5. Random Variables

5.1 Probability Functions and Distribution Functions

In many applications, the outcomes of a probabilistic experiment are numbers or have some numbers associated with them, and we can use these numbers to obtain important information beyond what we have seen so far. We can, for instance, describe in various ways how large or small these numbers are likely to be and compute likely averages and measures of spread. For example, in three tosses of a coin, the number of heads obtained can range from 0 to 3, and there is one of these numbers associated with each possible outcome. Informally, the quantity "number of heads" is called a random variable and the numbers 0 to 3 its possible values. In general, such an association of numbers with each member of a set is called a function. For most functions whose domain is a sample space, we have a new name:

Definition 5.1.1. ***Random Variable.*** *A random variable (abbreviated r.v.) is a real-valued function on a sample space.*

Random variables are usually denoted by capital letters from the end of the alphabet, such as X, Y, Z, and sets like $\{s : X(s) = x\}$, $\{s : X(s) \leq x\}$, and $\{s : X(s) \in I\}$, for any number x and any interval I, are events[1] in S. They are usually abbreviated as $\{X = x\}$, $\{X \leq x\}$, and $\{X \in I\}$ and have probabilities associated with them. The assignment of probabilities to all such events, for a given random variable X, is called the *probability distribution of* X. Furthermore, in the notation for such probabilities, it is customary to drop the braces, that is, to write $\mathrm{P}(X = x)$, for instance, rather than $\mathrm{P}(\{X = x\})$.

Hence, the preceding example can be formalized thus:

[1] Actually, in infinite sample spaces, there exist complicated functions for which not all such sets are events, and so we define a r.v. as not just any real-valued function X, but a so-called measurable function, that is, one for which all such sets *are* events. We shall ignore this issue; it is explored in more advanced books.

G. Schay, *Introduction to Probability with Statistical Applications*,
DOI 10.1007/978-3-319-30620-9_5

Example 5.1.1. Three Tosses of a Coin.

Let $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ describe three tosses of a coin, and let X denote the number of heads obtained. Then the values of X, for each outcome s in S, are given in the following table:

s :	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(s)$:	3	2	2	1	2	1	1	0

Thus, in the case of three independent tosses of a fair coin, $\mathrm{P}(X = 0) = 1/8$, $\mathrm{P}(X = 1) = 3/8$, $\mathrm{P}(X = 2) = 3/8$, and $\mathrm{P}(X = 3) = 1/8$. ♦

The following functions are generally used to describe the probability distribution of a random variable:

Definition 5.1.2. ***Probability Function.*** *For any probability space and any random variable X on it, the function $f(x) = P(X = x)$, defined for all possible values*[2] *x of X, is called the probability function (abbreviated p.f.) of X.*

Definition 5.1.3. ***Distribution Function.*** *For any probability space and any random variable X on it, the function $F(x) = P(X \leq x)$, defined for all real numbers x, is called the distribution function (abbreviated d.f.) of X.*

Example 5.1.2. Three Tosses of a Coin, Continued.

Let X be the number of heads obtained in three independent tosses of a fair coin, as in the previous example. Then the p.f. of X is given by

$$f(x) = \begin{cases} 1/8 \text{ if } & x = 0 \\ 3/8 \text{ if } & x = 1 \\ 3/8 \text{ if } & x = 2 \\ 1/8 \text{ if } & x = 3 \end{cases} \tag{5.1}$$

and the d.f. of X is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/8 & \text{if } 0 \leq x < 1 \\ 4/8 & \text{if } 1 \leq x < 2 \\ 7/8 & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases} \tag{5.2}$$

The graphs of these functions are shown in Figures 5.1 and 5.2 below.

It is also customary to picture the probability function by a histogram, which is a bar chart with the probabilities represented by areas. For the X above, this is shown in Figure 5.3. (In this case, the bars all have width one, and so their heights and areas are equal.) ♦

[2] Sometimes $f(x)$ is considered to be a function on all of $\mathbb{R}$, with $f(x) = 0$ if x is not a possible value of X. This is a minor distinction, and it should be clear from the context which definition is meant.

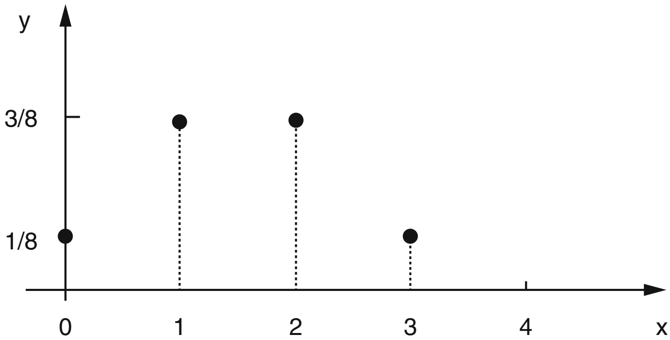

Fig. 5.1. Graph of the p.f. f of a binomial random variable with parameters $n = 3$ and $p = 1/2$

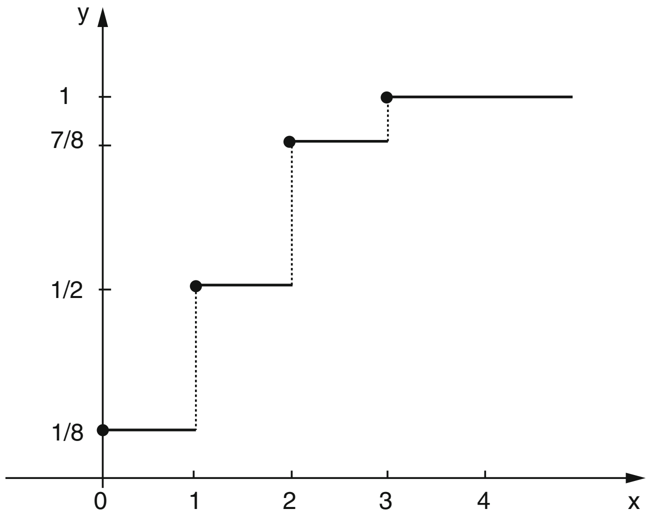

Fig. 5.2. Graph of the d.f. F of a binomial random variable with parameters $n = 3$ and $p = 1/2$

Certain frequently occurring random variables and their distributions have special names. Two of these are generalizations of the number of heads in the above example. The first one is for a single toss, but with a not necessarily fair coin, and the second one for an arbitrary number of tosses.

Definition 5.1.4. ***Bernoulli Random Variables.*** *A random variable X is called a Bernoulli random variable with parameter p, if it has two possible values, 0 and 1, with $P(X = 1) = p$ and $P(X = 0) = 1 - p = q$, where p is any number from the interval $[0,1]$. An experiment whose outcome is a Bernoulli random variable is called a Bernoulli trial.*

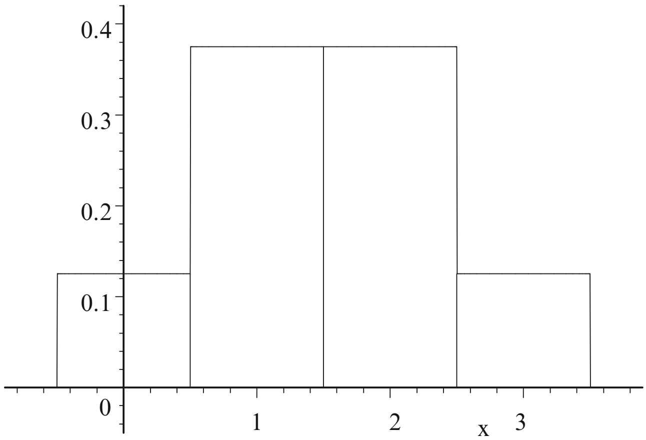

Fig. 5.3. Histogram of the p.f. f of a binomial random variable with parameters $n = 3$ and $p = 1/2$

Definition 5.1.5. ***Binomial Random Variables.*** *A random variable X is called a binomial random variable with parameters n and p, if it has the binomial distribution (see Example 4.3.4) with probability function*

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad \text{if} \quad x = 0,\ 1,\ 2, \ldots,\ n. \tag{5.3}$$

The distribution function of a binomial random variable is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k q^{n-k} & \text{if } 0 \le x < n \\ 1 & \text{if } x \ge n. \end{cases} \tag{5.4}$$

Here $\lfloor x \rfloor$ denotes the floor or greatest integer function, that is, $\lfloor x \rfloor$ = the greatest integer $\le x$.

Example 5.1.3. Sum of Two Dice.

Let us consider again the tossing of two dice, with 36 equiprobable elementary events, and let X be the sum of the points obtained. Then $f(x)$ and $F(x)$ are given by the following tables. (Count the appropriate squares in Figure 2.4 on p. 21.)

x :	2	3	4	5	6	7	8	9	10	11	12
$f(x)$:	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

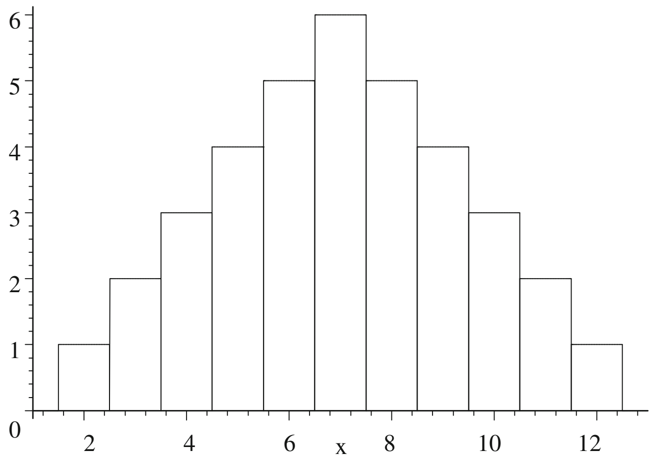

Fig. 5.4. Histogram of the d.f. of the sum thrown with two dice. The y-scale shows multiples of $1/36$

$x \in$	$(-\infty, 2)$	$[2, 3)$	$[3, 4)$	$[4, 5)$	$[5, 6)$	$[6, 7)$
$F(x)$:	0	$1/36$	$3/36$	$6/36$	$10/36$	$15/36$

$[7, 8)$	$[8, 9)$	$[9, 10)$	$[10, 11)$	$[11, 12)$	$[12, \infty)$
$21/36$	$26/36$	$30/36$	$33/36$	$35/36$	1

The histogram of $f(x)$ and the graph of $F(x)$ are given by Figures 5.4 and 5.5. ♦

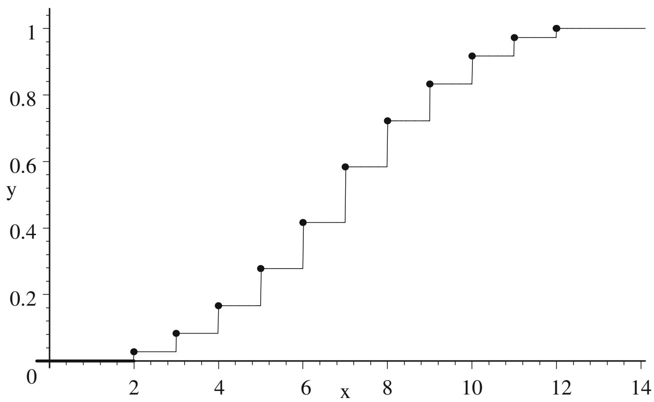

Fig. 5.5. Graph of the d.f. of the sum thrown with two dice

A random variable is said to be *discrete* if it has only a finite or a countably infinite number of possible values. The random variables we have seen so far are discrete. In the next section, we shall discuss the most important class of non-discrete random variables: continuous ones.

Another important type of discrete variable is named in the following definition:

Definition 5.1.6. ***Discrete Uniform Random Variables.*** *A random variable X and its distribution are called discrete uniform if X has a finite number of possible values, say $x_1, x_2, \ldots, x_n$, for any positive integer n, and $P(X = x_i) = \frac{1}{n}$ for all $i = 1, 2, \ldots, n$.*

Random variables with a countably infinite number of possible values occur in many applications, as in the next example.

Example 5.1.4. Throwing a Die Until a Six Comes Up.

Suppose we throw a fair die repeatedly, with the throws being independent of each other, until a six comes up. Let X be the number of throws. Clearly, X can take on any positive integer value, for it is possible (though unlikely) that we do not get a six in 100 throws, or 1000 throws, or in any large number of throws.

The probability function of X can be computed easily as follows:

$f(1) = \mathrm{P}(X = 1) = \mathrm{P}(\text{six on the first throw}) = \frac{1}{6}$,

$f(2) = \mathrm{P}(X = 2) = \mathrm{P}(\text{non} - \text{six on the first throw and six on the second}) = \frac{5}{6} \cdot \frac{1}{6}$,

$f(3) = \mathrm{P}(X = 3) = \mathrm{P}$ (non-six on the first two throws and six on the third) $= \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6}$, and so on.

Thus

$$f(k) = \mathrm{P}(X = k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} \quad \text{for} \quad k = 1, 2, \ldots \ . \tag{5.5}$$

♦

The above example is a special case of another named family of random variables:

Definition 5.1.7. ***Geometric Random Variables.*** *Suppose we perform independent Bernoulli trials with parameter p, with $0 < p < 1$, until we obtain a success. The number X of trials is called a geometric random variable with parameter p. It has the probability function*

$$f(k) = P(X = k) = pq^{k-1} \quad \textit{for} \quad k = 1, 2, \ldots \ . \tag{5.6}$$

The name "geometric" comes from the fact that the $f(k)$ values are the terms of a geometric series. Using the formula for the sum of a geometric series, we can confirm that they form a probability distribution:

$$\sum_{k=1}^{\infty} f(k) = \sum_{k=1}^{\infty} pq^{k-1} = \frac{p}{1-q} = 1. \tag{5.7}$$

From the preceding examples, we can glean some general observations about the probability and distribution functions of discrete random variables.

If $x_1, x_2, \ldots$ are the possible values of a discrete random variable X, then $f(x_i) \geq 0$ for all these values and $f(x) = 0$ otherwise. Furthermore, $\sum f(x_i) = 1$, because this sum equals the probability that X takes on any of its possible values, which is certain. Hence the total area of all the bars in the histogram of $f(x)$ is 1. Also, we can easily read off the histogram the probability of X falling in any given interval I, as the total area of those bars that cover the x_i values in I. For instance, for the X of Example 5.1.3, $\mathrm{P}(3 < X \leq 6) = \mathrm{P}(X = 4) + \mathrm{P}(X = 5) + \mathrm{P}(X = 6) = \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{1}{3}$, which is the total area of the bars over 4, 5, and 6.

The above observations, when applied to infinite intervals of the type $(-\infty, x]$, lead to the equation $F(x) = \mathrm{P}(X \in (-\infty, x]) = \sum_{x_i \leq x} \mathrm{P}(X = x_i) =$ sum of the areas of the bars over each $x_i \leq x$ and to the following properties of the distribution function:

Theorem 5.1.1. ***Properties of Distribution Functions.*** *The distribution function F of any random variable X has the following properties:*

1. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$, *since as $x \to -\infty$, the interval $(-\infty, x] \to \emptyset$.*
2. $F(\infty) = \lim_{x \to \infty} F(x) = 1$, *since as $x \to \infty$, the interval $(-\infty, x] \to \mathbb{R}$.*
3. *F is a nondecreasing function, since if $x < y$, then*

$$\begin{aligned} F(y) &= \mathrm{P}(X \in (-\infty, y]) = \mathrm{P}(X \in (-\infty, x]) + \mathrm{P}(X \in (x, y]) \\ &= F(x) + \mathrm{P}(X \in (x, y]), \end{aligned} \tag{5.8}$$

and so, $F(y)$ being the sum of $F(x)$ and a nonnegative term, we have $F(y) \geq F(x)$.

4. *F is continuous from the right at every x.*

These four properties of F hold not just for discrete random variables but for all types. Their proofs are outlined in Exercise 5.1.13 and those following it. Also, in more advanced courses, it is proved that any function with these four properties is the distribution function of some random variable.

While the distribution function can be used for any random variable, the probability function is useful only for discrete ones. To describe continuous random variables, we need another function, the so-called density function, instead, as will be seen in the next section.

The next theorem shows that the distribution function of a random variable X completely determines the distribution of X, that is, the probabilities $\mathrm{P}\{X \in I\}$ for all intervals I.

Theorem 5.1.2. ***Probabilities of a Random Variable Falling in Various Intervals.*** *For any random variable X and any real numbers x and y,*

1. $P(X \in (x,y]) = F(y) - F(x)$,
2. $P(X \in (x,y)) = \lim_{t \to y^-} F(t) - F(x)$,
3. $P(X \in [x,y]) = F(y) - \lim_{t \to x^-} F(t)$,
4. $P(X \in [x,y)) = \lim_{t \to y^-} F(t) - \lim_{t \to x^-} F(t)$.

For discrete random variables, the probability function and the distribution function determine each other: Let x_i, for $i = 1, 2, \ldots$, denote the possible values of X. Then clearly, for any x,

$$F(x) = \sum_{x_i \leq x} f(x_i) \tag{5.9}$$

and

$$f(x) = F(x) - \lim_{t \to x^-} F(t). \tag{5.10}$$

The first of these equations shows that $F(x)$ is constant between successive x_i values, and the latter equation shows that $f(x_i)$ equals the value of the jump of F at $x = x_i$.

Exercises

Exercise 5.1.1.

Let X be the number of hearts in a randomly dealt poker hand of five cards. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.2.

Let X be the number of heads obtained in five independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.3.

Let X be the number of heads minus the number of tails obtained in four independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.4.

Let X be the absolute value of the difference between the number of heads and the number of tails obtained in four independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.5.

Let X be the larger of the number of heads and the number of tails obtained in five independent tosses of a fair coin. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.6.

Let X be the number of heads minus the number of tails obtained in n independent tosses of a fair coin. Find a formula for its probability function and one for its distribution function.

Exercise 5.1.7.

Suppose we perform independent Bernoulli trials with parameter p, until we obtain two consecutive successes or two consecutive failures. Draw a tree diagram and find the probability function of the number of trials.

Exercise 5.1.8.

Suppose two players, A and B, play a game consisting of independent trials, each of which can result in a win for A or for B or in a draw D, until one player wins a trial. In each trial, $\mathrm{P}(A \text{ wins}) = p_{1,}$ $\mathrm{P}(B \text{ wins}) = p_{2,}$ and $\mathrm{P}(\text{draw}) = q = 1 - (p_1 + p_2)$. Let $X = n$ if A wins the game in the nth trial, and $X = 0$ if A does not win the game ever. Draw a tree diagram and find the probability function of X. Find also the probability that A wins (in any number of trials) and the probability that B wins. Show also that the probability of an endless sequence of draws is 0.

Exercise 5.1.9.

Let X be the number obtained in a single roll of a fair die. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.10.

We roll two fair dice, a blue and a red one, independently of each other. Let X be the number obtained on the blue die minus the number obtained on the red die. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.11.

We roll two fair dice independently of each other. Let X be the absolute value of the difference of the numbers obtained on them. Draw a histogram for its probability function and a graph for its distribution function.

Exercise 5.1.12.

Let the distribution function of a random variable X be given by

$$F(x) = \begin{cases} 0 & \text{if } x < -2 \\ 1/4 & \text{if } -2 \leq x < 2 \\ 7/8 & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases} \tag{5.11}$$

Find the probability function of X and graph both F and f.

Exercise 5.1.13.

Let $A_1, A_2, \ldots$ be a nondecreasing sequence of events on a sample space S, that is, let $A_n \subset A_{n+1}$ for $n = 1, 2, \ldots$, and let $A = \cup_{k=1}^{\infty} A_k$. Prove that $\mathrm{P}(A) = \lim_{n\to\infty} \mathrm{P}(A_n)$. *Hint*: Write A as the disjoint union $A_1 \cup [\cup_{k=2}^{\infty}(A_k - A_{k-1})]$, and apply the axiom of countable additivity.

Exercise 5.1.14.

Let $A_1, A_2, \ldots$ be a nonincreasing sequence of events on a sample space S, that is, let $A_n \supset A_{n+1}$ for $n = 1, 2, \ldots$, and let $A = \cap_{k=1}^{\infty} A_k$. Prove that $\mathrm{P}(A) = \lim_{n\to\infty} \mathrm{P}(A_n)$. *Hint*: Apply DeMorgan's laws to the result of the preceding exercise.

Exercise 5.1.15.

Prove that for the distribution function of any random variable, $\lim_{x\to-\infty} F(x) = 0$. *Hint*: Use the result of the preceding exercise and the theorem from real analysis that if $\lim_{n\to\infty} F(x_n) = L$ for every sequence $\langle x_n \rangle$ decreasing to $-\infty$, then $\lim_{x\to-\infty} F(x) = L$.

Exercise 5.1.16.

Prove that for the distribution function of any random variable, $\lim_{x\to\infty} F(x) = 1$. *Hint*: Use the result of Exercise 5.1.13 and the theorem from real analysis that if $\lim_{n\to\infty} F(x_n) = L$ for every sequence $\langle x_n \rangle$ increasing to ∞, then $\lim_{x\to\infty} F(x) = L$.

Exercise 5.1.17.

Prove that the distribution function F of any random variable is continuous from the right at every x. *Hint*: Use a modified version of the hints of the preceding exercises.

5.2 Continuous Random Variables

In this section, we consider random variables X whose possible values constitute a finite or infinite interval and whose distribution function is not a step function, but a continuous function. Such random variables are called *continuous* random variables.

The continuity of F implies that in Equation 5.10 $\lim_{t\to x^-} F(t) = \lim_{t\to x} F(t) = F(x)$, for every x, and so $f(x) = 0$, for every x. Thus, the probability function does not describe the distribution of such random variables because, in this case, the probability of X taking on any single value is zero. The latter statement can also be seen directly in the case of choosing a number at random from an interval, say from $[0, 1]$: If the probability of every value x were some positive c, then the total probability for obtaining any $x \in [0, 1]$ would be $\infty \cdot c = \infty$, in contradiction to the axiom requiring the total to be 1. On the other hand, we have no problem with $f(x) = 0$, for every x, since $\infty \cdot 0$ is indeterminate.

However, even if the probability of X taking on any single value is zero, the probability of X taking on any value in an *interval* need not be zero. Now, for a discrete random variable, the histogram of $f(x)$ readily displayed the probabilities of X falling in an interval I as the sum of the areas of the rectangles over I. Hence, a very natural generalization of such histograms suggests itself for continuous random variables: Just consider a continuous curve instead of the jagged top of the rectangles, and let the probability of X falling in I be the area under the curve over I. Thus we make the following formal definition:

Definition 5.2.1. ***Probability Density.*** *Let X be a continuous random variable with a given distribution function F. If there exists a nonnegative function*[3] *f that is integrable over $\mathbb{R}$ and for which*

$$\int_{-\infty}^{x} f(t)dt = F(x), \text{for all } x, \tag{5.12}$$

then f is called a probability density function[4] *(or briefly, density or p.d.f.) of X, and X is called absolutely continuous.*

Thus, if X has a density function, then

[3] Note that we are using the same letter f for this function as for the p.f. of a discrete r.v. This notation cannot lead to confusion though, since here we are dealing with continuous random variables rather than discrete ones. On the other hand, using the same letter for both functions will enable us to combine the two cases in some formulas later.

[4] The function f is not unique, because the integral remains unchanged if we change the integrand in a countable number of points. Usually, however, there is a version of f that is continuous wherever possible, and we shall call this version *the* density function of X, ignoring the possible ambiguity at points of discontinuity.

$$\mathrm{P}\left(X \in [x, y]\right) = F(y) - F(x) = \int_x^y f(t)dt, \tag{5.13}$$

and the probability remains the same whether we include or exclude one or both endpoints x and y of the interval. Also, if we set $x = \infty$ in Equation 5.12, we see that every p.d.f. must satisfy

$$\int_{-\infty}^{\infty} f(t)dt = 1. \tag{5.14}$$

In fact, any nonnegative piecewise continuous function f satisfying Equation 5.14 is a suitable density function and can be used to obtain the distribution function of a continuous random variable via Equation 5.12.

While the density function is not a probability, it is often used with differential notation to write the probability of X falling in an infinitesimal interval as[5]

$$\mathrm{P}\left(X \in [x, x + dx]\right) = \int_x^{x+dx} f(t)dt \sim f(x)dx. \tag{5.15}$$

By the fundamental theorem of calculus, the definition of the density function shows that, wherever f is continuous, F is differentiable and

$$F'(x) = f(x). \tag{5.16}$$

There exist, however, continuous random variables whose F is everywhere continuous but not differentiable and which therefore do not have a density function. Such random variables occur only very rarely in applications, and we do not discuss them in this book. In fact, we shall use the term continuous random variable—as most introductory books do—to denote random variables that possess a density function, instead of the precise term "absolutely continuous."

Let us turn now to examples of continuous random variables.

Example 5.2.1. Uniform Random Variable.

Consider a finite interval $[a, b]$, with $a < b$, and pick a point[6] X at random from it, that is, let the possible values of X be the numbers of $[a, b]$, and let X fall in each subinterval $[c, d]$ of $[a, b]$ with a probability that is proportional to the length of $[c, d]$ but does not depend on the location of $[c, d]$ within $[a, b]$. This distribution is achieved by the density function[7]

---

[5] The symbol $\sim$ means that the ratio of the expressions on either side of it tends to 1 as dx tends to 0 or, equivalently, that the limits of each side divided by dx are equal.

[6] We frequently use the words "point" and "number" interchangeably, ignoring the distinction between a number and its representation on the number line, just as the word "interval" is commonly used for both numbers and points.

[7] f is not unique: its values can be changed at a countable number of points, such as a and b, for instance, without affecting the probabilities, which are integrals of f.

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if} \quad a < x < b \\ 0 & \text{if } x \leq a \text{ or } x \geq b \end{cases} . \tag{5.17}$$

See Figure 5.6. Then, for $a \leq c \leq d \leq b$,

$$\mathrm{P}\left(X \in [c, d]\right) = \int_c^d f(t)dt = \frac{d-c}{b-a} , \tag{5.18}$$

which is indeed proportional to the length $d - c$ and does not depend on c and d in any other way.

The corresponding distribution function is given by

$$F\left(x\right) = \begin{cases} 0 & \text{if} \quad x < a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if} \quad x \geq b \end{cases} . \tag{5.19}$$

See Figure 5.7. ♦

Definition 5.2.2. *Uniform Random Variable.* *A random variable X with the above density is called uniform over $[a, b]$ or uniformly distributed over $[a, b]$. Its distribution is called the uniform distribution over $[a, b]$ and its density and distribution functions the uniform density and distribution functions over $[a, b]$.*

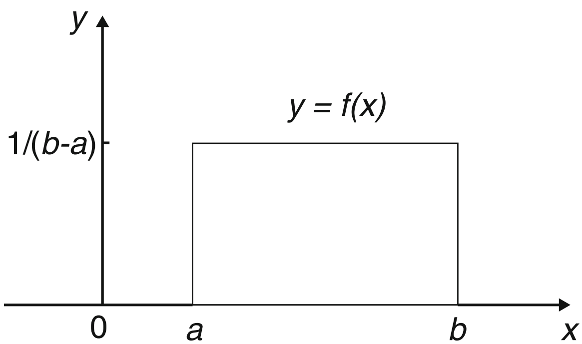

Fig. 5.6. The uniform density function over $[a, b]$

Often we know only the general shape of the density function, and we need to find the value of an unknown constant in its equation. Such constants can be determined by the requirement that f must satisfy $\int_{-\infty}^{\infty} f(t)dt = 1$, because the integral here equals the probability that X takes on any value whatsoever. The next two examples are of this type.

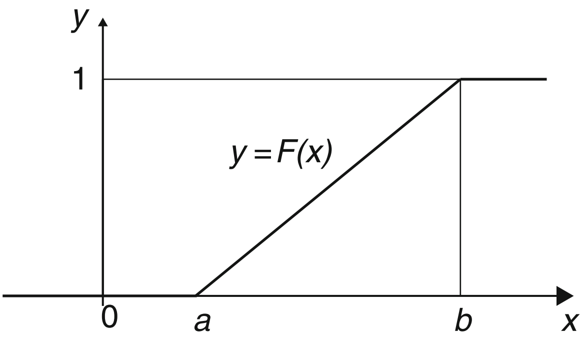

Fig. 5.7. The uniform distribution function over $[a, b]$

Example 5.2.2. Normalizing a p.d.f.

Let X be a random variable with p.d.f.

$$f(x) = \begin{cases} Cx^2 \text{ if } x \in [-1,1] \\ 0 \quad \text{if } x \notin [-1,1]\,. \end{cases} \tag{5.20}$$

Find the constant C and the distribution function of X.

Then,

$$1 = \int_{-\infty}^{\infty} f(t)dt = \int_{-1}^{1} Cx^2 dx = C\frac{x^3}{3}|_{-1}^{1} = \frac{2}{3}C. \tag{5.21}$$

Hence, $C = 3/2$. For $x \in [-1,1]$, the d.f. is

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-1}^{x} \frac{3}{2}t^2 dt = \frac{1}{2}x^3 + \frac{1}{2}. \tag{5.22}$$

Thus,

$$F(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{2}x^3 + \frac{1}{2} & \text{if } -1 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases} \tag{5.23}$$

♦

Example 5.2.3. Exponential Waiting Time.

Assume that the time T in minutes you have to wait on a certain summer night to see a shooting star has a probability density of the form

$$f(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ Ce^{-t/10} & \text{if } t > 0 \end{cases}. \tag{5.24}$$

Find the value of C and the distribution function of T and compute the probability that you have to wait more than 10 minutes.

Now,

$$1 = \int_{-\infty}^{\infty} f(t)dt = \int_{0}^{\infty} Ce^{-t/10}dt = -10Ce^{-t/10}|_0^\infty = 10C, \tag{5.25}$$

and so $C = 1/10$. Thus

$$f(t) = \begin{cases} 0 & \text{if } t \le 0 \\ \frac{1}{10}e^{-t/10} & \text{if } t > 0 \end{cases} \tag{5.26}$$

and, for $t > 0$,

$$F(t) = \mathrm{P}(T \le t) = \int_0^t \frac{1}{10}e^{-u/10}du = 1 - e^{-t/10}. \tag{5.27}$$

Consequently,

$$\mathrm{P}(T > 10) = 1 - F(10) = e^{-1} \simeq 0.368. \tag{5.28}$$

♦

The distribution of the example above is typical of many waiting time distributions occurring in real life, at least approximately. For instance, the time between the decay of atoms in a radioactive sample, the time one has to wait for the phone to ring in an office, and the time between customers showing up at some store are of this type; just the constants differ. (The reasons for the prevalence of this distribution will be discussed later under the heading “Poisson process.”)

Definition 5.2.3. ***Exponential Random Variable.*** *A random variable T is called exponential with parameter $\lambda > 0$ if it has density*

$$f(t) = \begin{cases} 0 & \text{if } t < 0 \\ \lambda e^{-\lambda t} & \text{if } t \ge 0 \end{cases} \tag{5.29}$$

and distribution function

$$F(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 - e^{-\lambda t} & \text{if } t \ge 0. \end{cases} \tag{5.30}$$

There exist random variables that are neither discrete nor continuous; they are said to be of *mixed type.* Here is an example:

Example 5.2.4. A Mixed Random Variable.

Suppose we toss a fair coin, and if it comes up H, then $X = 1$, and if it comes up T, then X is determined by spinning a pointer and noting its final position on a scale from 0 to 2, that is, X is then uniformly distributed over the interval $[0, 2]$.

Let

$$F_1(x) = \mathrm{P}(X \leq x|H) = \begin{cases} 0 \text{ if } x < 1 \\ 1 \text{ if } x \geq 1 \end{cases} \tag{5.31}$$

and

$$F_2(x) = \mathrm{P}(X \leq x|T) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2}x & \text{if } 0 \leq x < 2 \\ 1 & \text{if } 2 \leq x. \end{cases} \tag{5.32}$$

Then, according to the theorem of total probability, the distribution function F is given by

$$F(x) = \frac{1}{2}F_1(x) + \frac{1}{2}F_2(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{4}x & \text{if } 0 \leq x < 1 \\ \frac{1}{4}x + \frac{1}{2} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } 2 \leq x \end{cases} \tag{5.33}$$

and its graph is given by Figure 5.8.

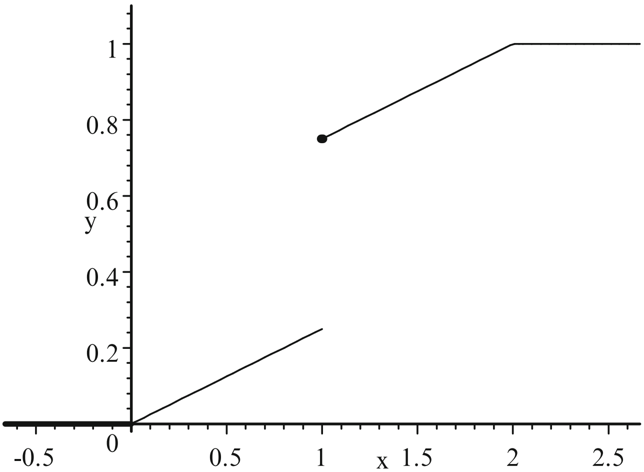

Fig. 5.8. A mixed-type distribution function

Note that

$$F'(x) = f(x) = \begin{cases} 0 \text{ if } x < 0 \\ \frac{1}{4} \text{ if } 0 < x < 1 \\ \frac{1}{4} \text{ if } 1 < x < 2 \\ 0 \text{ if } 2 < x \end{cases} \tag{5.34}$$

exists everywhere except at $x = 0$, 1, and 2, but because of the jump of F at 1, it is not a true density function. Indeed,

$$F(x) = \begin{cases} \int_{-\infty}^{x} f(t)dt & \text{if } x < 1 \\ \int_{-\infty}^{x} f(t)dt + \frac{1}{2} & \text{if } 1 \leq x, \end{cases} \tag{5.35}$$

and so $F(x) \neq \int_{-\infty}^{x} f(t)dt$ for all x, as required by the definition of density functions. ♦

Exercises

Exercise 5.2.1.

A continuous random variable X has a density of the form

$$f(x) = \begin{cases} Cx & \text{if } 0 \leq x \leq 4 \\ 0 & \text{if } x < 0 \text{ or } x > 4 \end{cases}. \tag{5.36}$$

1. Find C.
2. Sketch the density function of X.
3. Find the distribution function of X and sketch its graph.
4. Find the probability $\mathrm{P}(X < 1)$.
5. Find the probability $\mathrm{P}(2 < X)$.

Exercise 5.2.2.

A continuous random variable X has a density of the form $f(x) = Ce^{-|x|}$, defined on all of $\mathbb{R}$:

1. Find C.
2. Sketch the density function of X.
3. Find the distribution function of X and sketch its graph.
4. Find the probability $\mathrm{P}(-2 < X < 1)$.
5. Find the probability $\mathrm{P}(2 < |X|)$.

Exercise 5.2.3.

A continuous random variable X has a density of the form

$$f(x) = \begin{cases} \frac{C}{x^2} & \text{if } x \geq 1 \\ 0 & \text{if } x < 1 \end{cases}. \tag{5.37}$$

1. Find C.
2. Sketch the density function of X.
3. Find the distribution function of X and sketch its graph.
4. Find the probability $\mathrm{P}(X < 2)$.
5. Find the probability $\mathrm{P}(2 < |X|)$.

Exercise 5.2.4.

A continuous random variable X has a density of the form

$$f(x) = \begin{cases} \frac{C}{x^2} & \text{if } |x| \geq 1 \\ 0 & \text{if } |x| < 1 \end{cases} . \tag{5.38}$$

1. Find C.
2. Sketch the density function of X.
3. Find the distribution function of X and sketch its graph.
4. Find the probability $\mathrm{P}(X < 2)$.
5. Find the probability $\mathrm{P}(2 < |X|)$.

Exercise 5.2.5.

Let X be a mixed random variable with distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{6}x & \text{if } 0 \leq x < 1 \\ \frac{1}{3} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } 2 \leq x \end{cases} . \tag{5.39}$$

1. Devise an experiment whose outcome is this X.
2. Find the probability $\mathrm{P}(X < 1/2)$.
3. Find the probability $\mathrm{P}(X < 3/2)$.
4. Find the probability $\mathrm{P}(1/2 < X < 2)$.
5. Find the probability $\mathrm{P}(X = 1)$.
6. Find the probability $\mathrm{P}(X > 1)$.
7. Find the probability $\mathrm{P}(X = 2)$.

Exercise 5.2.6.

Let X be a mixed random variable with distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{3}x + \frac{1}{6} & \text{if } 0 \leq x < 1 \\ \frac{2}{3} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } 2 \leq x \end{cases} . \tag{5.40}$$

1. Devise an experiment whose outcome is this X.
2. Find the probability $\mathrm{P}(X < 1/2)$.
3. Find the probability $\mathrm{P}(X < 3/2)$.
4. Find the probability $\mathrm{P}(1/2 < X < 2)$.
5. Find the probability $\mathrm{P}(X = 1)$.
6. Find the probability $\mathrm{P}(X > 1)$.
7. Find the probability $\mathrm{P}(X = 3/2)$.

Exercise 5.2.7.

Let X be a mixed random variable with distribution function F given by the graph in Figure 5.9:

1. Find a formula for $F(x)$.
2. Find the probability $\mathrm{P}(X < 1/2)$.
3. Find the probability $\mathrm{P}(X < 3/2)$.
4. Find the probability $\mathrm{P}(1/2 < X < 2)$.
5. Find the probability $\mathrm{P}(X = 1)$.
6. Find the probability $\mathrm{P}(X > 1)$.
7. Find the probability $\mathrm{P}(X = 2)$.

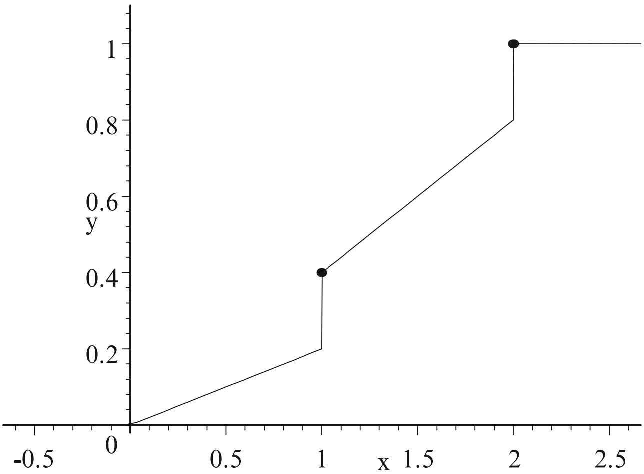

Fig. 5.9.

Exercise 5.2.8.

Let X be a mixed random variable with distribution function F given by the graph in Figure 5.10:

1. Find a formula for $F(x)$.
2. Find the probability $\mathrm{P}(X < 1/2)$.
3. Find the probability $\mathrm{P}(X < 3/2)$.
4. Find the probability $\mathrm{P}(1/2 < X < 2)$.
5. Find the probability $\mathrm{P}(X = 1)$.
6. Find the probability $\mathrm{P}(X > 1)$.
7. Find the probability $\mathrm{P}(X = 2)$.

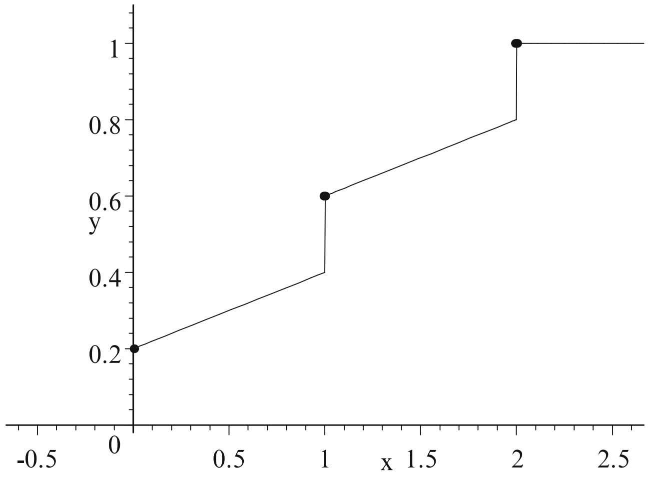

Fig. 5.10.

5.3 Functions of Random Variables

In many applications we need to find the distribution of a function of a random variable. For instance, we may know from measurements the distribution of the radius of stars, and we may want to know the distribution of their volumes. (Probabilities come in—as in several examples of Chapter 4—from a random choice of a single star.) Or we may know the income distributions in different countries and want to change scales to be able to compare them. We shall encounter many more examples in the rest of the book. We start off with the change of scale example in a general setting.

Example 5.3.1. Linear Functions of Random Variables.

Let X be a random variable with a known distribution function F_X and define a new random variable as $Y = aX + b$, where $a \neq 0$ and b are given constants.

If X is discrete, then we can obtain the probability function f_Y of Y very easily by solving the equation in its definition:

$$f_Y(y) = \mathrm{P}(Y = y) = \mathrm{P}(aX + b = y) = \mathrm{P}\left(X = \frac{y-b}{a}\right) = f_X\left(\frac{y-b}{a}\right). \tag{5.41}$$

Equivalently, if x is a possible value of X, that is, $f_X(x) \neq 0$, then $f_Y(y) = f_X(x)$ for $y = ax + b$, which is the corresponding possible value of Y.

If X is continuous, then we cannot imitate the above procedure, because the density function is not a probability. We can, however, obtain the distribution function F_Y of Y similarly, by solving the inequality in its definition: For $a > 0$,

$$F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}(aX + b \le y) = \mathrm{P}\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right), \tag{5.42}$$

and for $a < 0$,

$$F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}(aX+b \le y) = \mathrm{P}\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right). \tag{5.43}$$

If X is continuous with density f_X, then F_X is differentiable and $f_X = F'_X$. As Equations 5.42 and 5.43 show, then F_Y is also differentiable. Hence Y too is continuous, with density function

$$f_Y(y) = F'_Y(y) = \pm\frac{d}{dy}F_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|}F'_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right). \tag{5.44}$$

♦

Example 5.3.2. Shifting and Stretching a Discrete Uniform Variable.

Let X denote the number obtained in the roll of a die and let $Y = 2X+10$. Then the p.f. of X is

$$f_X(x) = \begin{cases} 1/6 \text{ if } & x = 1, 2, \ldots, 6 \\ 0 & \text{otherwise.} \end{cases} \tag{5.45}$$

Thus, using Equation 5.41 with this f_X and with $a = 2$ and $b = 10$, we get the p.f. of Y as

$$f_Y(y) = f_X\left(\frac{y-10}{2}\right) = \begin{cases} 1/6 \text{ if } & y = 12, 14, \ldots, 22 \\ 0 & \text{otherwise.} \end{cases} \tag{5.46}$$

We can obtain the same result more simply, by tabulating the possible x and $y = 2x + 10$ values and the corresponding probabilities:

x	1	2	3	4	5	6
y	12	14	16	18	20	22
$f_X(x) = f_Y(y)$	1/6	1/6	1/6	1/6	1/6	1/6

♦

Example 5.3.3. Shifting and Stretching a Uniform Variable.

Let X be uniform on the interval $[-1, 1]$ and let $Y = 2X + 10$. Then the p.d.f. of X is

$$f_X(x) = \begin{cases} 1/2 \text{ if } & x \in [-1, 1] \\ 0 & \text{otherwise.} \end{cases} \tag{5.47}$$

If $X = -1$, then $Y = 2(-1) + 10 = 8$, and if $X = 1$, then $Y = 2 \cdot 1 + 10 = 12$. Thus, the interval $[-1, 1]$ gets changed into $[8, 12]$, and so Equation 5.44, with the present f_X and with $a = 2$ and $b = 10$, yields

$$f_Y(y) = \frac{1}{2} f_X\left(\frac{y - 10}{2}\right) = \begin{cases} 1/4 \text{ if } y \in [8, 12] \\ 0 \quad \text{otherwise.} \end{cases} \tag{5.48}$$

Notice that here the p.d.f. got shifted and stretched in much the same way as the p.f. in the preceding example, but there the values of the p.f. remained $\frac{1}{6}$, while here the values of the p.d.f. have become halved. The reason for this difference is clear: In the discrete case, the number of possible values has not changed (both X and Y had six), but in the continuous case, the p.d.f. got stretched by a factor of 2 (from width 2 to width 4) and so, to compensate for that, in order to have a total area of 1, we had to halve the density.

The foregoing examples can easily be generalized to the case in which $Y = g(X)$, for any invertible function g. Rather than summarizing the results in a theorem, we just give prescriptions for the procedures and illustrate them with examples:

1. For discrete X tabulate the possible values x of X together with $y = g(x)$ and $f_X(x) = f_Y(y)$.
2. For continuous X, solve the inequality in $F_Y(y) = \mathrm{P}(Y \le y) = \mathrm{P}(g(X) \le y)$ to obtain $F_Y(y)$ in terms of $F_X\left(g^{-1}(y)\right)$. (We must be careful to reverse the inequality when solving for X if g is decreasing.) To obtain the p.d.f. $f_Y(y)$, differentiate $F_Y(y)$.

Example 5.3.4. Squaring a Binomial.

Let X be binomial with parameters $n = 3$ and $p = \frac{1}{2}$ and let $Y = X^2$. Then we can obtain f_Y by tabulating the possible X and $Y = X^2$ values and the corresponding probabilities:

x	0	1	2	3
y	0	1	4	9
$f_X(x) = f_Y(y)$	1/8	3/8	3/8	1/8

♦

Example 5.3.5. Squaring a Positive Uniform Random Variable.

Let X be uniform on the interval $[1, 3]$ and let $Y = X^2$. Then the p.d.f. of X is

$$f_X(x) = \begin{cases} 1/2 & \text{if } x \in [1,3] \\ 0 & \text{otherwise.} \end{cases} \tag{5.49}$$

Now, $g(X) = X^2$ is one-to-one for the possible values of X, which are positive, and so, for $y \geq 0$,

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(X^2 \leq y) = \mathrm{P}\left(X \leq \sqrt{y}\right) = F_X\left(\sqrt{y}\right). \tag{5.50}$$

Hence, by the chain rule,

$$f_Y(y) = \frac{d}{dy} F_X\left(\sqrt{y}\right) = f_X\left(\sqrt{y}\right) \frac{d\sqrt{y}}{dy} = \begin{cases} \frac{1}{2}\frac{1}{2\sqrt{y}} & \text{if } \ y \in [1,9] \\ 0 & \text{otherwise.} \end{cases} \tag{5.51}$$

We can check that this f_Y is indeed a density function:

$$\int_1^9 \frac{1}{4\sqrt{y}} dy = \left.\frac{\sqrt{y}}{2}\right|_1^9 = \frac{1}{2}\left(\sqrt{9} - \sqrt{1}\right) = 1. \tag{5.52}$$

♦

Example 5.3.6. Random Number Generation.

An important application of the procedures for changing of random variables described above is to the computer simulation of physical systems with random inputs. Most mathematical and statistical software packages produce so-called random numbers (or more precisely, pseudorandom numbers) that are uniformly distributed on the interval $[0,1]$. (Though such numbers are generated by deterministic algorithms, they are for most practical purposes a good substitute for samples of independent, uniform random variables on the interval $[0,1]$.) Often, however, we need random numbers with a different distribution and want to transform the uniform random numbers to new numbers that have the desired distribution.

Suppose we need random numbers that have the continuous distribution function F and that F is strictly increasing where it is not 0 or 1. (The restrictions on F can be removed, but we do not want to get into this.) Then F has a strictly increasing inverse F^{-1} over $[0,1]$, which we can use as the function g in Part 2 of the general procedure given above. Thus, letting $Y = F^{-1}(X)$, with X being uniform on $[0,1]$, we have

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(F^{-1}(X) \leq y) = \mathrm{P}(X \leq F(y)) = F(y), \tag{5.53}$$

where the last step follows from the fact that $\mathrm{P}(X \leq x) = x$ on $[0,1]$ for an X that is uniform on $[0,1]$, with the substitution $x = F(y)$. (See Equation 5.19.)

Thus, if x_1, x_2,... are random numbers uniform on $[0,1]$, produced by the generator, then the numbers $y_1 = F^{-1}(x_1)$, $y_2 = F^{-1}(x_2)$,... are random numbers with the distribution function F. ♦

If g is not one-to-one, we can still follow the procedures of the examples above, but, for some y, we have more than one solution of the equation $y = g(x)$ or of the corresponding inequality, and we must consider all of those solutions, as in the following examples.

Example 5.3.7. The X^2 Function.

Let X be a random variable with a known distribution function F_X and define a new random variable as $Y = X^2$.

If X is discrete, then we can obtain the probability function f_Y of Y as

$$f_Y(y) = \mathrm{P}(X^2 = y) = \begin{cases} \mathrm{P}\left(X = \pm\sqrt{y}\right) = f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right) & \text{if } \ y > 0 \\ \mathrm{P}\left(X = 0\right) = f_X\left(0\right) & \text{if } \ y = 0 \\ 0 & \text{if } \ y < 0. \end{cases} \tag{5.54}$$

For continuous X, the distribution function F_Y of Y is given by

$$F_Y(y) = \mathrm{P}(X^2 \le y) = \begin{cases} \mathrm{P}\left(-\sqrt{y} \le X \le \sqrt{y}\right) = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right) & \text{if } \ y > 0 \\ 0 & \text{if } \ y \le 0, \end{cases} \tag{5.55}$$

and for discrete X, we have

$$F_Y(y) = \begin{cases} \mathrm{P}\left(-\sqrt{y} \le X \le \sqrt{y}\right) = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right) + f_X\left(-\sqrt{y}\right) & \text{if } \ y \ge 0 \\ 0 & \text{if } \ y < 0. \end{cases} \tag{5.56}$$

If X is continuous and has density function f_X, then differentiating Equation 5.55 we get

$$f_Y(y) = F_Y'(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right] & \text{if } \ y > 0 \\ 0 & \text{if } y \le 0. \end{cases} \tag{5.57}$$

♦

Example 5.3.8. Distribution of $(X-2)^2$ for a Binomial.

Let X be binomial with parameters $n = 3$ and $p = \frac{1}{2}$, and let $Y = (X-2)^2$. Rather than developing a formula like Equation 5.54, the best way to proceed is to tabulate the possible values of X and Y and the corresponding probabilities, as in Example 5.3.4:

x	0	1	2	3
y	4	1	0	1
$f_X(x)$	1/8	3/8	3/8	1/8

Now, $Y = 1$ occurs when $X = 1$ or 3. Since these cases are mutually exclusive, $\mathrm{P}(Y = 1) = \mathrm{P}(X = 1) + \mathrm{P}(X = 3) = 3/8 + 1/8 = 1/2$. Hence, the table of f_Y is

y	0	1	4
$f_Y(y)$	3/8	1/2	1/8

♦

Example 5.3.9. Distribution of X^2 for a Uniform X.

Let X be uniform on the interval $[-1,1]$ and let $Y = X^2$. Then, by Formula 5.19,

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ \dfrac{x+1}{2} & \text{if } -1 \leq x < 1 \\ 1 & \text{if} \quad x \geq 1. \end{cases} \tag{5.58}$$

Substituting this F_X into Equation 5.55 and observing that $\dfrac{\sqrt{y}+1}{2} - \dfrac{-\sqrt{y}+1}{2} = \sqrt{y}$, we get

$$F_Y(y) = \begin{cases} 0 & \text{if} \quad y < 0 \\ \sqrt{y} & \text{if } 0 \leq y < 1 \\ 1 & \text{if} \quad y \geq 1. \end{cases} \tag{5.59}$$

We can obtain the density of Y by differentiating F_Y, as

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.60}$$

♦

Example 5.3.10. Distribution of X^2 for a Nonuniform X.

Let X be a random variable with p.d.f. $f(x) = \frac{3x^2}{2}$ on the interval $[-1,1]$ and 0 elsewhere and $Y = X^2$. Find the distribution function and the density function of Y.
Solution: If $0 < y < 1$, then

$$\begin{aligned} F_Y(y) &= \mathrm{P}(X^2 \leq y) = \mathrm{P}\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{3x^2}{2} dx = \left[\frac{x^3}{2}\right]_{-\sqrt{y}}^{\sqrt{y}} = y^{3/2}. \end{aligned} \tag{5.61}$$

Thus

$$F_Y(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ y^{3/2} & \text{if } 0 < y < 1 \\ 1 & \text{if } y \geq 1. \end{cases} \tag{5.62}$$

Hence

$$f_Y(y) = F_Y'(y) = \begin{cases} \frac{3\sqrt{y}}{2} & \text{if } 0 < y < 1 \\ 0 & \text{if } y \leq 0 \text{ or } y \geq 1. \end{cases} \tag{5.63}$$

♦

Example 5.3.11. Coordinates of a Uniform Random Variable on a Circle.

Suppose that a point is moving around a circle of radius r centered at the origin of the xy coordinate system with constant speed, and we observe it at a random instant. What is the distribution of each of the point's coordinates at that time?

Since the point is observed at a random instant, its position is uniformly distributed on the circle. Thus its polar angle Θ is a uniform random variable on the interval $[0, 2\pi]$, with constant density $f_\Theta(\theta) = \frac{1}{2\pi}$ there and 0 elsewhere. We want to find the distributions of $X = r\cos\Theta$ and $Y = r\sin\Theta$.

Now, for a given $x = r\cos\theta$, there are two solutions modulo 2π: $\theta_1 = \arccos\frac{x}{r}$ and $\theta_2 = 2\pi - \arccos\frac{x}{r}$. So if $X \le x$, then Θ falls in the angle on the left between these two values. Thus

$$F_X(x) = \mathrm{P}(X \le x) = \begin{cases} 0 & \text{if} \quad x < -r \\ \frac{\theta_2-\theta_1}{2\pi} = 1 - \frac{1}{\pi}\arccos\frac{x}{r} & \text{if} \quad -r \le x < r \\ 1 & \text{if} \quad r \le x. \end{cases} \tag{5.64}$$

Hence

$$f_X(x) = F'_X(x) = \begin{cases} \frac{1}{\pi\sqrt{r^2-x^2}} & \text{if } -r < x < r \\ 0 & \text{otherwise.} \end{cases} \tag{5.65}$$

The density of X can also be obtained directly from Figure 5.11 by using Equation 5.15. For $x > 0$ and $dx > 0$, the variable X falls into the interval $[x, x + dx]$ if and only if Θ falls into either of the intervals of size $d\theta$ at θ_1 and θ_2. (For negative x or dx, we need obvious modifications.) Thus, $f_X(x)dx = 2 \cdot \frac{1}{2\pi}d\theta$, and so $f_X(x) = \frac{1}{\pi} \cdot \frac{d\theta}{dx} = \frac{1}{\pi} \cdot 1 \Big/ \frac{dx}{d\theta} = \frac{1}{\pi\sqrt{r^2-x^2}}$ as before.

We leave the analogous computation for the distribution of the y-coordinate as an exercise. ♦

Exercises

Exercise 5.3.1.

Let X be a discrete uniform random variable with possible values $-5, -4, \ldots, 4, 5$. Find the probability function and the distribution function of $Y = X^2 - 3X$.

Exercise 5.3.2.

Let X be a binomial random variable with parameters $p = \frac{1}{2}$ and $n = 6$. Find the probability function and the distribution function of $Y = X^2 - 2X$.

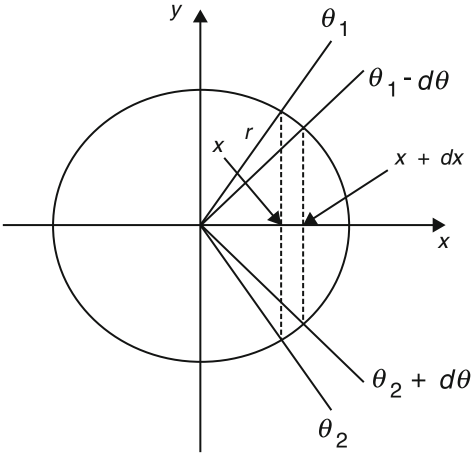

Fig. 5.11. Density of the x-coordinate of a random point on a circle

Exercise 5.3.3.

Let X be a Bernoulli random variable with $p = \frac{1}{2}$ and $Y = \arctan X$. Find the probability function and the distribution function of Y.

Exercise 5.3.4.

Let X be a discrete random variable with probability function f_X. Find formulas for the probability function and the distribution function of $Y = (X - a)^2$, where a is an arbitrary constant.

Exercise 5.3.5.

Let X be a random variable uniformly distributed on the interval (0,1) and $Y = \ln X$. Find the distribution function and the density function of Y.

Exercise 5.3.6.

Let X be a random variable uniformly distributed on the interval [−1,1] and $Y = |X|$. Find the distribution function and the density function of Y.

Exercise 5.3.7.

Let X be a continuous random variable with density function f_X. Find formulas for the distribution function and the density function of $Y = |X|$.

Exercise 5.3.8.

Assume that the distribution of the radius R of stars has a density function f_R. Find formulas for the density and the distribution function of their volume $V = \frac{4}{3}R^3\pi$.

Exercise 5.3.9.

Find the distribution function and the density function of Y in Example 5.3.11.

Exercise 5.3.10.

Let X be a continuous random variable with density f_X. Find formulas for the distribution function and the density function of $Y = (X - a)^2$, where a is an arbitrary constant.

Exercise 5.3.11.

Let X be a continuous random variable with a continuous distribution function F that is strictly increasing where it is not 0 or 1. Show that the random variable $Y = F(X)$ is uniformly distributed on the interval [0,1].

Exercise 5.3.12.

Let X be a random variable uniformly distributed on the interval [−2,2] and $Y = (X - 1)^2$:

a) Find the density function and the distribution function of X.
b) Find the distribution function and the density function of Y.

5.4 Joint Distributions

In many applications, we need to consider two or more random variables simultaneously. For instance, the two-way classification of voters in Example 4.3.3 can be regarded to involve two random variables, if we assign numbers to the various age groups and party affiliations.

In general, we want to consider joint probabilities of events defined by two or more random variables on the same sample space. The probabilities of all such events constitute *the joint distribution* or the *bivariate* (for two variables) or *multivariate* (for two or more variables) *distribution* of the given random variables and can be described by their joint p.f., d.f., or p.d.f., much as for single random variables.

Definition 5.4.1. ***Joint Probability Function.***

Let X and Y be two discrete random variables on the same sample space. The function of two variables defined by[8] $f(x,y) = \mathrm{P}(X = x, Y = y)$, *for all possible values*[9] x *of* X *and* y *of* Y, *is called the joint or bivariate probability function of* X *and* Y *or of the pair* (X, Y).

[8] $\mathrm{P}(X = x,\, Y = y)$ stands for $\mathrm{P}(X = x \text{ and } Y = y) = \mathrm{P}(\{X = x\} \cap \{Y = y\})$.

[9] Sometimes $f(x,y)$ is defined for all real numbers x, y, with $f(x,y) = 0$ if $\mathrm{P}(X = x) = 0$ or $\mathrm{P}(Y = y) = 0$.

Similarly, for a set of n random variables on the same sample space, with n a positive integer greater than 2, we define the joint or multivariate probability function of $(X_1, X_2, \ldots, X_n)$ as the function given by $f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$, for all possible values x_i of each X_i or for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

If for two random variables we sum $f(x, y)$ over all possible values y of Y, then we get the (*marginal*)[10] probability function f_X (or f_1) of X. Indeed,

$$\begin{aligned}\sum_y f(x,y) &= \sum_y \mathrm{P}(X = x, Y = y) = \mathrm{P}(\{X = x\} \cap (\cup_y \{Y = y\})) \\ &= \mathrm{P}(\{X = x\} \cap S) = \mathrm{P}(X = x) = f_X(x). \end{aligned} \tag{5.66}$$

Similarly, if we sum $f(x, y)$ over all possible values x of X, then we get the probability function f_Y (or f_2) of Y, and if we sum $f(x, y)$ over all possible values x of X and y of Y both, in either order then, of course, we get 1.

For n random variables, if we sum $f(x_1, x_2, \ldots, x_n)$ over all possible values x_i of any X_i, then we get the joint (marginal) probability function of the $n-1$ random variables X_j with $j \neq i$, and if we sum over all possible values of any k of them, then we get the joint (marginal) probability function of the remaining $n - k$ random variables.

Definition 5.4.2. ***Joint Distribution Function.***

Let X and Y be two arbitrary random variables on the same sample space. The function of two variables defined by $F(x, y) = \mathrm{P}(X \leq x, Y \leq y)$, for all real x and y, is called the joint or bivariate distribution function of X and Y or of the pair (X, Y).

The functions[11] *$F_X(x) = F(x, \infty)$ and $F_Y(y) = F(\infty, y)$ are called the (marginal) distribution functions of X and Y.*

Similarly, for a set of n random variables on the same sample space, with n a positive integer greater than 2, we define the joint or multivariate distribution function of $(X_1, X_2, \ldots, X_n)$ as the function given by $F(x_1, x_2, \ldots, x_n) = \mathrm{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$, for all real numbers $x_1, x_2, \ldots, x_n$.

If we substitute ∞ for any of the arguments of $F(x_1, x_2, \ldots, x_n)$, we get the marginal d.f.'s of the random variables that correspond to the remaining arguments.

For joint distributions, we have the following obvious theorem:

Theorem 5.4.1. ***Joint Distribution of Two Functions of Two Discrete Random Variables.*** *If X and Y are two discrete random variables with joint probability function $f_{X,Y}(x, y)$ and $U = g(X, Y)$ and $V = h(X, Y)$ any two functions, then the joint probability function of U and V is given by*

[10] The adjective "marginal" is really unnecessary; we just use it occasionally to emphasize the relation to the joint distribution.

[11] $F(x, \infty)$ is a shorthand for $\lim_{y \to \infty} F(x, y)$, etc.

$$f_{U,V}(u,v) = \sum_{(x,y):g(x,y)=u,\ h(x,y)=v} \sum f_{X,Y}(x,y). \tag{5.67}$$

Example 5.4.1. Sum and Absolute Difference of Two Dice.

Roll two fair dice as in 2.3.3, and let X and Y denote the numbers obtained with them. Find the joint probability function of $U = X+Y$ and $V = |X-Y|$.

First, we construct a table of the values of U and V, for all possible outcomes x and y (Table 5.1):

$y\backslash x$	1	2	3	4	5	6
1	2,0	3,1	4,2	5,3	6,4	7,5
2	3,1	4,0	5,1	6,2	7,3	8,4
3	4,2	5,1	6,0	7,1	8,2	9,3
4	5,3	6,2	7,1	8,0	9,1	10,2
5	6,4	7,3	8,2	9,1	10,0	11,1
6	7,5	8,4	9,3	10,2	11,1	12,0

Table 5.1. The values of $U = X + Y$ and $V = X - Y$ for the numbers X and Y showing on two dice

By assumption, each pair of x and y values has probability $1/36$, and so each pair (u, v) of U and V values has as its probability $1/36$ times the number of boxes in which it appears. Hence, for instance, $f_{U,V}(3, 1) = \mathrm{P}(U = 3, V = 1) = \mathrm{P}(X = 1, Y = 2) + \mathrm{P}(X = 2, Y = 1) = \frac{2}{36}$. Thus, the joint probability function $f_{U,V}(u, v)$ of U and V is given by the table below (Table 5.2), with the marginal probability function $f_U(u)$ shown as the row sums on the right margin and the marginal probability function $f_V(v)$ shown as the column sums on the bottom margin:

$u\backslash v$	0	1	2	3	4	5	$f_U(u)$
2	1/36	0	0	0	0	0	1/36
3	0	2/36	0	0	0	0	2/36
4	1/36	0	2/36	0	0	0	3/36
5	0	2/36	0	2/36	0	0	4/36
6	1/36	0	2/36	0	2/36	0	5/36
7	0	2/36	0	2/36	0	2/36	6/36
8	1/36	0	2/36	0	2/36	0	5/36
9	0	2/36	0	2/36	0	0	4/36
10	1/36	0	2/36	0	0	0	3/36
11	0	2/36	0	0	0	0	2/36
12	1/36	0	0	0	0	0	1/36
$f_V(v)$	6/36	10/36	8/36	6/36	4/36	2/36	1

Table 5.2. The joint and marginal probability functions of $U = X + Y$ and $V = X - Y$ for the numbers X and Y showing on two dice

Example 5.4.2. Maximum and Minimum of Three Integers.

Choose three numbers X_1, X_2, X_3 without replacement and with equal probabilities from the set $\{1, 2, 3, 4\}$, and let $X = \max\{X_1, X_2, X_3\}$ and $Y = \min\{X_1, X_2, X_3\}$. Find the joint probability function of X and Y.

First, we list the set of all 24 possible outcomes (Table 5.3), together with the values of X and Y:

X_1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4
X_2	2	2	3	3	4	4	1	1	3	3	4	4	1	1	2	2	4	4	1	1	2	2	3	3
X_3	3	4	2	4	2	3	3	4	1	4	1	3	2	4	1	4	1	2	2	3	1	3	1	2
X	3	4	3	4	4	4	3	4	3	4	4	4	3	4	3	4	4	4	4	4	4	4	4	4
Y	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	2	1	1	1	2	1	2

Table 5.3. The values of $X = \max(X_1, X_2, X_3)$ and $Y = \min(X_1, X_2, X_3)$

Now, each possible outcome has probability $1/24$, and so we just have to count the number of times each pair of X, Y values occurs and multiply it by $1/24$ to get the probability function $f(x, y)$ of (X, Y). This p.f. is given in the following table (Table 5.4), together with the marginal probabilities $f_Y(y)$ on the right and $f_X(x)$ at the bottom:

$y\backslash x$	3	4	Any x
1	1/4	1/2	3/4
2	0	1/4	1/4
Any y	1/4	3/4	1

Table 5.4. The joint p.f. and marginals of $X = \max(X_1, X_2, X_3)$ and $Y = \min(X_1, X_2, X_3)$

♦

Example 5.4.3. Multinomial Distribution.

Suppose we have k types of objects and we perform n independent trials of choosing one of these objects, with probabilities $p_1, p_2, \ldots, p_k$ for the different types in each of the trials, where $p_1 + p_2 + \cdots + p_k = 1$. Let $N_1, N_2, \ldots, N_k$ denote the numbers of objects obtained in each category. Then clearly, the joint probability function of $N_1, N_2, \ldots, N_k$ is given by

$$\begin{aligned} f(n_1, n_2, \ldots, n_k) &= \mathrm{P}(N_1 = n_1, N_2 = n_2, \ldots, N_k = n_k) \\ &= \binom{n}{n_1, n_2, \ldots, n_k} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \end{aligned} \tag{5.68}$$

for every choice of nonnegative integers $n_1, n_2, \ldots, n_k$ with $n_1 + n_2 + \cdots + n_k = n$ and $f(n_1, n_2, \ldots, n_k) = 0$ otherwise. ♦

Next, we consider the joint distributions of continuous random variables.

Definition 5.4.3. ***Joint Density Function.***

Let X and Y be two continuous random variables on the same probability space. If there exists an integrable nonnegative function $f(x,y)$ on $\mathbb{R}^2$ such that

$$\mathrm{P}(a < X < b, c < Y < d) = \int_c^d \int_a^b f(x,y)dxdy \tag{5.69}$$

for all real numbers a, b, c, d, then f is called a joint or bivariate probability density function[12] *of X and Y or of the pair (X,Y), and X and Y are said to be jointly continuous.*

Similarly, for a set of n continuous random variables on the same probability space, with n a positive integer greater than 2, if there exists an integrable nonnegative function $f(x_1, x_2, \ldots, x_n)$ on $\mathbb{R}^n$ such that, for any coordinate rectangle[13] *R of $\mathbb{R}^n$,*

$$P((X_1, X_2, \ldots, X_n) \in R) = \int \cdots \int_R f(x_1, x_2, \ldots, x_n)dx_1 \ldots dx_n, \tag{5.70}$$

then f is called a joint or multivariate probability density function of $X_1, X_2, \ldots, X_n$ or of the point or vector $(X_1, X_2, \ldots, X_n)$, and $X_1, X_2, \ldots, X_n$ are said to be jointly continuous.

Similarly as for discrete variables, in the continuous bivariate case, $\int_{-\infty}^{\infty} f(x,y)dx = f_Y(y)$ is the (marginal) density of Y, and $\int_{-\infty}^{\infty} f(x,y)dy = f_X(x)$ is the (marginal) density of X. In the multivariate case, integrating the joint density over any k of its arguments from $-\infty$ to ∞, we get the (marginal) joint density of the remaining $n-k$ random variables.

The relationship between the p.d.f. and the d.f. is analogous to the one for a single random variable: For a continuous bivariate distribution,

$$F(x,y) = \mathrm{P}(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)dsdt, \tag{5.71}$$

and

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}, \tag{5.72}$$

wherever the derivative on the right exists and is continuous. Similar relations exist for multivariate distributions.

An important class of joint distributions is obtained by generalizing the notion of a uniform distribution on an interval to higher dimensions:

---

12 The same ambiguities arise as in the one-dimensional case. (See footnote 4 on page 115.)

13 That is, a Cartesian product of n intervals, one from each coordinate axis.

Definition 5.4.4. ***Uniform Distribution on Various Regions.*** *Let D be a region of $\mathbb{R}^n$, with n-dimensional volume V. Then the point $(X_1, X_2, \ldots, X_n)$ is said to be chosen at random or uniformly distributed on D, if its distribution is given by the density function*[14]

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} \frac{1}{V} & \text{if } (x_1, x_2, \ldots, x_n) \in D \\ 0 & \text{otherwise} \end{cases}. \tag{5.73}$$

Example 5.4.4. Uniform Distribution on the Unit Square.

Let D be the closed unit square of $\mathbb{R}^2$, that is, $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Then the random point (X, Y) is uniformly distributed on D, if its distribution is given by the density function

$$f(x, y) = \begin{cases} 1 & \text{if } (x, y) \in D \\ 0 & \text{otherwise} \end{cases}. \tag{5.74}$$

Clearly, the marginal densities are the uniform densities on the $[0, 1]$ intervals of the x and y axes, respectively. ♦

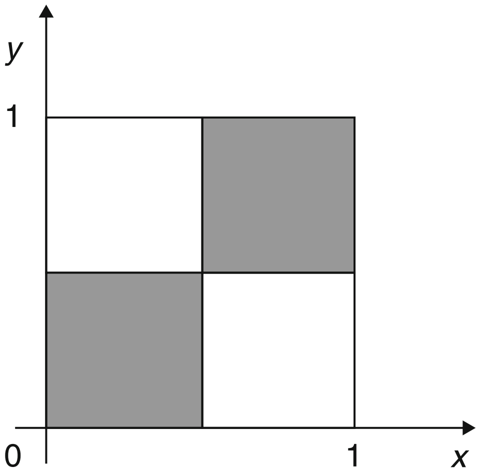

Fig. 5.12. D is the shaded area

Example 5.4.5. Uniform Distribution on Part of the Unit Square.

Let D be the union of the lower left quarter and of the upper right quarter of the unit square of $\mathbb{R}^2$, that is, $D = \{(x, y) : 0 \leq x \leq 1/2, 0 \leq y \leq 1/2\} \cup \{(x, y) : 1/2 \leq x \leq 1, 1/2 \leq y \leq 1\}$ as shown in Figure 5.12.

[14] Note that it makes no difference for this assignment of probabilities whether we consider the region D open or closed or, more generally, whether we include or omit any set of points of dimension less than n.

Then, clearly, the area of D is $1/2$, and so the density function of a random point (X, Y), uniformly distributed on D, is given by

$$f(x,y) = \begin{cases} 2 \text{ if } (x,y) \in D \\ 0 \text{ otherwise} \end{cases}. \tag{5.75}$$

The surprising thing about this distribution is that the marginal densities are again the uniform densities on the $[0, 1]$ intervals of the x and y axes, just as in the previous example, although the joint density is very different and not even continuous on the unit square. ♦

Example 5.4.6. Uniform Distribution on a Diagonal of the Unit Square.

Let D be again the unit square of $\mathbb{R}^2$, that is, $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$, and let the random point (X, Y) be uniformly distributed on the diagonal $y = x$ between the vertices $(0, 0)$ and $(1, 1)$, that is, on the line segment $L = \{(x, y) : y = x, 0 \leq x \leq 1\}$. In other words, assign probabilities to regions A in the plane by

$$\mathrm{P}((X,Y) \in A) = \frac{\text{length}(A \cap L)}{\sqrt{2}}. \tag{5.76}$$

Clearly, here again, the marginal densities are the uniform densities on the $[0, 1]$ intervals of the x and y axes, respectively. Note, however, that X and Y are not jointly continuous (nor discrete) and do not have a joint density function, in spite of X and Y being continuous separately. ♦

Example 5.4.7. Uniform Distribution on the Unit Disk.

Let D be the open unit disk of $\mathbb{R}^2$, that is, $D = \{(x, y) : x^2+y^2 < 1\}$. Then the random point (X, Y) is uniformly distributed on D, if its distribution is given by the density function

$$f(x,y) = \begin{cases} 1/\pi \text{ if } (x,y) \in D \\ 0 \quad\;\; \text{otherwise} \end{cases}. \tag{5.77}$$

The marginal density of X is obtained from its definition $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$. Now, for any fixed $x \in (-1, 1)$, $f(x, y) \neq 0$ if and only if $-\sqrt{1 - x^2} < y < \sqrt{1 - x^2}$, and so for such x

$$\int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi}\sqrt{1-x^2}. \tag{5.78}$$

Thus,

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2} & \text{if } x \in (-1,1) \\ 0 & \text{otherwise} \end{cases}. \tag{5.79}$$

By symmetry, the marginal density of Y is the same, just with x replaced by y :

$$f_Y(y) = \begin{cases} \frac{2}{\pi}\sqrt{1-y^2} & \text{if } y \in (-1,1) \\ 0 & \text{otherwise} \end{cases}. \tag{5.80}$$

♦

Frequently, as for single random variables, we know the general form of a joint distribution except for an unknown coefficient, which we determine from the requirement that the total probability must be 1.

Example 5.4.8. A Distribution on a Triangle.

Let D be the triangle in $\mathbb{R}^2$ given by $D = \{(x,y) : 0 < x, 0 < y, x+y < 1\}$, and let (X,Y) have the density function

$$f(x,y) = \begin{cases} Cxy^2 & \text{if } (x,y) \in D \\ 0 & \text{otherwise} \end{cases}. \tag{5.81}$$

Find the value of C and compute the probability $\mathrm{P}(X < Y)$.

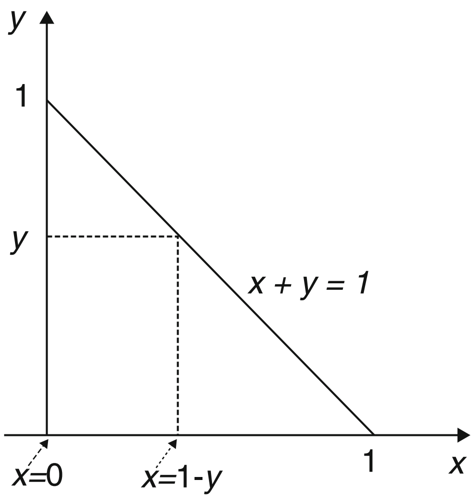

Fig. 5.13. The range of x for a given y

Then, by Figure 5.13,

$$
\begin{aligned}
1 &= \iint_{\mathbb{R}^2} f(x,y)dxdy = \iint_D Cxy^2 dxdy = \int_0^1 \int_0^{1-y} Cxy^2 dxdy \\
&= C\int_0^1 \frac{1}{2}(1-y)^2 y^2 dy = C\int_0^1 \frac{1}{2}(y^2 - 2y^3 + y^4)dy \\
&= C\frac{1}{2}\left(\frac{1}{3} - \frac{1}{2} + \frac{1}{5}\right) = \frac{C}{60}. \qquad (5.82)
\end{aligned}
$$

Thus $C = 60$.

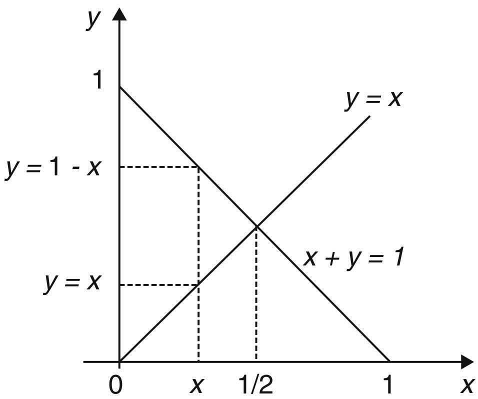

Fig. 5.14. The integration limits for $\mathrm{P}(X < Y)$

To compute the probability $\mathrm{P}(X < Y)$, we have to integrate f over those values (x, y) of (X, Y) for which $x < y$ holds, that is, for the half of the triangle D above the $y = x$ line. (See Figure 5.14.) Thus

$$
\begin{aligned}
\mathrm{P}(X < Y) &= 60\int_0^{1/2} \int_x^{1-x} xy^2 dydx = 60\int_0^{1/2} x\left[\frac{y^3}{3}\right]_x^{1-x} dx \\
&= 20\int_0^{1/2} x\left[(1-x)^3 - x^3\right] dx = 20\int_0^{1/2} \left(x - 3x^2 + 3x^3 - 2x^4\right) dx \\
&= 20\left[\frac{1}{2}\left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^3 + \frac{3}{4}\left(\frac{1}{2}\right)^4 - \frac{2}{5}\left(\frac{1}{2}\right)^5\right] = \frac{11}{16}. \qquad (5.83)
\end{aligned}
$$

♦

The second part of the above example is an instance of the following general principle: If (X, Y) is continuous with joint p.d.f. f and A is any set[15] in $\mathbb{R}^2$, then

$$\mathrm{P}((X,Y) \in A) = \iint_A f(x,y)dxdy. \tag{5.84}$$

In particular, if the set A is defined by a function g so that $A = \{(x, y) : g(x, y) \leq a\}$, for some constant a, then

$$\mathrm{P}(g(X,Y) \leq a) = \iint_{\{g(x,y)\leq a\}} f(x,y)dxdy. \tag{5.85}$$

Relations similar to Equations 5.84 and 5.85 hold for discrete random variables as well; we just have to replace the integrals by sums.

Equation 5.85 shows how to obtain the d.f. of a new random variable $Z = g(X, Y)$.This is illustrated in the following example.

Example 5.4.9. Distribution of the Sum of the Coordinates of a Point.

Let the random point (X, Y) be uniformly distributed on the unit square $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$, as in Example 5.4.4. Find the d.f. of $Z = X + Y$.

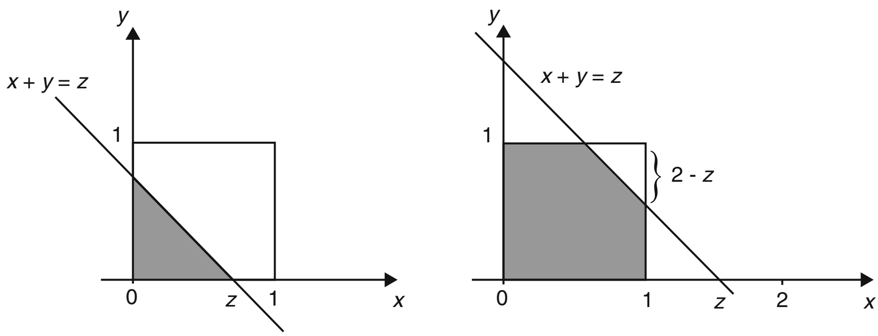

Fig. 5.15. The region $\{x + y \leq z\} \cap D$, depending on the value of z

[15] More precisely, A is any set in $\mathbb{R}^2$ such that $\{s : (X(s), Y(s)) \in A\}$ is an event.

By Equation 5.85 (see Figure 5.15),

$$F_Z(z) = \mathrm{P}(X+Y \le z) = \iint_{\{x+y\le z\}} f(x,y)dxdy = \iint_{\{x+y\le z\}\cap D} dxdy$$
$$= \text{Area of } D \text{ under the line } x+y=z$$
$$= \begin{cases} 0 & \text{if } z<0 \\ \frac{z^2}{2} & \text{if } 0\le z<1 \\ 1-\frac{(2-z)^2}{2} & \text{if } 1\le z<2 \\ 1 & \text{if } 2\le z \end{cases}. \tag{5.86}$$

and so the p.d.f. of Z is

$$f_Z(z) = F_Z'(z) = \begin{cases} 0 & \text{if } z<0 \\ z & \text{if } 0\le z<1 \\ 2-z & \text{if } 1\le z<2 \\ 0 & \text{if } 2\le z. \end{cases} \tag{5.87}$$

♦

The method of the foregoing example can be generalized as follows:

Theorem 5.4.2. ***Distribution of the Sum of Two Random Variables.*** *If X and Y are continuous with joint density f, then the d.f. and the density of $Z = X+Y$ are given by*

$$F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydx = \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f(x,y)dxdy \tag{5.88}$$

and

$$f_Z(z) = \int_{-\infty}^{\infty} f(x,z-x)dx = \int_{-\infty}^{\infty} f(z-y,y)dy. \tag{5.89}$$

If X and Y are discrete with joint p.f. f, then the p.f. of $Z = X+Y$ is given by

$$f_Z(z) = \sum_{x=-\infty}^{\infty} f(x,z-x) = \sum_{y=-\infty}^{\infty} f(z-y,y). \tag{5.90}$$

Proof. In the continuous case,

$$F_Z(z) = \mathrm{P}(X+Y\le Z) = \int\int_{x+y\le z} f(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydx \tag{5.91}$$

and, by the fundamental theorem of calculus,

$$f_Z(z) = F'_Z(z) = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-x} f(x,y)dy \right) dx = \int_{-\infty}^{\infty} f(x, z-x)dx. \tag{5.92}$$

In the discrete case,

$$f_Z(z) = \sum_{x+y=z} f(x,y) = \sum_{x=-\infty}^{\infty} f(x, z-x)dx. \tag{5.93}$$

In each formula, the second form can be obtained by interchanging the roles of x and y. ■

Exercises

Exercise 5.4.1.

Roll two dice as in Example 5.4.1. Find the joint probability function of $U = X + Y$ and $V = X - Y$.

Exercise 5.4.2.

Roll two dice as in Example 5.4.1. Find the joint probability function of $U = \max(X, Y)$ and $V = \min(X, Y)$.

Exercise 5.4.3.

Roll six dice. Find the probabilities of obtaining:

1. Each of the six possible numbers once,
2. One 1, two 2's, and three 3's.

Exercise 5.4.4.

Let the random point (X, Y) be uniformly distributed on the triangle $D = \{(x, y) : 0 \le x \le y \le 1\}$. Find the marginal densities of X and Y and plot their graphs.

Exercise 5.4.5.

Let the random point (X, Y) be uniformly distributed on the unit disk $D = \{(x, y) : x^2 + y^2 < 1\}$. Find the d.f. and the p.d.f. of the point's distance $Z = \sqrt{X^2 + Y^2}$ from the origin.

Exercise 5.4.6.

Let (X, Y) be continuous with density $f(x, y) = Ce^{-x-2y}$ for $x \geq 0, y \geq 0$ and 0 otherwise. Find:

1. The value of the constant C,
2. The marginal densities of X and Y,
3. The joint d.f. $F(x, y)$,
4. $\mathrm{P}(X < Y)$.

Exercise 5.4.7.

Let (X, Y) be continuous with density $f(x, y) = Cxy^2$ on the triangle $D = \{(x, y) : 0 \leq x \leq y \leq 1\}$ and 0 otherwise. Find:

1. The value of the constant C,
2. The marginal densities of X and Y,
3. The joint d.f. $F(x, y)$,
4. $\mathrm{P}(X > Y^2)$.

Exercise 5.4.8.

Let the random point (X, Y) be uniformly distributed on the square $D = \{(x, y) : -1 \leq x \leq 1, -1 \leq y \leq 1\}$. Find the d.f. and the p.d.f. of $Z = X + Y$.

Exercise 5.4.9.

Show that, for any random variables X and Y and any real numbers $x_1 < x_2$ and $y_1 < y_2$,

$$\mathrm{P}(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) + F(x_1, y_1) - F(x_2, y_1).$$

Exercise 5.4.10.

Let the random point (X, Y) be uniformly distributed on the unit square $D = \{(x, y) : 0 \leq x \leq 1, \ 0 \leq y \leq 1\}$. Find the d.f. and the density of $Z = X - Y$.

Exercise 5.4.11.

Find formulas analogous to those in Theorem 5.4.2 for:

1. $Z = X - Y$,
2. $Z = 2X - Y$,
3. $Z = XY$.

5.5 Independence of Random Variables

The notion of independence of events can easily be extended to random variables, by applying the product rule to their joint distributions.

Definition 5.5.1. ***Independence of Two Random Variables.***

Two random variables X and Y are said to be independent of each other if, for all intervals A and B,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \tag{5.94}$$

Equivalently, we can reformulate the defining condition in terms of F or f:

Theorem 5.5.1. ***Alternative Conditions for Independence of Two Random Variables.*** *Two random variables X and Y are independent of each other if and only if their joint d.f. is the product of their marginal d.f.'s:*

$$F(x, y) = F_X(x)F_Y(y) \text{ for all } x, y. \tag{5.95}$$

Two discrete or absolutely continuous random variables X and Y are independent of each other if and only if their joint p.f. or p.d.f. is the product of their marginal p.f.'s or p.d.f.'s[16]*:*

$$f(x, y) = f_X(x)f_Y(y) \text{ for all } x, y. \tag{5.96}$$

Proof. If in Definition 5.5.1 we choose $A = (-\infty, x]$ and $B = (-\infty, y]$, then we get Equation 5.95. Conversely, if Equation 5.95 holds, then Equation 5.94 follows for any intervals from Theorem 5.1.2.

For discrete variables, Equation 5.96 follows from Definition 5.5.1 by substituting the one-point intervals $A = [x, x]$ and $B = [y, y]$, and for continuous variables by differentiating Equation 5.95. Conversely, we can obtain Equation 5.95 from Equation 5.96 by summation or integration. ■

Example 5.5.1. Two Discrete Examples.

In Example 5.4.2 we obtained the following table (Table 5.5) for the joint p.f. f and the marginals of two discrete random variables X and Y:

$y\backslash x$	3	4	Any x
1	1/4	1/2	3/4
2	0	1/4	1/4
Any y	1/4	3/4	1

Table 5.5. The joint p.f. and marginals of two discrete dependent random variables

[16] More precisely, two absolutely continuous r.v.'s are independent if and only if there exist versions of the densities for which Equation 5.96 holds. (See footnote 4 on page 115.)

These variables are *not independent*, because $f(x, y) \neq f_X(x)f_Y(y)$ for all x, y. For instance, $f(3,1) = \frac{1}{4}$ but $f_X(3)f_Y(1) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$. (Note that we need to establish only one instance of $f(x, y) \neq f_X(x)f_Y(y)$ to *disprove* independence, but to *prove* independence we need to show $f(x, y) = f_X(x)f_Y(y)$ for all x, y.)

We can easily construct a table for an f with the same x, y values and the same marginals that represents the distribution of *independent* X and Y. All we have to do is to make each entry $f(x, y)$ equal to the product of the corresponding numbers on the margins (Table 5.6):

$y \backslash x$	3	4	Any x
1	3/16	9/16	3/4
2	1/16	3/16	1/4
Any y	1/4	3/4	1

Table 5.6. The joint p.f. and marginals of two discrete independent random variables

These examples show that there are usually many possible joint distributions for given marginals, but only one of those represents independent random variables. ♦

Example 5.5.2. Independent Uniform Random Variables.

Let the random point (X, Y) be uniformly distributed on the rectangle $D = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$. Then

$$f(x,y) = \begin{cases} \frac{1}{(b-a)(d-c)} & \text{if } (x,y) \in D \\ 0 & \text{otherwise} \end{cases} \tag{5.97}$$

and the marginal densities are obtained by integration as

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \begin{cases} \int_c^d \frac{dy}{(b-a)(d-c)} = \frac{1}{(b-a)} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{5.98}$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \begin{cases} \int_a^b \frac{dx}{(b-a)(d-c)} = \frac{1}{(d-c)} & \text{if } c \leq y \leq d \\ 0 & \text{otherwise} \end{cases}. \tag{5.99}$$

Hence X and Y are uniformly distributed on their respective intervals and are independent, because $f(x, y) = f_X(x)f_Y(y)$ for all x, y, as the preceding formulas show.

Clearly, the converse of our result is also true: If X and Y are uniformly distributed on their respective intervals and are independent, then $f_X(x)f_Y(y)$ yields the p.d.f. 5.97 of a point (X, Y) uniformly distributed on the corresponding rectangle. ♦

Example 5.5.3. Uniform (X, Y) ***on the Unit Disk.***

Let the random point (X, Y) be uniformly distributed on the unit disk $D = \{(x, y) : x^2 + y^2 < 1\}$. In Example 5.4 we obtained

$$f(x, y) = \begin{cases} 1/\pi & \text{if } (x, y) \in D \\ 0 & \text{otherwise} \end{cases}, \tag{5.100}$$

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1 - x^2} & \text{if } x \in (-1, 1) \\ 0 & \text{otherwise} \end{cases} \tag{5.101}$$

and

$$f_Y(y) = \begin{cases} \frac{2}{\pi}\sqrt{1 - y^2} & \text{if } y \in (-1, 1) \\ 0 & \text{otherwise} \end{cases}. \tag{5.102}$$

Now, clearly, $f(x, y) \neq f_X(x)f_Y(y)$ for all x, y, and so X and Y are not independent.

Note that this result is in agreement with the nontechnical meaning of dependence: From the shape of the disk, it follows that some values of X more or less determine the corresponding values of Y (and vice versa). For instance, if X is close to ± 1, then Y must be close to 0, and so X and Y are not expected to be independent of each other. ♦

Example 5.5.4. Constructing a Triangle.

Suppose we pick two random points X and Y independently and uniformly on the interval $[0, 1]$. What is the probability that we can construct a triangle from the resulting three segments as its sides?

A triangle can be constructed if and only if the sum of any two sides is longer than the third side. In our case, this condition means that each side must be shorter than $\frac{1}{2}$. (Prove this!) Thus X and Y must satisfy either

$$0 < X < \frac{1}{2}, \quad 0 < Y - X < \frac{1}{2}, \quad \frac{1}{2} < Y < 1, \tag{5.103}$$

or

$$0 < Y < \frac{1}{2}, \quad 0 < X - Y < \frac{1}{2}, \quad \frac{1}{2} < X < 1. \tag{5.104}$$

By Example 5.5.2 the given selection of the two points X and Y on a line is equivalent to the selection of the single point (X, Y) with a uniform distribution on the unit square of the plane. Now, the two sets of inequalities describe the two triangles at the center, shown in Figure 5.16, and the required probability is their combined area: $\frac{1}{4}$.

♦

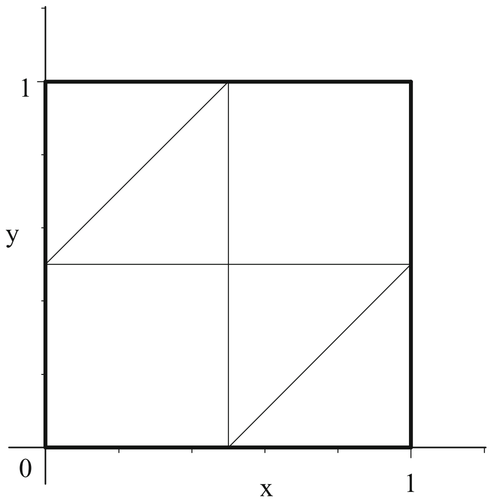

Fig. 5.16.

Example 5.5.5. Buffon's Needle Problem.

In 1777 a French scientist Comte de Buffon published the following problem: Suppose a needle is thrown at random on a floor marked with equidistant parallel lines. What is the probability that the needle will cross one of the lines?

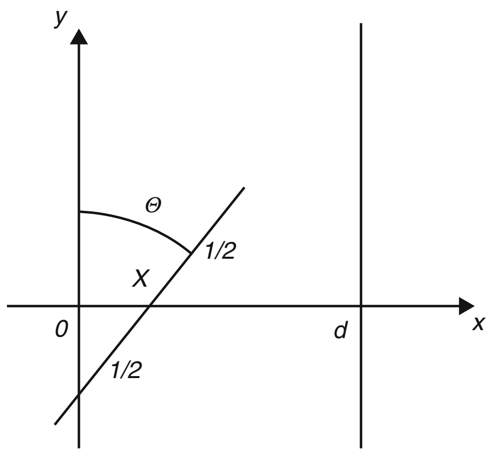

Fig. 5.17.

Let the distance between the lines be d and the length of the needle l. (See Fig. 5.17.) Choose a coordinate system in which the lines are vertical and one of the lines is the y-axis. Let the center of the needle have the random coordinates (X, Y). Clearly, the Y-coordinate is irrelevant to the problem, and we may assume that the center lies on the x-axis. Because of the periodicity, we may also assume that the center falls in the first strip, that is, that $0 \leq X \leq d$. Now, if the needle makes a random angle Θ with the y-axis, then it will cross the y-axis if and only if $0 \leq X \leq (l/2)\sin\Theta$, and it will cross the $y = d$ line if and only if $d - (l/2)\sin\Theta \leq X \leq d$. The random throw of the needle implies that X and Θ are uniform r.v.'s on the $[0, d]$ and the $[0, \pi]$ intervals, respectively, which is, by Example 5.5.2, equivalent to the random point (Θ, X) being uniform on the $[0, \pi] \times [0, d]$ rectangle in the (θ, x) plane. The needle will cross one of the lines if and only if the random point (Θ, X) falls into either one of the D-shaped regions in Figure 5.18. Since the area of the rectangle is πd and the area of each D-shaped region is $\int_0^\pi (l/2)\sin\theta d\theta = l$, the required probability is

$$\mathrm{P}(\text{the needle will cross a line}) = \frac{2l}{\pi d}. \tag{5.105}$$

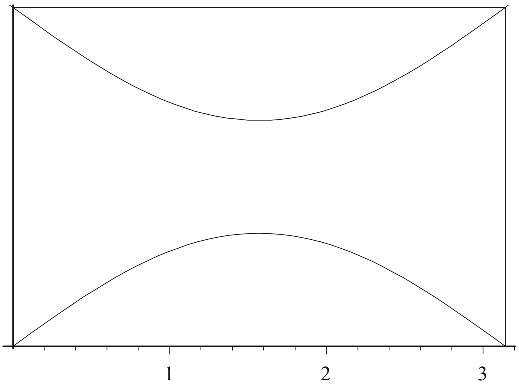

Fig. 5.18. The (θ, x) plane, with the $[0, \pi] \times [0, d]$ rectangle and the $y = (l/2)\sin\theta$ and $y = d - (l/2)\sin\theta$ curves

In principle, this experiment can be used for the determination of π. However, it is difficult to arrange completely random throws of a needle, and it would be a very inefficient way to obtain π; we have much better methods. On the other hand, computer simulations of needle throws have been performed to obtain approximations to π, just to illustrate the result.♦

Example 5.5.6. Bertrand's Paradox.

In 1889, another Frenchman, Joseph Bertrand, considered the following problem:

Consider a circle of radius r and select one of its chords at random. What is the probability that the selected chord will be longer than the side of an equilateral triangle inscribed into the given circle.

Bertrand presented three solutions with "paradoxically" different results:

1. Since the length of a chord is uniquely determined by the position of its center, we just choose a point for it at random, that is, with a uniform distribution inside the circle. We can see from Figure 5.19 that the chord will be longer than the side of an inscribed equilateral triangle if and only if its center falls inside the circle of radius $r/2$ concentric with the given circle. Thus

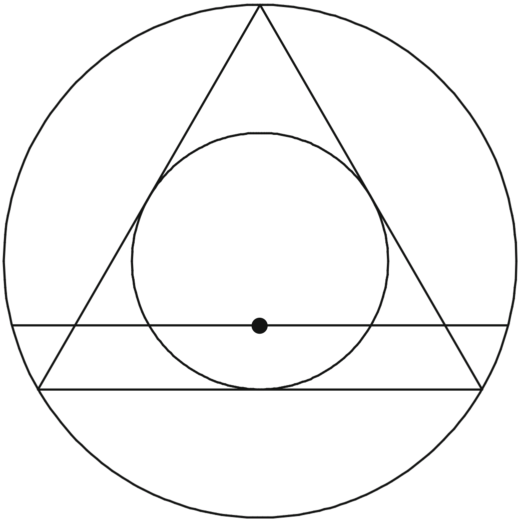

Fig. 5.19.

$$\mathrm{P}(\text{the chord is longer than the side of the triangle}) = \frac{\pi\,(r/2)^2}{\pi r^2} = \frac{1}{4}. \tag{5.106}$$

2. By symmetry, we may consider only horizontal chords, and then we may assume that their center is uniformly distributed on the vertical diameter of the given circle. The chord will be longer than the side of the triangle if and only if its center falls on the thick vertical segment in Figure 5.20. Thus

$$\mathrm{P}(\text{the chord is longer than the side of the triangle}) = \frac{1}{2}. \tag{5.107}$$

3. We may also choose a random chord by fixing one of its endpoints and choosing the other one at random on the circle, that is, uniformly distributed on the perimeter. Let the fixed point be on top, as shown in Figure 5.21. Clearly, the chord will be longer than the side of the triangle

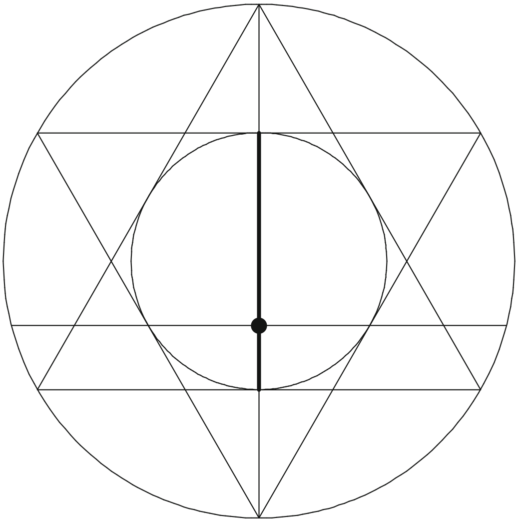

Fig. 5.20.

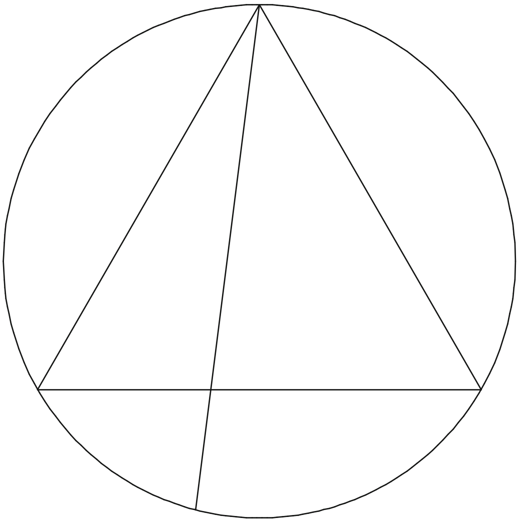

Fig. 5.21.

if and only if its random end falls on the bottom one third of the circle. Thus

$$\mathrm{P}(\text{the chord is longer than the side of the triangle}) = \frac{1}{3}. \tag{5.108}$$

The resolution of the paradox lies in realizing that the statement of the problem is ambiguous. Choosing a chord at random is not well defined; each of the three choices presented above is a reasonable but different way of doing so. ♦

Next, we present some theorems about independence of random variables.

Theorem 5.5.2. ***A Constant Is Independent of Any Random Variable.*** *Let $X = c$, where c is any constant, and let Y be any r.v. Then X and Y are independent.*

Proof. Let $X = c$, and let Y be any r.v. Then Equation 5.95 becomes

$$\mathrm{P}(c \le x, Y \le y) = \mathrm{P}(c \le x)\mathrm{P}(Y \le y), \tag{5.109}$$

and this equation is true because for $x \ge c$ and any y, it reduces to $\mathrm{P}(Y \le y) = \mathrm{P}(Y \le y)$, and for $x < c$ it reduces to $0 = 0$. ■

Theorem 5.5.3. ***No Nonconstant Random Variable Is Independent of Itself.*** *Let X be any nonconstant random variable and let $Y = X$. Then X and Y are dependent.*

Proof. Let A and B be two disjoint intervals for which $\mathrm{P}(X \in A) > 0$ and $\mathrm{P}(X \in B) > 0$ hold. Since X is not constant, such intervals clearly exist. If $Y = X$, then $\mathrm{P}(X \in A, Y \in B) = 0$, but $\mathrm{P}(X \in A)\mathrm{P}(Y \in B) > 0$, and so Equation 5.94 does not hold for all intervals A and B. ■

Theorem 5.5.4. ***Independence of Functions of Random Variables.*** *Let X and Y be independent random variables, and let g and h be any real-valued measurable functions (see the footnote on page 105) on range(X) and range(Y), respectively. Then $g(X)$ and $h(Y)$ are independent.*

Proof. We give the proof for discrete X and Y only.

Let A and B be arbitrary intervals. Then

$$\begin{aligned}
\mathrm{P}(g(X) \in A, h(Y) \in B) &= \sum_{\{x:g(x)\in A\}} \sum_{\{y:h(y)\in B\}} \mathrm{P}(X = x, Y = y) \\
&= \sum_{\{x:g(x)\in A\}} \sum_{\{y:h(y)\in B\}} \mathrm{P}(X = x)\mathrm{P}(Y = y) \\
&= \sum_{\{x:g(x)\in A\}} \mathrm{P}(X = x) \sum_{\{y:h(y)\in B\}} \mathrm{P}(Y = y) \\
&= \mathrm{P}(g(X) \in A)\mathrm{P}(h(Y) \in B).
\end{aligned} \tag{5.110}$$

■

We can extend the definition of independence to several random variables as well, but we need to distinguish different types of independence, depending on the number of variables involved:

Definition 5.5.2. ***Independence of Several Random Variables.***

Let $X_1, X_2, \ldots, X_n$, for $n = 2, 3, \ldots$, be arbitrary random variables. They are (totally) independent, if

$$\mathrm{P}(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \mathrm{P}(X_1 \in A_1)\mathrm{P}(X_2 \in A_2) \cdots \mathrm{P}(X_n \in A_n) \tag{5.111}$$

for all intervals $A_1, A_2, \ldots, A_n$.

They are pairwise independent if

$$\mathrm{P}(X_i \in A_i, X_j \in A_j) = \mathrm{P}(X_i \in A_i)\mathrm{P}(X_j \in A_j) \tag{5.112}$$

for all $i \neq j$ and all intervals A_i, A_j.

Note that in the case of total independence, it is not necessary to require the product rule for all subsets of the n random variables (as we had to for general events), because the product rule for any number less than n follows from Equation 5.111 by setting $A_i = \mathbb{R}$ for all values of i that we want to omit. On the other hand, pairwise independence is a weaker requirement than total independence: Equation 5.112 does not imply Equation 5.111. Also, we could have defined various types of independence between total and pairwise, but such types generally do not occur in practice.

We have the following theorems for several random variables, analogous to Theorem 5.5.1 and Theorem 5.5.4, which we state without proof.

Theorem 5.5.5. ***Alternative Conditions for Independence of Several Random Variables.*** *Any random variables $X_1, X_2, \ldots, X_n$, for $n = 2, 3, \ldots$, are independent of each other if and only if their joint d.f. is the product of their marginal d.f.'s:*

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n) \text{ for all } x_1, x_2, \ldots, x_n. \tag{5.113}$$

Also, any discrete or absolutely continuous random variables $X_1, X_2, \ldots, X_n$, for $n = 2, 3, \ldots$, are independent of each other if and only if their joint p.f. or p.d.f. is the product of their marginal p.f.'s or (appropriate versions of their) p.d.f.'s:

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n) \text{ for all } x_1, x_2, \ldots, x_n. \tag{5.114}$$

Theorem 5.5.6. ***Independence of Functions of Random Variables.*** *Let $X_1, X_2, \ldots, X_n$, for $n = 2, 3, \ldots$, be independent random variables, and let the g_i be real-valued measurable functions on range(X_i) for $i = 1, 2, \ldots, n$. Then $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ are independent.*

Theorem 5.5.6 could be further generalized in an obvious way by taking the g_i to be functions of several, non-overlapping variables. For example, in the case of three random variables, we have the following theorem:

Theorem 5.5.7. ***Independence of*** $g(X,Y)$ ***and*** Z. *If* Z *is independent of* (X,Y)*, then* Z *is independent of* $g(X,Y)$*, too, for any measurable function* g.

Proof. We give the proof for jointly continuous X, Y, and Z only.

For arbitrary t and z,

$$\begin{aligned}
\mathrm{P}(g(X,Y) \le t, Z \le z) &= \int_{-\infty}^{z} \iint_{g(x,y)\le t} f(x,y,\varsigma)\, dx dy d\varsigma \\
&= \int_{-\infty}^{z} \iint_{g(x,y)\le t} f_{X,Y}(x,y)\, f_Z(\varsigma)\, dx dy d\varsigma \\
&= \iint_{g(x,y)\le t} f_{X,Y}(x,y)\, dx dy \int_{-\infty}^{z} f_Z(\varsigma)\, d\varsigma \\
&= \mathrm{P}(g(X,Y) \le t)\, \mathrm{P}(Z \le z). \qquad (5.115)
\end{aligned}$$

By Theorem 5.5.1, Equation 5.115 proves the independence of $g(X,Y)$ and Z. ■

In some applications, we need to find the distribution of the maximum or of the minimum of several independent random variables. This can be done as follows:

Theorem 5.5.8. ***Distribution of Maximum and Minimum of Several Random Variables.*** *Let* $X_1, X_2, \ldots, X_n$*, for* $n = 2, 3, \ldots,$ *be independent, identically distributed (abbreviated i.i.d.) random variables with common d.f.* F_X*, and let* $Y = \max\{X_1, X_2, \ldots, X_n\}$ *and* $Z = \min\{X_1, X_2, \ldots, X_n\}$.[17] *Then the distribution functions of* Y *and* Z *are given by*

$$F_Y(y) = [F_X(y)]^n \text{ for all } y \in \mathbb{R} \qquad (5.116)$$

and

$$F_Z(z) = 1 - [1 - F_X(z)]^n \text{ for all } z \in \mathbb{R}. \qquad (5.117)$$

Proof. For any $y \in \mathbb{R}$, $Y = \max\{X_1, X_2, \ldots, X_n\} \le y$ holds if and only if, for every i, $X_i \le y$. Thus, we have

$$\begin{aligned}
F_Y(y) &= \mathrm{P}(X_1 \le y, X_2 \le y, \ldots, X_n \le y) \\
&= \mathrm{P}(X_1 \le y)\, \mathrm{P}(X_2 \le y) \cdots \mathrm{P}(X_n \le y) = [F_X(y)]^n. \qquad (5.118)
\end{aligned}$$

[17] Note that the max and the min have to be taken pointwise, that is, for each sample point s, we have to consider the max and the min of $\{X_1(s), X_2(s), \ldots, X_n(s)\}$, and so Y and Z will in general be different from each of the X_i.

Similarly,

$$\begin{aligned}F_Z(z) &= \mathrm{P}\left(Z \leq z\right) = 1 - \mathrm{P}\left(Z > z\right) \\ &= 1 - \mathrm{P}\left(X_1 > z, X_2 > z, \ldots, X_n > z\right) \\ &= 1 - \mathrm{P}\left(X_1 > z\right)\mathrm{P}\left(X_2 > z\right)\cdots\mathrm{P}\left(X_n > z\right) \\ &= 1 - \left[1 - F_X(z)\right]^n . \end{aligned} \tag{5.119}$$

■

Example 5.5.7. Maximum of Two Independent Uniformly Distributed Points.

Let X_1 and X_2 be independent, uniform random variables on the interval $[0,1]$. Find the d.f. and the p.d.f. of $Y = \max\{X_1, X_2\}$.

By Equation 5.19,

$$F_X\left(x\right) = \begin{cases} 0 \text{ if } & x < 0 \\ x \text{ if } 0 \leq & x < 1 \\ 1 \text{ if } & x \geq 1 \end{cases}, \tag{5.120}$$

and so, by Theorem 5.5.8,

$$F_Y\left(y\right) = \begin{cases} 0 \text{ if } & y < 0 \\ y^2 \text{ if } 0 \leq & y < 1 \\ 1 \text{ if } & y \geq 1 \end{cases}. \tag{5.121}$$

Hence the p.d.f. of Y is given by

$$f_Y\left(y\right) = \begin{cases} 2y \text{ if } & 0 \leq y < 1 \\ 0 \text{ if } & y < 0 \text{ or } y \geq 1 \end{cases}, \tag{5.122}$$

which shows that the probability of $Y = \max\{X_1, X_2\}$ falling in a subinterval of length dy is no longer constant over $[0,1]$, as for X_1 or X_2, but increases linearly.

The two functions above can also be seen in Figure 5.22. The sample space is the set of points $s = (x_1, x_2)$ of the unit square, and, for any sample point s, $X_1(s) = x_1$ and $X_2(s) = x_2$. The sample points are uniformly distributed on the unit square, and so the areas of subsets give the corresponding probabilities. Since for any sample point s above the diagonal $x_1 < x_2$ holds, $Y(s) = x_2$ there and, similarly, below the diagonal $Y(s) = x_1$. Thus, the set $\{s : Y(s) \leq y\}$ is the shaded square of area y^2, and the thin strip of width dy, to the right and above the square, has an area $\approx 2ydy$. ♦

Another, very important function of two independent random variables is their sum. We have the following theorem for its distribution:

Theorem 5.5.9. ***Sum of Two Independent Random Variables.*** *Let X and Y be independent random variables and $Z = X + Y$. If X and Y are discrete, then the p.f. of Z is given by*

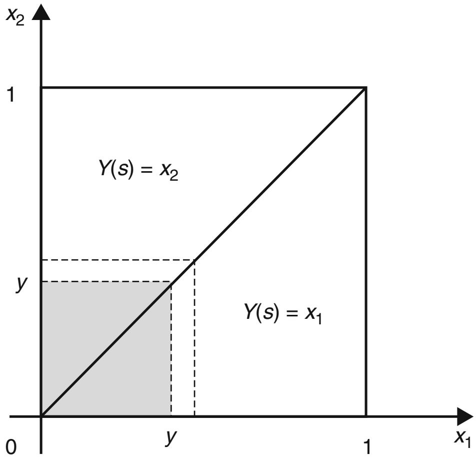

Fig. 5.22. The d.f and the p.d.f. of $Y = \max\{X_1, X_2\}$ for two i.i.d. uniform r.v.'s on $[0, 1]$

$$f_Z(z) = \sum_{x+y=z} f_X(x) f_Y(y) = \sum_x f_X(x) f_Y(z-x), \tag{5.123}$$

where, for a given z, the summation is extended over all possible values of X and Y for which $x + y = z$, if such values exist. Otherwise $f_Z(z)$ is taken to be 0. The expression on the right is called the convolution of f_X and f_Y.

If X and Y are continuous with densities f_X and f_Y, then the density of $Z = X + Y$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx, \tag{5.124}$$

where the integral is again called the convolution of f_X and f_Y.

Proof. These results follow from Theorem 5.4.2 by substituting $f(x, y) = f_X(x) f_Y(y)$. However, in the continuous case, we also give another, more direct and visual proof: If X and Y are independent and continuous with densities f_X and f_Y, then Z falls between z and $z + dz$ if and only if the point (X, Y) falls in the oblique strip between the lines $x + y = z$ and $x + y = z + dz$, as shown in Figure 5.23. The area of the shaded parallelogram is $dxdz$, and the probability of (X, Y) falling into it is[18]

$$\mathrm{P}(x \le X < x + dx, z \le Z < z + dz) \sim f(x, y) dxdz = f_X(x) f_Y(z-x) dxdz. \tag{5.125}$$

[18] Recall that the symbol $\sim$ means that the ratio of the expressions on either side of it tends to 1 as dx and dz tend to 0.

Hence the probability of the strip is obtained by integrating this expression over all x as

$$\mathrm{P}\left(z \leq Z < z + dz\right) \sim \left[\int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx\right] dz, \tag{5.126}$$

and, since $\mathrm{P}(z \leq Z < z + dz) \sim f_Z(z)\, dz$, Equation 5.126 implies Equation 5.124. ■

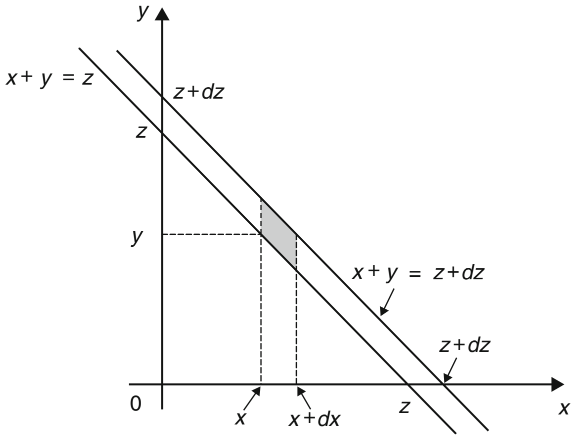

Fig. 5.23. The probability of (X, Y) falling in the oblique strip is dz times the convolution

The convolution formulas for two special classes of random variables are worth mentioning separately:

Corollary 5.5.1. *If the possible values of X and Y are the natural numbers $i, j = 0, 1, 2, \ldots,$ then the p.f. of $Z = X + Y$ is given by*

$$f_Z(k) = \sum_{i=0}^{k} f_X(i) f_Y(k-i) \text{ for } k = 0, 1, 2, \ldots, \tag{5.127}$$

and if X and Y are continuous nonnegative random variables, then the p.d.f. of $Z = X + Y$ is given by

$$f_Z(z) = \int_0^z f_X(x) f_Y(z-x) dx. \tag{5.128}$$

Example 5.5.8. Sum of Two Binomial Random Variables.

Let X and Y be independent, binomial r.v.'s with parameters n_1, p and n_2, p, respectively. Then $Z = X + Y$ is binomial with parameters $n_1 + n_2, p$ because, by Equation 5.127 and Equation 3.37,

$$
\begin{aligned}
f_Z(k) &= \sum_{i=0}^{k} \binom{n_1}{i} p^i q^{n_1-i} \binom{n_2}{k-i} p^{k-i} q^{n_2-k+i} \\
&= \sum_{i=0}^{k} \binom{n_1}{i} \binom{n_2}{k-i} p^k q^{n_1+n_2-k} \\
&= \binom{n_1+n_2}{k} p^k q^{n_1+n_2-k} \text{ for } k = 0, 1, 2, \ldots, n_1+n_2. \qquad (5.129)
\end{aligned}
$$

This result should be obvious without any computation as well, since X counts the number of successes in n_1 independent trials and Y the number of successes in n_2 trials, independent of each other and of the first n_1 trials, and so $Z = X + Y$ counts the number of successes in $n_1 + n_2$ independent trials, all with the same probability p.

On the other hand, for sampling without replacement, the trials are not independent, and the analogous sum of two independent hypergeometric random variables does not turn out to be hypergeometric. ♦

Example 5.5.9. Sum of Two Uniform Random Variables.

Let X and Y be i.i.d. random variables, uniform on $[0,1]$ as in Example 5.4.9. This time, however, we are going to use Equation 5.128 to obtain the density of $Z = X + Y$.

The common density of X and Y is

$$
f(x) = \begin{cases} 1 \text{ if } 0 \le x \le 1 \\ 0 \text{ otherwise} \end{cases} \qquad (5.130)
$$

Hence $f_Y(z-x) = 1$ if $0 \le z - x \le 1$ or, equivalently, if $z - 1 \le x \le z$ and is 0 otherwise. Thus the density of Z is the convolution from Equation 5.128:

$$
f_Z(z) = \int_0^z f_X(x) f_Y(z-x)dx = \int_{[0,1]\cap[z-1,z]} 1dx = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{if } 0 \le z < 1 \\ 2-z & \text{if } 1 \le z < 2 \\ 0 & \text{if } 2 \le z. \end{cases} \qquad (5.131)
$$

This result is the same, of course, as the corresponding one in Example 5.4.9. ♦

Theorem 5.5.10. ***Product and Ratio of Two Independent Random Variables.*** *Let X and Y be independent, continuous, positive random variables with given densities f_X and f_Y, with $f_X(x) = 0$ for $x < 0$ and $f_Y(y) = 0$ for $y < 0$:*

1. *The density function of* $Z = XY$ *is given by*

$$f_Z(z) = \begin{cases} \int_0^\infty f_X(\frac{z}{y}) f_Y(y) \frac{1}{y} dy & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases} \tag{5.132}$$

and alternatively by

$$f_Z(z) = \begin{cases} \int_0^\infty f_X(x) f_Y(\frac{z}{x}) \frac{1}{x} dx & \text{if } z > 0 \\ 0 & \text{if } z \leq 0. \end{cases} \tag{5.133}$$

2. *The density function of* $Z = X/Y$ *is given by*

$$f_Z(z) = \begin{cases} \int_0^\infty f_X(zy) f_Y(y) y dy & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases} \tag{5.134}$$

and alternatively by

$$f_Z(z) = \begin{cases} \int_0^\infty f_X(x) f_Y(\frac{x}{z}) \frac{x}{z^2} dx & \text{if } z > 0 \\ 0 & \text{if } z \leq 0. \end{cases} \tag{5.135}$$

Proof.

1. For $z > 0$

$$\begin{aligned} F_Z(z) &= \mathrm{P}(XY \leq z) = \iint_{xy \leq z} f_X(x) f_Y(y) dx dy \\ &= \int_0^\infty \left[\int_0^{z/y} f_X(x) dx \right] f_Y(y) dy = \int_0^\infty F_X\left(\frac{z}{y}\right) f_Y(y) dy, \end{aligned} \tag{5.136}$$

and so, by the chain rule,

$$f_Z(z) = F'_Z(z) = \int_0^\infty f_X\left(\frac{z}{y}\right) f_Y(y) \frac{1}{y} dy. \tag{5.137}$$

If $z \leq 0$, then, by the assumed positivity of X and Y, $\mathrm{P}(XY \leq z) = 0$ and $f_Z(z) = 0$. The alternative formula can be obtained by interchanging x and y.

2. For $z > 0$

$$\begin{aligned} F_Z(z) &= \mathrm{P}\left(\frac{X}{Y} \leq z\right) = \iint_{x/y \leq z} f_X(x) f_Y(y) dx dy \\ &= \int_0^\infty \left[\int_0^{zy} f_X(x) dx \right] f_Y(y) dy = \int_0^\infty F_X(zy) f_Y(y) dy, \end{aligned} \tag{5.138}$$

and so

$$f_Z(z) = F'_Z(z) = \int_0^\infty f_X(zy) f_Y(y) y dy. \tag{5.139}$$

Alternatively,

$$\begin{aligned} F_Z(z) &= \mathrm{P}\left(\frac{X}{Y} \le z\right) = \iint_{x/y \le z} f_X(x) f_Y(y) dx dy \\ &= \int_0^\infty f_X(x) \left[\int_{x/z}^\infty f_Y(y) dy\right] dx = \int_0^\infty f_X(x) \left[1 - F_Y\left(\frac{x}{z}\right)\right] dx, \end{aligned} \tag{5.140}$$

and so

$$f_Z(z) = F'_Z(z) = \int_0^\infty f_X(x) f_Y\left(\frac{x}{z}\right) \frac{x}{z^2} dx. \tag{5.141}$$

If $z \le 0$, then clearly $\mathrm{P}(X/Y \le z) = 0$ and $f_Z(z) = 0$. ■

Example 5.5.10. Ratio of Two Exponential Random Variables.

Let X and Y be two exponential random variables with unequal parameters λ_1 and λ_2, respectively. Find the density of their ratio $Z = X/Y$.

Now, $f_X(x) = \lambda_1 e^{-\lambda_1 x}$ and $f_Y(y) = \lambda_2 e^{-\lambda_2 y}$ for $x, y > 0$. Thus, from Equation 5.134 and using integration by parts, we obtain the density of $Z = X/Y$, for $z > 0$, as

$$\begin{aligned} f_Z(z) &= \int_0^\infty \lambda_2 e^{-\lambda_2 y} \lambda_1 e^{-\lambda_1 z y} y dy = \lambda_1 \lambda_2 \int_0^\infty e^{-(\lambda_2 + \lambda_1 z) y} y dy \\ &= \lambda_1 \lambda_2 \left[\left.\frac{e^{-(\lambda_2 + \lambda_1 z) y}}{-(\lambda_2 + \lambda_1 z)} y\right|_0^\infty - \int_0^\infty \frac{e^{-(\lambda_2 + \lambda_1 z) y}}{-(\lambda_2 + \lambda_1 z)} dy\right] \\ &= \lambda_1 \lambda_2 \left[0 - \left.\frac{e^{-(\lambda_2 + \lambda_1 z) y}}{(\lambda_2 + \lambda_1 z)^2}\right|_0^\infty\right] = \frac{\lambda_1 \lambda_2}{(\lambda_2 + \lambda_1 z)^2}. \end{aligned} \tag{5.142}$$

Exercises

Exercise 5.5.1.

Two cards are dealt from a regular deck of 52 cards without replacement. Let X denote the number of spades and Y the number of hearts obtained. Are X and Y independent?

Exercise 5.5.2.

We roll two dice once. Let X denote the number of 1's and Y the number of 6's obtained. Are X and Y independent?

Exercise 5.5.3.

Let the random point (X, Y) be uniformly distributed on $D = \{(x, y) : 0 \le x \le 1/2, 0 \le y \le 1/2\} \cup \{(x, y) : 1/2 \le x \le 1, 1/2 \le y \le 1\}$ as in 5.4.5. Are X and Y independent?

Exercise 5.5.4.

Let X and Y be continuous random variables with density

$$f(x, y) = \begin{cases} xe^{-x(y+1)} & \text{if } x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5.143}$$

Are X and Y independent?

Exercise 5.5.5.

Recall that the *indicator function* or *indicator random variable* I_A of an event A in any sample space S is defined by

$$I_A(s) = \begin{cases} 1 \text{ if } s \in A \\ 0 \text{ if } s \in \overline{A}. \end{cases} \tag{5.144}$$

1. Prove that $I_{A \cup B} = I_A + I_B - I_{AB}$.
2. Prove that A and B are independent events if and only if I_A and I_B are independent random variables.

Exercise 5.5.6.

Let the random point (X, Y) be uniformly distributed on the unit disk as in 5.4. Show that the polar coordinates $R \in [0, 1]$ and $\Theta \in [0, 2\pi]$ of the point are independent. (*Hint*: Determine the joint d.f. $F_{R,\Theta}(r, \theta)$ and the marginals $F_R(r) = F_{R,\Theta}(r, 2\pi)$ and $F_\Theta(\theta) = F_{R,\Theta}(1, \theta)$ from a picture, and use Equation 5.95.)

Exercise 5.5.7.

Alice and Bob visit the school library, each at a random time uniformly distributed between 2PM and 6PM, independently of each other, and stay there for an hour. What is the probability that they meet?

Exercise 5.5.8.

A point X is chosen at random on the interval $[0, 1]$ and independently another point Y on the interval $[1, 2]$. What is the probability that we can construct a triangle from the resulting three segments $[0, X], [X, Y], [Y, 2]$ as sides?

Exercise 5.5.9.

We choose a point at random on the perimeter of a circle and then independently another point at random in the interior of the circle. What is the probability that the two points will be nearer to each other than the radius of the circle?

Exercise 5.5.10.

Let X be a discrete uniform r.v. on the set $\{000, 011, 101, 110\}$ of four binary integers, and let X_i denote the ith digit of X, for $i = 1, 2, 3$. Show that X_1, X_2, X_3 are independent pairwise but not totally independent.

Can you generalize this example to more than three random variables?

Exercise 5.5.11.

Let X and Y be i.i.d. uniform on $(0, 1)$:

1. Find the joint density of $Z = XY$.
2. Find the joint density of $Z = X/Y$.

Exercise 5.5.12.

What is the probability that in ten independent tosses of a fair coin, we get two heads in the first four tosses and five heads altogether?

Exercise 5.5.13.

Consider light bulbs with independent, exponentially distributed lifetimes with parameter $\lambda = \frac{1}{100 \text{ days}}$:

1. Find the probability that such a bulb survives to 200 days.
2. Find the probability that such a bulb dies before 40 days.
3. Find the probability that the bulb with the longest lifetime in a batch of 10 survives to 200 days.
4. Find the probability that the bulb with the shortest lifetime in a batch of 10 dies before 40 days.

Exercise 5.5.14.

Let $X_1, X_2, \ldots, X_n$, for $n = 2, 3, \ldots$, be i.i.d. random variables with common d.f. F_X. Find a formula for the joint d.f. $F_{Y,Z}$ of $Y = \max\{X_1, X_2, \ldots, X_n\}$ and $Z = \min\{X_1, X_2, \ldots, X_n\}$ in terms of F_X.

Exercise 5.5.15.

Show that the p.d.f. of the sum $S = T_1 + T_2$ of two i.i.d exponential r.v.'s with parameter λ is given by

$$f_S(s) = \begin{cases} 0 & \text{if } s < 0 \\ \lambda^2 s e^{-\lambda s} & \text{if } s \geq 0 \end{cases} \tag{5.145}$$

Exercise 5.5.16.

Find the p.d.f. of the sum $S = T_1 + T_2$ of two independent exponential r.v.'s with parameters λ and $\mu \neq \lambda$, respectively.

Exercise 5.5.17.

Show that the p.d.f. of the difference $Z = T_1 - T_2$ of two i.i.d exponential r.v.'s with parameter λ is $f_Z(z) = \frac{\lambda}{2} e^{-\lambda|z|}$.

Exercise 5.5.18.

Let X_i for $i = 1, 2, \ldots$ be i.i.d. random variables, uniform on $[0, 1]$, and let f_n denote the p.d.f. of $S_n = \sum_{i=1}^{n} X_i$ for $n \geq 1$:

1. Show that $f_{n+1}(z) = \int_{z-1}^{z} f_n(x)\, dx$.
2. Evaluate $f_3(z)$ and sketch its graph.

Exercise 5.5.19.

Let X and Y be i.i.d. random variables, uniform on $[0, 1]$. Find the density of $Z = X - Y$.

5.6 Conditional Distributions

In many applications, we need to consider the distribution of a random variable under certain conditions. For conditions with nonzero probabilities, we can just apply the definition of conditional probabilities to events associated with random variables. Thus, we make the following definition:

Definition 5.6.1. ***Conditional Distributions for Conditions with Nonzero Probabilities.***

Let A be any event with $\mathrm{P}(A) \neq 0$ and X any random variable. Then we define the conditional distribution function of X under the condition A by

$$F_{X|A}(x) = \mathrm{P}(X \leq x | A) \ \text{ for all } x \in \mathbb{R}. \tag{5.146}$$

If X is a discrete random variable, then we define the conditional probability function of X under the condition A by

$$f_{X|A}(x) = \mathrm{P}(X = x|A) \text{ for all } x \in \mathbb{R}. \tag{5.147}$$

If X is a continuous random variable and there exists a nonnegative function $f_{X|A}$ that is integrable over R and for which

$$\int_{-\infty}^{x} f_{X|A}(t)dt = F_{X|A}(x), \text{ for all } x, \tag{5.148}$$

then $f_{X|A}$ is called the conditional density function of X under the condition A.

If Y is a discrete random variable and $A = \{Y = y\}$, then we write

$$F_{X|Y}(x, y) = \mathrm{P}(X \leq x|Y = y) \text{ for all } x \in \mathbb{R} \text{ and all possible values } y \text{ of } Y \tag{5.149}$$

and call $F_{X|Y}$ the conditional distribution function of X given Y.

If both X and Y are discrete, then the conditional probability function of X given Y is defined by

$$f_{X|Y}(x, y) = \mathrm{P}(X = x|Y = y) \text{ for all possible values } x \text{ and } y \text{ of } X \text{ and } Y. \tag{5.150}$$

If X is continuous, Y is discrete, $A = \{Y = y\}$, and $f_{X|A}$ in Equation 5.148 exists, then $f_{X|A}$ is called the conditional density function of X given $Y = y$ and is denoted by $f_{X|Y}(x, y)$ for all $x \in \mathbb{R}$ and all possible values y of Y.

If X is a continuous random variable with a conditional density function $f_{X|A}$, then, by the fundamental theorem of calculus, Equation 5.148 gives that

$$f_{X|A}(x) = F'_{X|A}(x), \tag{5.151}$$

wherever $F'_{X|A}$ is continuous. At such points, we also have

$$f_{X|A}(x)dx \sim \mathrm{P}(x \leq X < x + dx \mid A) = \frac{\mathrm{P}(\{x \leq X < x + dx\} \cap A)}{\mathrm{P}(A)}. \tag{5.152}$$

By the definitions of conditional probabilities and joint distributions, Equation 5.150 for discrete X and Y can also be written as

$$f_{X|Y}(x, y) = \frac{f(x, y)}{f_Y(y)} \text{ for all possible values } x \text{ and } y \text{ of } X \text{ and } Y, \tag{5.153}$$

where $f(x, y)$ is the joint p.f. of X and Y and $f_Y(y)$ the marginal p.f. of Y.

Example 5.6.1. Sum and Absolute Difference of Two Dice.

In Example 5.4.1 we considered the random variables $U = X + Y$ and $V = |X - Y|$, where X and Y were the numbers obtained with rolling two dice. Now, we want to find the values of the conditional probability functions $f_{U|V}$ and $f_{V|U}$. For easier reference, we first reproduce the table of the joint probability function $f(u, v)$ and the marginals here (Table 5.7):

$u\backslash v$	0	1	2	3	4	5	$f_U(u)$
2	1/36	0	0	0	0	0	1/36
3	0	2/36	0	0	0	0	2/36
4	1/36	0	2/36	0	0	0	3/36
5	0	2/36	0	2/36	0	0	4/36
6	1/36	0	2/36	0	2/36	0	5/36
7	0	2/36	0	2/36	0	2/36	6/36
8	1/36	0	2/36	0	2/36	0	5/36
9	0	2/36	0	2/36	0	0	4/36
10	1/36	0	2/36	0	0	0	3/36
11	0	2/36	0	0	0	0	2/36
12	1/36	0	0	0	0	0	1/36
$f_V(v)$	6/36	10/36	8/36	6/36	4/36	2/36	1

Table 5.7. The joint and marginal probability functions of $U = X + Y$ and $V = |X - Y|$, for the numbers X and Y showing on two dice

According to Equation 5.153, the table of the conditional probability function $f_{U|V}(u, v)$ can be obtained from the table above by dividing each $f(u, v)$ value by the marginal probability below it, and similarly, the table of the conditional probability function $f_{V|U}(u, v)$ can be obtained by dividing each $f(u, v)$ value by the marginal probability to the right of it:

$u\backslash v$	0	1	2	3	4	5
2	1/6	0	0	0	0	0
3	0	1/5	0	0	0	0
4	1/6	0	1/4	0	0	0
5	0	1/5	0	1/3	0	0
6	1/6	0	1/4	0	1/2	0
7	0	1/5	0	1/3	0	1
8	1/6	0	1/4	0	1/2	0
9	0	1/5	0	1/3	0	0
10	1/6	0	1/4	0	0	0
11	0	1/5	0	0	0	0
12	1/6	0	0	0	0	0

Table 5.8. The conditional probability function of $U = X + Y$ given $V = |X - Y|$, for the numbers X and Y showing on two dice

$u\backslash v$	0	1	2	3	4	5
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	1/3	0	2/3	0	0	0
5	0	1/2	0	1/2	0	0
6	1/5	0	2/5	0	2/5	0
7	0	1/3	0	1/3	0	1/3
8	1/5	0	2/5	0	2/5	0
9	0	1/2	0	1/2	0	0
10	1/3	0	2/3	0	0	0
11	0	1	0	0	0	0
12	1	0	0	0	0	0

Table 5.9. The conditional probability function of $V = |X - Y|$ given $U = X + Y$, for the numbers X and Y showing on two dice

The conditional probabilities in these tables make good sense. For instance, if $V = |X - Y| = 1$, then $U = X + Y$ can be only $3 = 1+2 = 2+1, 5 = 2+3 = 3+2, 7 = 3+4 = 4+3, 9 = 4+5 = 5+4$, or $11 = 5+6 = 6+5$. Since each of these five possible U values can occur under the condition $V = 1$ in exactly two ways, their conditional probabilities must be 1/5 each, as shown in the second column of Table 5.8.

Similarly, if $U = X + Y = 3$, then we must have $(X, Y) = (1, 2)$ or $(X, Y) = (2, 1)$, and in either case $V = |X - Y| = 1$. Thus, $f_{V|U}(3, 1) = 1$ as shown for $(u, v) = (3, 1)$ in Table 5.9. ♦

For a continuous random variable Y, $\mathrm{P}(A|Y = y)$ and the conditional density $f_{X|Y}(x, y)$ are undefined because $\mathrm{P}(Y = y) = 0$. Nevertheless we can define $\mathrm{P}(A|Y = y)$ as a limit with Y falling in an infinitesimal interval at y, rather than being equal to y. For $f_{X|Y}(x, y)$ we can use Equation 5.153 as a model, with f and f_Y reinterpreted as densities.

Definition 5.6.2. ***Conditional Probabilities and Densities for Given Values of a Continuous Random Variable.***

For a continuous random variable Y and any event A, we define

$$\mathrm{P}(A|Y = y) = \lim_{h \to 0^+} \mathrm{P}(A|y \le Y < y + h), \tag{5.154}$$

if the limit exists. In particular, if $A = \{X \le x\}$, for any random variable X and any real x, then the conditional p.d.f. of X, given $Y = y$, is defined as

$$F_{X|Y}(x, y) = \lim_{h \to 0^+} \mathrm{P}(X \le x|y \le Y < y + h), \tag{5.155}$$

if the limit exists, and, if X is discrete, then the conditional p.f. of X, given $Y = y$, is defined as

$$f_{X|Y}(x, y) = \lim_{h \to 0^+} \mathrm{P}(X = x|y \le Y < y + h), \tag{5.156}$$

if the limit exists.

Furthermore, for continuous random variables X and Y with joint density $f(x,y)$ and Y having marginal density $f_Y(y)$, we define the conditional density $f_{X|Y}$ by

$$f_{X|Y}(x,y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.157}$$

for all real x and y.

Example 5.6.2. Conditional Density for (X,Y) Uniform on Unit Disk.

Let (X,Y) be uniform on the unit disk $D = \{(x,y) : x^2 + y^2 < 1\}$ as in Example 5.4.7. Hence

$$f_{X|Y}(x,y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{1-y^2}} & \text{if } (x,y) \in D \\ 0 & \text{otherwise} \end{cases}. \tag{5.158}$$

For a fixed $y \in (-1,1)$, this expression is constant over the x-interval $\left(-\sqrt{1-y^2}, \sqrt{1-y^2}\right)$, and therefore, not unexpectedly, it is the density of the uniform distribution over that interval. ♦

Note that $f_{X|Y}$ can also be interpreted as a limit. Indeed,

$$\begin{aligned} &\lim_{h\to 0^+} \mathrm{P}\left(x \le X < x + dx \,|\, y \le Y < y + h\right) \\ &= \lim_{h\to 0^+} \frac{\mathrm{P}\left(x \le X < x + dx,\, y \le Y < y + h\right)}{\mathrm{P}\left(y \le Y < y + h\right)} \\ &\sim \lim_{h\to 0^+} \frac{f(x,y)\,h dx}{f_Y(y)\,h} = \frac{f(x,y)\,dx}{f_Y(y)} = f_{X|Y}(x,y)\,dx, \end{aligned} \tag{5.159}$$

wherever $f(x,y)$ and $f_Y(y)$ exist and are continuous and $f_Y(y) \neq 0$. On the other hand, $\mathrm{P}(A|Y = y)$ can be interpreted also *without* a limit as

$$\mathrm{P}(A|Y = y) = \frac{\mathrm{P}(A)\, f_{Y|A}(y)}{f_Y(y)}, \tag{5.160}$$

wherever $f_{Y|A}(y)$ and $f_Y(y)$ exist and are continuous and $f_Y(y) \neq 0$, because then

$$\begin{aligned} &\lim_{h\to 0^+} \mathrm{P}\left(A|y \le Y < y + h\right) \\ &= \lim_{h\to 0^+} \frac{\mathrm{P}\left(A \cap \{y \le Y < y + h\}\right)}{\mathrm{P}\left(y \le Y < y + h\right)} = \lim_{h\to 0^+} \frac{\mathrm{P}(A)\,\mathrm{P}\left(y \le Y < y + h \,|\, A\right)}{\mathrm{P}\left(y \le Y < y + h\right)} \\ &= \lim_{h\to 0^+} \frac{\mathrm{P}(A)\, f_{Y|A}(y) h}{f_Y(y)\, h} = \frac{\mathrm{P}(A)\, f_{Y|A}(y)}{f_Y(y)}. \end{aligned} \tag{5.161}$$

Equation 5.160 is valid also when $f_{Y|A}(y)$ and $f_Y(y)$ exist and are continuous and $f_Y(y) \neq 0$, but $\mathrm{P}(A) = 0$. For, in this case, $A \cap \{y \leq Y < y + h\} \subset A$, and so $\mathrm{P}(A \cap \{y \leq Y < y + h\}) = 0$, which implies $\mathrm{P}(A|y \leq Y < y + h) = 0$ and $\mathrm{P}(A|Y = y) = 0$ as well. Thus, Equation 5.160 reduces to $0 = 0$.

Equation 5.160 can be written in multiplicative form as

$$\mathrm{P}(A|Y = y) f_Y(y) = \mathrm{P}(A) f_{Y|A}(y). \tag{5.162}$$

This equation is valid also when $f_Y(y) = 0$, as well, because in that case $f_{Y|A}(y) = 0$ as well. This fact follows from Equation 5.152 with Y in place of X:

$$\begin{aligned} f_{Y|A}(y)dy &\sim \frac{\mathrm{P}(\{y \leq Y < y + dy\} \cap A)}{\mathrm{P}(A)} \\ &\leq \frac{\mathrm{P}(y \leq Y < y + dy)}{\mathrm{P}(A)} \sim \frac{f_Y(y)\,dy}{\mathrm{P}(A)} = 0. \end{aligned} \tag{5.163}$$

Similarly, Equation 5.157 too can be written in multiplicative form as

$$f_{X|Y}(x, y) f_Y(y) = f(x, y). \tag{5.164}$$

This equation is valid when $f_Y(y) = 0$, as well, because $f_Y(y) = 0$ implies $f(x, y) = 0$. Interchanging x and y, we also have

$$f_{Y|X}(x, y) f_X(x) = f(x, y). \tag{5.165}$$

Returning to $f_{X|Y}$, we can see that, for any fixed y such that $f_Y(y) \neq 0$, it is a density as a function of x. Consequently, it can be used to define conditional probabilities for X, given $Y = y$, as

$$\mathrm{P}(a < X < b|Y = y) = \int_a^b f_{X|Y}(x, y)\,dx = \frac{1}{f_Y(y)} \int_a^b f(x, y)\,dx \tag{5.166}$$

and, in particular, the *conditional distribution function of* X, *given* $Y = y$, as

$$F_{X|Y}(x, y) = \int_{-\infty}^x f_{X|Y}(t, y)\,dt = \frac{1}{f_Y(y)} \int_{-\infty}^x f(t, y)\,dt. \tag{5.167}$$

Using Definition 5.6.2, we can generalize the Theorem of Total Probability (Theorem 4.5.2) as follows:

Theorem 5.6.1. *Theorem of Total Probability, Continuous Versions.* *For a continuous random variable Y and any event A, if $f_{Y|A}$ and f_Y exist for all y, then*

$$P(A) = \int_{-\infty}^{\infty} P(A|Y = y) f_Y(y)\,dy \tag{5.168}$$

and if X and Y are both continuous and $f_{X|Y}$ and f_Y exist for all x, y, then

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x, y) f_Y(y)\,dy. \tag{5.169}$$

Proof. Integrating both sides of Equation 5.162 from $-\infty$ to ∞, we obtain Equation 5.168, because $\int_{-\infty}^{\infty} f_{Y|A}(y)dy = 1$ from Equation 5.148.

Similarly, integrating both sides of Equation 5.164 with respect to y from $-\infty$ to ∞, we obtain Equation 5.169. ∎

We have new versions of Bayes' theorem as well:

Theorem 5.6.2. ***Bayes' Theorem, Continuous Versions.*** *For a continuous random variable Y and any event A with nonzero probability, if $P(A|Y=y)$ and f_Y exist for all y, then*

$$f_{Y|A}(y) = \frac{P(A|Y=y)\, f_Y(y)}{\int_{-\infty}^{\infty} P(A|Y=y)\, f_Y(y)\, dy}. \tag{5.170}$$

Here f_Y is called the prior density of Y, and $f_{Y|A}$ its posterior density, referring to the fact that these are the densities of Y before and after the observation of A.

Furthermore, if X and Y are both continuous, $f_{X|Y}$ and f_Y exist for all x, y, and $f_X(x) \neq 0$, then

$$f_{Y|X}(y,x) = \frac{f_{X|Y}(x,y)\, f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x,y)\, f_Y(y)\, dy}. \tag{5.171}$$

Again, f_Y is called the prior density of Y, and $f_{Y|X}$ its posterior density.

Proof. From Equation 5.162 we get, when $\mathrm{P}(A) \neq 0$,

$$f_{Y|A}(y) = \frac{\mathrm{P}(A|Y=y)\, f_Y(y)}{\mathrm{P}(A)}. \tag{5.172}$$

Substituting the expression for $\mathrm{P}(A)$ here from Equation 5.168, we obtain Equation 5.170.

Similarly, from Equations 5.164 and 5.165, we obtain, when $f_X(x) \neq 0$,

$$f_{Y|X}(y,x) = \frac{f_{X|Y}(x,y)\, f_Y(y)}{f_X(x)}, \tag{5.173}$$

and substituting the expression for $f_X(x)$ here from Equation 5.169, we obtain Equation 5.171. ∎

Example 5.6.3. Bayes Estimate of a Bernoulli Parameter.

Suppose that X is a Bernoulli random variable with an unknown parameter P, which is uniformly distributed on the interval $[0,1]$. In other words,[19]

$$f_{X|P}(x,p) = p^x (1-p)^{1-x} \quad \text{for } x = 0, 1 \tag{5.174}$$

[19] We assume $0^0 = 1$ where necessary.

and

$$f_P(p) = \begin{cases} 1 \text{ for } p \in [0,1] \\ 0 \text{ otherwise} \end{cases}. \tag{5.175}$$

We make an observation of X and want to find the posterior density $f_{P|X}(p,x)$ of P. (This problem is a very simple example of the so-called Bayesian method of statistical estimation. It will be generalized to several observations instead of just one in Example 7.4.4.)

By Equation 5.171,

$$f_{P|X}(p,x) = \begin{cases} \frac{p^x(1-p)^{1-x}}{\int_0^1 p^x(1-p)^{1-x}dp} & \text{for } p \in [0,1] \text{ and } x = 0,1 \\ 0 & \text{otherwise} \end{cases}. \tag{5.176}$$

For $x = 1$ we have $\int_0^1 p^x (1-p)^{1-x} dp = \int_0^1 pdp = \frac{1}{2}$, and for $x = 0$, similarly, $\int_0^1 p^x (1-p)^{1-x} dp = \int_0^1 (1-p)\, dp = \frac{1}{2}$. Hence,

$$f_{P|X}(p,x) = \begin{cases} 2p & \text{for } p \in [0,1] \text{ and } x = 1 \\ 2(1-p) & \text{for } p \in [0,1] \text{ and } x = 0 \\ 0 & \text{otherwise} \end{cases}. \tag{5.177}$$

Thus, the observation changes the uniform prior density into a triangular posterior density that gives more weight to p-values near the observed value of X. ♦

Before closing this section, we want to present one more theorem, which follows from the definitions at once:

Theorem 5.6.3. ***Conditions for Independence of Random Variables.***

If A is any event with $\mathrm{P}(A) \neq 0$ and X any random variable, then A and X are independent of each other if and only if

$$F_{X|A}(x) = F_X(x) \text{ for all } x \in \mathbb{R}. \tag{5.178}$$

If X and Y are any random variables, then they are independent of each other if and only if

$$F_{X|Y}(x,y) = F_X(x) \tag{5.179}$$

for all $x \in \mathbb{R}$ and, for discrete Y, at all possible values y of Y and, for continuous Y, at all y values where $f_{X|Y}(x,y)$ exists.

If A is any event with $\mathrm{P}(A) \neq 0$ and X any discrete random variable, then A and X are independent of each other if and only if

$$f_{X|A}(x) = f_X(x) \text{ for all } x \in \mathbb{R}. \tag{5.180}$$

If X and Y are any random variables, both discrete or both absolutely continuous, then they are independent of each other if and only if

$$f_{X|Y}(x,y) = f_X(x) \tag{5.181}$$

for all $x \in \mathbb{R}$ and all y values where $f_Y(y) \neq 0$.

In closing this section, let us mention that all the conditional functions considered above can easily be generalized to more than two random variables, as will be seen in some exercises and later chapters.

Exercises

Exercise 5.6.1.

Roll four dice. Let X denote the number of 1's and Y the number of 6's obtained. Find the values of the p.f. $f_{X|Y}(x,y)$ and display them in a 5×5 table.

Exercise 5.6.2.

Roll two dice. Let X and Y denote the numbers obtained and let $Z = X + Y$:

1. Find the values of the p.f. $f_{X|Z}(x,z)$ and display them in a 6×11 table.
2. Find the values of the conditional joint p.f. $f_{(X,Y)|Z}(x,y,z)$ for $z = 2$, and show that X and Y are independent under this condition.
3. Find the values of the conditional joint p.f. $f_{(X,Y)|Z}(x,y,z)$ for $z = 3$, and show that X and Y are not independent under this condition.

Exercise 5.6.3.

As in Example 5.5.4, pick two random points X and Y independently and uniformly on the interval $[0,1]$, and let A denote the event that we can construct a triangle from the resulting three segments as its sides. Find the probability $\mathrm{P}(A|X = x)$ as a function of x and the conditional density function $f_{X|A}(x)$.

Exercise 5.6.4.

As in Example 5.6.3, let X be a Bernoulli random variable with an unknown parameter P, which is uniformly distributed on the interval $(0,1)$. Suppose we make two independent observations X_1 and X_2 of X, so that

$$f_{(X_1,X_2)|P}(x_1,x_2,p) = p^{x_1+x_2}(1-p)^{2-x_1-x_2} \quad \text{for } x_1, x_2 = 0, 1. \tag{5.182}$$

Find and graph $f_{P|(X_1,X_2)}(p,x_1,x_2)$ for all four possible values of (x_1,x_2).

Exercise 5.6.5.

Let (X,Y) be uniform on the triangle $D = \{(x,y) : 0 < x, 0 < y, x + y < 1\}$. Find the conditional densities $f_{X|Y}(x,y)$ and $f_{Y|X}(x,y)$.

Exercise 5.6.6.

Let $D = \{(x, y) : 0 < x, 0 < y, x + y < 1\}$ and (X, Y) have density

$$f(x, y) = \begin{cases} 60xy^2 & \text{if } (x, y) \in D \\ 0 & \text{otherwise} \end{cases} . \tag{5.183}$$

(See Example 5.4.8.) Find the conditional densities $f_{X|Y}(x, y)$ and $f_{Y|X}(x, y)$.

Exercise 5.6.7.

Let (X, Y) be uniform on the open unit square $D = \{(x, y) : 0 < x < 1, 0 < y < 1\}$ and $Z = X + Y$ (see Example 5.4.9):

1. Find the conditional distribution functions $F_{X|Z}(x, z)$ and $F_{Y|Z}(y, z)$ and the conditional densities $f_{X|Z}(x, z)$ and $f_{Y|Z}(y, z)$.
2. Let A be the event $\{Z < 1\}$. Find $F_{X|A}(x)$ and $f_{X|A}(x)$.

http://www.springer.com/978-3-319-30618-6

Introduction to Probability with Statistical Applications
Schay, G.
2016, XII, 385 p. 49 illus., Hardcover
ISBN: 978-3-319-30618-6
A product of Birkhäuser Basel