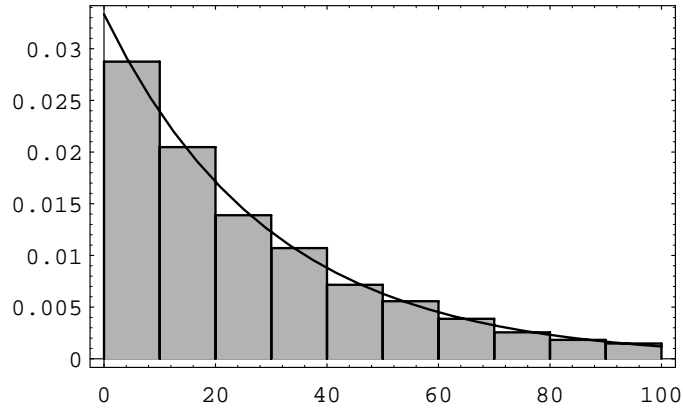


Figure 2.21: Tiempo de vida residual de un disco rígido.

un valor de X , calculamos el valor de la expresión

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

donde rnd representa un número real entre 0 y 1. (Que esta expresión tiene densidad exponencial se mostrará en el Capítulo 4.3.) La Figura 2.21 muestra el área del gráfico de barra para 10,000 valores simulados de Y .

El valor medio de Y en esta simulación es 29.74, que es más cercano al tiempo de vida medio original de 30 meses que al valor de 15 meses que se supuso anteriormente. La distribución de Y se ve parecida a la distribución de X , también. De hecho el caso es que X e Y tienen la misma distribución. Esta propiedad se llama la *propiedad sin memoria*, porque la cantidad de tiempo que tenemos que esperar por una ocurrencia no depende de cuánto ya hemos esperado. La única función de densidad continua con esta propiedad es la densidad exponencial. \square

Asignación de Probabilidades

Una pregunta fundamental en la práctica es: Cómo escogeremos la función de densidad de probabilidad para describir cualquier experimento dado? La respuesta depende en gran parte de la cantidad y tipo de información que disponemos sobre el experimento. En algunos casos, podemos ver que los resultados son igualmente probables. En algunos casos, podemos ver que el experimento ya se parece a otro descrito por una densidad conocida. En algunos casos, podemos ejecutar el experimento un número grande de veces y hacer una suposición razonable de la densidad en base a la distribución observada de resultados, como hicimos en el Capítulo 1. En general, el problema de escoger la función de densidad correcta para un experimento dado es un problema central para el experimentador y no siempre es fácil resolver (vea el Ejemplo 2.6). Nosotros no examinaremos esta pregunta en detalle aquí pero en cambio asumiremos que la densidad correcta ya es conocida para cada uno de los experimentos bajo estudio.

La introducción de coordenadas convenientes para describir un espacio muestral continuo, y una densidad conveniente para describir sus probabilidades, no siempre es tan obvia, como muestra nuestro último ejemplo.

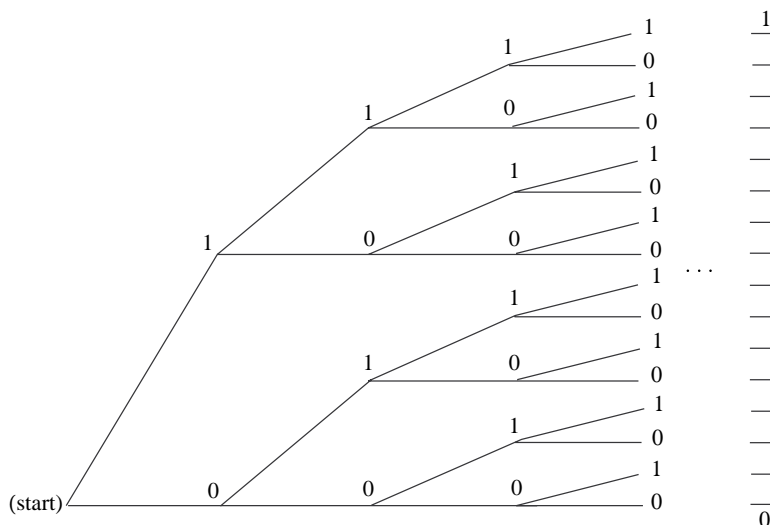


Figure 2.22: Tree for infinite number of tosses of a coin.

Infinite Tree

Ejemplo 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 $\omega = \{H H T H T T H \dots\}$. If we write 1 for H and 0 for T, then $\omega = \{1 1 0 1 0 0 1 \dots\}$. 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 $x = .1101001\dots$ 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 $\{T H H H H H \dots\}$ and $\{H T T T T T \dots\}$ 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 $.10****\dots$, 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 $1/2 = .10000\dots$ and $3/4 = .11000\dots$,

and moreover, every value of x between $1/2$ and $3/4$ has a binary expansion of the form $.10****\dots$. 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 $011011011\dots$ (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

Ejercicios

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?