2

Random variables and their distributions

Summary. Quantities governed by randomness correspond to functions on the probability space called random variables. The value taken by a random variable is subject to chance, and the associated likelihoods are described by a function called the distribution function. Two important classes of random variables are discussed, namely discrete variables and continuous variables. The law of averages, known also as the law of large numbers, states that the proportion of successes in a long run of independent trials converges to the probability of success in any one trial. This result provides a mathematical basis for a philosophical view of probability based on repeated experimentation. Worked examples involving random variables and their distributions are included, and the chapter terminates with sections on random vectors and on Monte Carlo simulation.

2.1 Random variables

We shall not always be interested in an experiment itself, but rather in some consequence of its random outcome. For example, many gamblers are more concerned with their losses than with the games which give rise to them. Such consequences, when real valued, may be thought of as functions which map Ω into the real line $\mathbb{R}$, and these functions are called 'random† variables'.

(1) Example. A fair coin is tossed twice: $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$. For $\omega \in \Omega$, let $X(\omega)$ be the number of heads, so that

$$X(\mathrm{HH}) = 2, \quad X(\mathrm{HT}) = X(\mathrm{TH}) = 1, \quad X(\mathrm{TT}) = 0.$$

Now suppose that a gambler wagers his fortune of £1 on the result of this experiment. He gambles cumulatively so that his fortune is doubled each time a head appears, and is annihilated on the appearance of a tail. His subsequent fortune W is a random variable given by

$$W(\mathrm{HH}) = 4, \quad W(\mathrm{HT}) = W(\mathrm{TH}) = W(\mathrm{TT}) = 0. \qquad \bullet$$

†Derived from the Old French word *randon* meaning 'haste'.

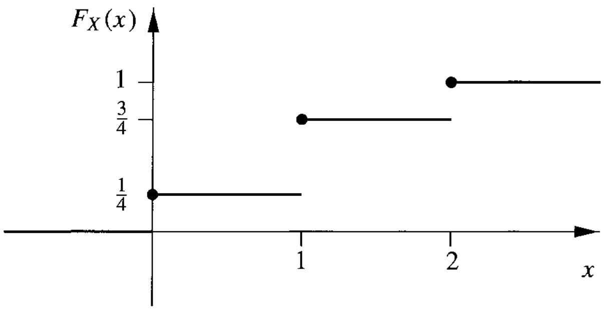

Figure 2.1. The distribution function F_X of the random variable X of Examples (1) and (5).

After the experiment is done and the outcome $\omega \in \Omega$ is known, a random variable $X : \Omega \to \mathbb{R}$ takes some value. In general this numerical value is more likely to lie in certain subsets of $\mathbb{R}$ than in certain others, depending on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the function X itself. We wish to be able to describe the distribution of the likelihoods of possible values of X. Example (1) above suggests that we might do this through the function $f : \mathbb{R} \to [0, 1]$ defined by

$$f(x) = \text{probability that } X \text{ is equal to } x,$$

but this turns out to be inappropriate in general. Rather, we use the *distribution function* $F : \mathbb{R} \to \mathbb{R}$ defined by

$$F(x) = \text{probability that } X \text{ does not exceed } x.$$

More rigorously, this is

$$F(x) = \mathbb{P}(A(x)) \tag{2}$$

where $A(x) \subseteq \Omega$ is given by $A(x) = \{\omega \in \Omega : X(\omega) \leq x\}$. However, $\mathbb{P}$ is a function on the collection $\mathcal{F}$ of events; we cannot discuss $\mathbb{P}(A(x))$ unless $A(x)$ belongs to $\mathcal{F}$, and so we are led to the following definition.

(3) Definition. A random variable is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. Such a function is said to be **$\mathcal{F}$-measurable**.

If you so desire, you may pay no attention to the technical condition in the definition and think of random variables simply as functions mapping Ω into $\mathbb{R}$. We shall always use upper-case letters, such as X, Y, and Z, to represent generic random variables, whilst lower-case letters, such as x, y, and z, will be used to represent possible numerical values of these variables. Do not confuse this notation in your written work.

Every random variable has a distribution function, given by (2); distribution functions are *very* important and useful.

(4) Definition. The **distribution function** of a random variable X is the function $F : \mathbb{R} \to [0, 1]$ given by $F(x) = \mathbb{P}(X \leq x)$.

This is the obvious abbreviation of equation (2). Events written as $\{\omega \in \Omega : X(\omega) \leq x\}$ are commonly abbreviated to $\{\omega : X(\omega) \leq x\}$ or $\{X \leq x\}$. We write F_X where it is necessary to emphasize the role of X.

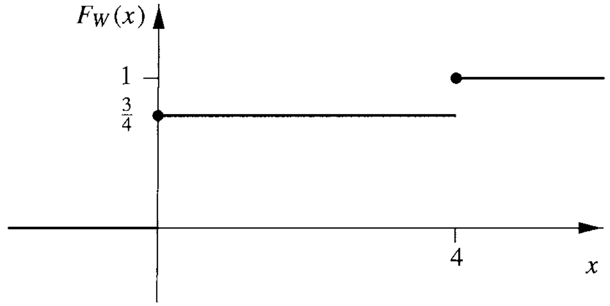

Figure 2.2. The distribution function F_W of the random variable W of Examples (1) and (5).

(5) Example (1) revisited. The distribution function F_X of X is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{4} & \text{if } 0 \le x < 1, \\ \frac{3}{4} & \text{if } 1 \le x < 2, \\ 1 & \text{if } x \ge 2, \end{cases}$$

and is sketched in Figure 2.1. The distribution function F_W of W is given by

$$F_W(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{3}{4} & \text{if } 0 \le x < 4, \\ 1 & \text{if } x \ge 4, \end{cases}$$

and is sketched in Figure 2.2. This illustrates the important point that the distribution function of a random variable X tells us about the values taken by X and their relative likelihoods, rather than about the sample space and the collection of events. ●

(6) Lemma. *A distribution function F has the following properties:*

(a) $\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1,$

(b) *if $x < y$ then $F(x) \le F(y)$,*

(c) *F is right-continuous, that is, $F(x+h) \to F(x)$ as $h \downarrow 0$.*

Proof.

(a) Let $B_n = \{\omega \in \Omega : X(\omega) \le -n\} = \{X \le -n\}$. The sequence $B_1, B_2, \dots$ is decreasing with the empty set as limit. Thus, by Lemma (1.3.5), $\mathbb{P}(B_n) \to \mathbb{P}(\varnothing) = 0$. The other part is similar.

(b) Let $A(x) = \{X \le x\}$, $A(x,y) = \{x < X \le y\}$. Then $A(y) = A(x) \cup A(x,y)$ is a disjoint union, and so by Definition (1.3.1),

$$\mathbb{P}(A(y)) = \mathbb{P}(A(x)) + \mathbb{P}(A(x,y))$$

giving

$$F(y) = F(x) + \mathbb{P}(x < X \le y) \ge F(x).$$

(c) This is an *exercise*. Use Lemma (1.3.5). ■

Actually, this lemma characterizes distribution functions. That is to say, F is the distribution function of some random variable if and only if it satisfies (6a), (6b), and (6c).

For the time being we can forget all about probability spaces and concentrate on random variables and their distribution functions. The distribution function F of X contains a great deal of information about X.

(7) Example. Constant variables. The simplest random variable takes a constant value on the whole domain Ω. Let $c \in \mathbb{R}$ and define $X : \Omega \to \mathbb{R}$ by

$$X(\omega) = c \quad \text{for all} \quad \omega \in \Omega.$$

The distribution function $F(x) = \mathbb{P}(X \le x)$ is the step function

$$F(x) = \begin{cases} 0 & x < c, \\ 1 & x \ge c. \end{cases}$$

Slightly more generally, we call X *constant* (*almost surely*) if there exists $c \in \mathbb{R}$ such that $\mathbb{P}(X = c) = 1$. ●

(8) Example. Bernoulli variables. Consider Example (1.3.2). Let $X : \Omega \to \mathbb{R}$ be given by

$$X(H) = 1, \quad X(T) = 0.$$

Then X is the simplest non-trivial random variable, having two possible values, 0 and 1. Its distribution function $F(x) = \mathbb{P}(X \le x)$ is

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - p & 0 \le x < 1, \\ 1 & x \ge 1. \end{cases}$$

X is said to have the *Bernoulli distribution* sometimes denoted $\text{Bern}(p)$. ●

(9) Example. Indicator functions. A particular class of Bernoulli variables is very useful in probability theory. Let A be an event and let $I_A : \Omega \to \mathbb{R}$ be the *indicator function* of A; that is,

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \in A^c. \end{cases}$$

Then I_A is a Bernoulli random variable taking the values 1 and 0 with probabilities $\mathbb{P}(A)$ and $\mathbb{P}(A^c)$ respectively. Suppose $\{B_i : i \in I\}$ is a family of disjoint events with $A \subseteq \bigcup_{i \in I} B_i$. Then

(10)
$$I_A = \sum_i I_{A \cap B_i},$$

an identity which is often useful. ●

(11) Lemma. *Let F be the distribution function of X. Then*

(a) $\mathbb{P}(X > x) = 1 - F(x)$,

(b) $\mathbb{P}(x < X \le y) = F(y) - F(x)$,

(c) $\mathbb{P}(X = x) = F(x) - \lim_{y \uparrow x} F(y)$.

Proof. (a) and (b) are *exercises*.

(c) Let $B_n = \{x - 1/n < X \le x\}$ and use the method of proof of Lemma (6). ■

Note one final piece of jargon for future use. A random variable X with distribution function F is said to have two 'tails' given by

$$T_1(x) = \mathbb{P}(X > x) = 1 - F(x), \quad T_2(x) = \mathbb{P}(X \le x) = F(-x),$$

where x is large and positive. We shall see later that the rates at which the T_i decay to zero as $x \to \infty$ have a substantial effect on the existence or non-existence of certain associated quantities called the 'moments' of the distribution.

Exercises for Section 2.1

1. Let X be a random variable on a given probability space, and let $a \in \mathbb{R}$. Show that

(i) aX is a random variable,

(ii) $X - X = 0$, the random variable taking the value 0 always, and $X + X = 2X$.

2. A random variable X has distribution function F. What is the distribution function of $Y = aX + b$, where a and b are real constants?

3. A fair coin is tossed n times. Show that, under reasonable assumptions, the probability of exactly k heads is $\binom{n}{k}(\frac{1}{2})^n$. What is the corresponding quantity when heads appears with probability p on each toss?

4. Show that if F and G are distribution functions and $0 \le \lambda \le 1$ then $\lambda F + (1 - \lambda)G$ is a distribution function. Is the product FG a distribution function?

5. Let F be a distribution function and r a positive integer. Show that the following are distribution functions:

(a) $F(x)^r$,

(b) $1 - \{1 - F(x)\}^r$,

(c) $F(x) + \{1 - F(x)\} \log\{1 - F(x)\}$,

(d) $\{F(x) - 1\}e + \exp\{1 - F(x)\}$.

2.2 The law of averages

We may recall the discussion in Section 1.3 of repeated experimentation. In each of N repetitions of an experiment, we observe whether or not a given event A occurs, and we write $N(A)$ for the total number of occurrences of A. One possible philosophical underpinning of probability theory requires that the proportion $N(A)/N$ settles down as $N \to \infty$ to some limit interpretable as the 'probability of A'. Is our theory to date consistent with such a requirement?

With this question in mind, let us suppose that $A_1, A_2, \ldots$ is a sequence of independent events having equal probability $\mathbb{P}(A_i) = p$, where $0 < p < 1$; such an assumption requires of course the existence of a corresponding probability space $(\Omega, \mathcal{F}, \mathbb{P})$, but we do not plan to get bogged down in such matters here. We think of A_i as being the event 'that A occurs on the ith experiment'. We write $S_n = \sum_{i=1}^{n} I_{A_i}$, the sum of the indicator functions of $A_1, A_2, \ldots, A_n$; S_n is a random variable which counts the number of occurrences of A_i for $1 \leq i \leq n$ (certainly S_n is a function of Ω, since it is the sum of such functions, and it is left as an *exercise* to show that S_n is $\mathcal{F}$-measurable). The following result concerning the ratio $n^{-1}S_n$ was proved by James Bernoulli before 1692.

(1) Theorem. *It is the case that $n^{-1}S_n$ converges to p as $n \to \infty$ in the sense that, for all $\epsilon > 0$,*

$$\mathbb{P}\big(p - \epsilon \leq n^{-1}S_n \leq p + \epsilon\big) \to 1 \quad \textit{as} \quad n \to \infty.$$

There are certain technicalities involved in the study of the convergence of random variables (see Chapter 7), and this is the reason for the careful statement of the theorem. For the time being, we encourage the reader to interpret the theorem as asserting simply that the proportion $n^{-1}S_n$ of times that the events $A_1, A_2, \ldots, A_n$ occur converges as $n \to \infty$ to their common probability p. We shall see later how important it is to be careful when making such statements.

Interpreted in terms of tosses of a fair coin, the theorem implies that the proportion of heads is (with large probability) near to $\frac{1}{2}$. As a caveat regarding the difficulties inherent in studying the convergence of random variables, we remark that it is *not* true that, in a 'typical' sequence of tosses of a fair coin, heads outnumber tails about one-half of the time.

Proof. Suppose that we toss a coin repeatedly, and heads occurs on each toss with probability p. The random variable S_n has the same probability distribution as the number H_n of heads which occur during the first n tosses, which is to say that $\mathbb{P}(S_n = k) = \mathbb{P}(H_n = k)$ for all k. It follows that, for small positive values of ϵ,

$$\mathbb{P}\left(\frac{1}{n}S_n \geq p + \epsilon\right) = \sum_{k \geq n(p+\epsilon)} \mathbb{P}(H_n = k).$$

We have from Exercise (2.1.3) that

$$\mathbb{P}(H_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for} \quad 0 \leq k \leq n,$$

and hence

$$\mathbb{P}\left(\frac{1}{n}S_n \geq p + \epsilon\right) = \sum_{k=m}^{n} \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

where $m = \lceil n(p+\epsilon) \rceil$, the least integer not less than $n(p+\epsilon)$. The following argument is standard in probability theory. Let $\lambda > 0$ and note that $e^{\lambda k} \geq e^{\lambda n(p+\epsilon)}$ if $k \geq m$. Writing $q = 1 - p$, we have that

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}S_n \geq p + \epsilon\right) &\leq \sum_{k=m}^{n} e^{\lambda[k - n(p+\epsilon)]} \binom{n}{k} p^k q^{n-k} \\
&\leq e^{-\lambda n \epsilon} \sum_{k=0}^{n} \binom{n}{k} (pe^{\lambda q})^k (qe^{-\lambda p})^{n-k} \\
&= e^{-\lambda n \epsilon} (pe^{\lambda q} + qe^{-\lambda p})^n,
\end{aligned}$$

by the binomial theorem. It is a simple *exercise* to show that $e^x \le x + e^{x^2}$ for $x \in \mathbb{R}$. With the aid of this inequality, we obtain

$$\text{(3)} \qquad \mathbb{P}\left(\frac{1}{n}S_n \ge p+\epsilon\right) \le e^{-\lambda n\epsilon}\left[pe^{\lambda^2 q^2} + qe^{\lambda^2 p^2}\right]^n \le e^{\lambda^2 n - \lambda n\epsilon}.$$

We can pick λ to minimize the right-hand side, namely $\lambda = \frac{1}{2}\epsilon$, giving

$$\text{(4)} \qquad \mathbb{P}\left(\frac{1}{n}S_n \ge p+\epsilon\right) \le e^{-\frac{1}{4}n\epsilon^2} \quad \text{for} \quad \epsilon > 0,$$

an inequality that is known as 'Bernstein's inequality'. It follows immediately that $\mathbb{P}(n^{-1}S_n \ge p+\epsilon) \to 0$ as $n \to \infty$. An exactly analogous argument shows that $\mathbb{P}(n^{-1}S_n \le p-\epsilon) \to 0$ as $n \to \infty$, and thus the theorem is proved. ■

Bernstein's inequality (4) is rather powerful, asserting that the chance that S_n exceeds its mean by a quantity of order n tends to zero *exponentially fast* as $n \to \infty$; such an inequality is known as a 'large-deviation estimate'. We may use the inequality to prove rather more than the conclusion of the theorem. Instead of estimating the chance that, for a specific value of n, S_n lies between $n(p-\epsilon)$ and $n(p+\epsilon)$, let us estimate the chance that this occurs *for all large n*. Writing $A_n = \{p-\epsilon \le n^{-1}S_n \le p+\epsilon\}$, we wish to estimate $\mathbb{P}(\bigcap_{n=m}^{\infty} A_n)$. Now the complement of this intersection is the event $\bigcup_{n=m}^{\infty} A_n^{c}$, and the probability of this union satisfies, by the inequalities of Boole and Bernstein,

$$\text{(5)} \qquad \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n^{c}\right) \le \sum_{n=m}^{\infty} \mathbb{P}(A_n^{c}) \le \sum_{n=m}^{\infty} 2e^{-\frac{1}{4}n\epsilon^2} \to 0 \quad \text{as} \quad m \to \infty,$$

giving that, as required,

$$\text{(6)} \qquad \mathbb{P}\left(p-\epsilon \le \frac{1}{n}S_n \le p+\epsilon \text{ for all } n \ge m\right) \to 1 \quad \text{as} \quad m \to \infty.$$

Exercises for Section 2.2

1. You wish to ask each of a large number of people a question to which the answer "yes" is embarrassing. The following procedure is proposed in order to determine the embarrassed fraction of the population. As the question is asked, a coin is tossed out of sight of the questioner. If the answer would have been "no" and the coin shows heads, then the answer "yes" is given. Otherwise people respond truthfully. What do you think of this procedure?

2. A coin is tossed repeatedly and heads turns up on each toss with probability p. Let H_n and T_n be the numbers of heads and tails in n tosses. Show that, for $\epsilon > 0$,

$$\mathbb{P}\left(2p-1-\epsilon \le \frac{1}{n}(H_n - T_n) \le 2p-1+\epsilon\right) \to 1 \qquad \text{as } n \to \infty.$$

3. Let $\{X_r : r \ge 1\}$ be observations which are independent and identically distributed with unknown distribution function F. Describe and justify a method for estimating $F(x)$.

2.3 Discrete and continuous variables

Much of the study of random variables is devoted to distribution functions, characterized by Lemma (2.1.6). The general theory of distribution functions and their applications is quite difficult and abstract and is best omitted at this stage. It relies on a rigorous treatment of the construction of the Lebesgue–Stieltjes integral; this is sketched in Section 5.6. However, things become much easier if we are prepared to restrict our attention to certain subclasses of random variables specified by properties which make them tractable. We shall consider in depth the collection of 'discrete' random variables and the collection of 'continuous' random variables.

(1) Definition. The random variable X is called **discrete** if it takes values in some countable subset $\{x_1, x_2, \dots\}$, only, of $\mathbb{R}$. The discrete random variable X has **(probability) mass function** $f : \mathbb{R} \to [0, 1]$ given by $f(x) = \mathbb{P}(X = x)$.

We shall see that the distribution function of a discrete variable has jump discontinuities at the values $x_1, x_2, \dots$ and is constant in between; such a distribution is called *atomic*. this contrasts sharply with the other important class of distribution functions considered here.

(2) Definition. The random variable X is called **continuous** if its distribution function can be expressed as

$$F(x) = \int_{-\infty}^{x} f(u)\, du \qquad x \in \mathbb{R},$$

for some integrable function $f : \mathbb{R} \to [0, \infty)$ called the **(probability) density function** of X.

The distribution function of a continuous random variable is certainly continuous (actually it is 'absolutely continuous'). For the moment we are concerned only with discrete variables and continuous variables. There is another sort of random variable, called 'singular', for a discussion of which the reader should look elsewhere. A common example of this phenomenon is based upon the Cantor ternary set (see Grimmett and Welsh 1986, or Billingsley 1995). Other variables are 'mixtures' of discrete, continuous, and singular variables. Note that the word 'continuous' is a misnomer when used in this regard: in describing X as continuous, we are referring to a property of its distribution function rather than of the random variable (function) X itself.

(3) Example. Discrete variables. The variables X and W of Example (2.1.1) take values in the sets $\{0, 1, 2\}$ and $\{0, 4\}$ respectively; they are both discrete. ●

(4) Example. Continuous variables. A straight rod is flung down at random onto a horizontal plane and the angle ω between the rod and true north is measured. The result is a number in $\Omega = [0, 2\pi)$. Never mind about $\mathcal{F}$ for the moment; we can suppose that $\mathcal{F}$ contains all nice subsets of Ω, including the collection of open subintervals such as (a, b), where $0 \le a < b < 2\pi$. The implicit symmetry suggests the probability measure $\mathbb{P}$ which satisfies $\mathbb{P}((a, b)) = (b - a)/(2\pi)$; that is to say, the probability that the angle lies in some interval is directly proportional to the length of the interval. Here are two random variables X and Y:

$$X(\omega) = \omega, \quad Y(\omega) = \omega^2.$$

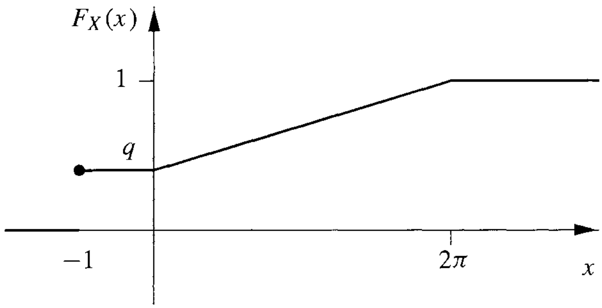

Figure 2.3. The distribution function F_X of the random variable X in Example (5).

Notice that Y is a function of X in that $Y = X^2$. The distribution functions of X and Y are

$$F_X(x) = \begin{cases} 0 & x \le 0, \\ x/(2\pi) & 0 \le x < 2\pi, \\ 1 & x \ge 2\pi, \end{cases} \qquad F_Y(y) = \begin{cases} 0 & y \le 0, \\ \sqrt{y}/(2\pi) & 0 \le y < 4\pi^2, \\ 1 & y \ge 4\pi^2. \end{cases}$$

To see this, let $0 \le x < 2\pi$ and $0 \le y < 4\pi^2$. Then

$$\begin{aligned} F_X(x) &= \mathbb{P}\big(\{\omega \in \Omega : 0 \le X(\omega) \le x\}\big) \\ &= \mathbb{P}\big(\{\omega \in \Omega : 0 \le \omega \le x\}\big) = x/(2\pi), \\ F_Y(y) &= \mathbb{P}\big(\{\omega : Y(\omega) \le y\}\big) \\ &= \mathbb{P}\big(\{\omega : \omega^2 \le y\}\big) = \mathbb{P}\big(\{\omega : 0 \le \omega \le \sqrt{y}\}\big) = \mathbb{P}(X \le \sqrt{y}) \\ &= \sqrt{y}/(2\pi). \end{aligned}$$

The random variables X and Y are continuous because

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du, \qquad F_Y(y) = \int_{-\infty}^{y} f_Y(u)\,du,$$

where

$$f_X(u) = \begin{cases} 1/(2\pi) & \text{if } 0 \le u \le 2\pi, \\ 0 & \text{otherwise,} \end{cases}$$
$$f_Y(u) = \begin{cases} u^{-\frac{1}{2}}/(4\pi) & \text{if } 0 \le u \le 4\pi^2, \\ 0 & \text{otherwise.} \end{cases}$$

●

(5) Example. A random variable which is neither continuous nor discrete. A coin is tossed, and a head turns up with probability $p(= 1-q)$. If a head turns up then a rod is flung on the ground and the angle measured as in Example (4). Then $\Omega = \{\mathrm{T}\} \cup \{(\mathrm{H}, x) : 0 \le x < 2\pi\}$, in the obvious notation. Let $X : \Omega \to \mathbb{R}$ be given by

$$X(\mathrm{T}) = -1, \qquad X((\mathrm{H}, x)) = x.$$

The random variable X takes values in $\{-1\} \cup [0, 2\pi)$ (see Figure 2.3 for a sketch of its distribution function). We say that X is continuous with the exception of a 'point mass (or *atom*) at -1'. ●

Exercises for Section 2.3

1. Let X be a random variable with distribution function F, and let $a = (a_m : -\infty < m < \infty)$ be a strictly increasing sequence of real numbers satisfying $a_{-m} \to -\infty$ and $a_m \to \infty$ as $m \to \infty$. Define $G(x) = \mathbb{P}(X \leq a_m)$ when $a_{m-1} \leq x < a_m$, so that G is the distribution function of a discrete random variable. How does the function G behave as the sequence a is chosen in such a way that $\sup_m |a_m - a_{m-1}|$ becomes smaller and smaller?

2. Let X be a random variable and let $g : \mathbb{R} \to \mathbb{R}$ be continuous and strictly increasing. Show that $Y = g(X)$ is a random variable.

3. Let X be a random variable with distribution function

$$\mathbb{P}(X \leq x\} = \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 < x \leq 1, \\ 1 & \text{if } x > 1. \end{cases}$$

Let F be a distribution function which is continuous and strictly increasing. Show that $Y = F^{-1}(X)$ is a random variable having distribution function F. Is it necessary that F be continuous and/or strictly increasing?

4. Show that, if f and g are density functions, and $0 \leq \lambda \leq 1$, then $\lambda f + (1 - \lambda)g$ is a density. Is the product fg a density function?

5. Which of the following are density functions? Find c and the corresponding distribution function F for those that are.

(a) $f(x) = \begin{cases} cx^{-d} & x > 1, \\ 0 & \text{otherwise}. \end{cases}$

(b) $f(x) = ce^x(1 + e^x)^{-2}$, $x \in \mathbb{R}$.

2.4 Worked examples

(1) Example. Darts. A dart is flung at a circular target of radius 3. We can think of the hitting point as the outcome of a random experiment; we shall suppose for simplicity that the player is guaranteed to hit the target somewhere. Setting the centre of the target at the origin of $\mathbb{R}^2$, we see that the sample space of this experiment is

$$\Omega = \{(x, y) : x^2 + y^2 < 9\}.$$

Never mind about the collection $\mathcal{F}$ of events. Let us suppose that, roughly speaking, the probability that the dart lands in some region A is proportional to its area $|A|$. Thus

$$\mathbb{P}(A) = |A|/(9\pi). \tag{2}$$

The scoring system is as follows. The target is partitioned by three concentric circles C_1, C_2, and C_3, centered at the origin with radii 1, 2, and 3. These circles divide the target into three annuli A_1, A_2, and A_3, where

$$A_k = \left\{(x, y) : k - 1 \leq \sqrt{x^2 + y^2} < k\right\}.$$

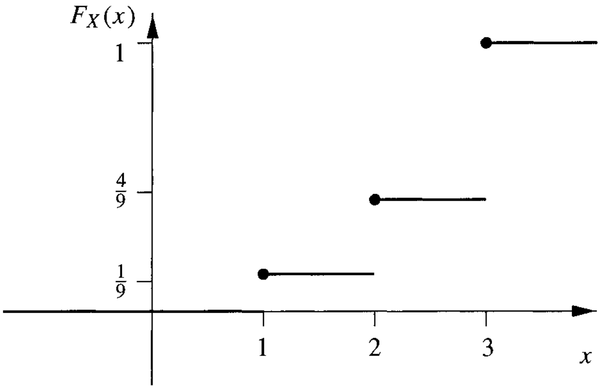

Figure 2.4. The distribution function F_X of X in Example (1).

We suppose that the player scores an amount k if and only if the dart hits A_k. The resulting score X is the random variable given by

$$X(\omega) = k \quad \text{whenever} \quad \omega \in A_k.$$

What is its distribution function?

Solution. Clearly

$$\mathbb{P}(X = k) = \mathbb{P}(A_k) = |A_k|/(9\pi) = \tfrac{1}{9}(2k - 1), \quad \text{for} \quad k = 1, 2, 3,$$

and so the distribution function of X is given by

$$F_X(r) = \mathbb{P}(X \le r) = \begin{cases} 0 & \text{if } r < 1, \\ \frac{1}{9}\lfloor r \rfloor^2 & \text{if } 1 \le r < 3, \\ 1 & \text{if } r \ge 3, \end{cases}$$

where $\lfloor r \rfloor$ denotes the largest integer not larger than r (see Figure 2.4). ●

(3) Example. Continuation of (1). Let us consider a revised method of scoring in which the player scores an amount equal to the distance between the hitting point ω and the centre of the target. This time the score Y is a random variable given by

$$Y(\omega) = \sqrt{x^2 + y^2}, \quad \text{if} \quad \omega = (x, y).$$

What is the distribution function of Y?

Solution. For any real r let C_r denote the disc with centre $(0, 0)$ and radius r, that is

$$C_r = \{(x, y) : x^2 + y^2 \le r\}.$$

Then

$$F_Y(r) = \mathbb{P}(Y \le r) = \mathbb{P}(C_r) = \tfrac{1}{9}r^2 \qquad \text{if} \qquad 0 \le r \le 3.$$

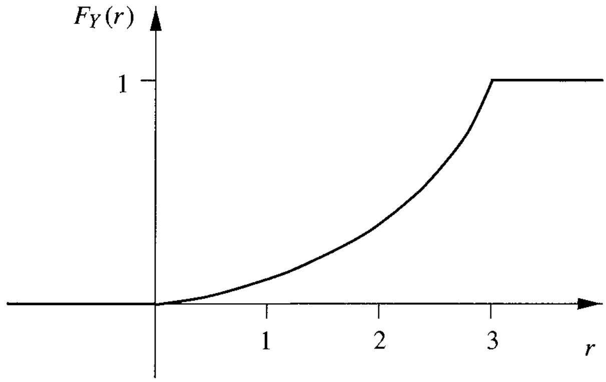

Figure 2.5. The distribution function F_Y of Y in Example (3).

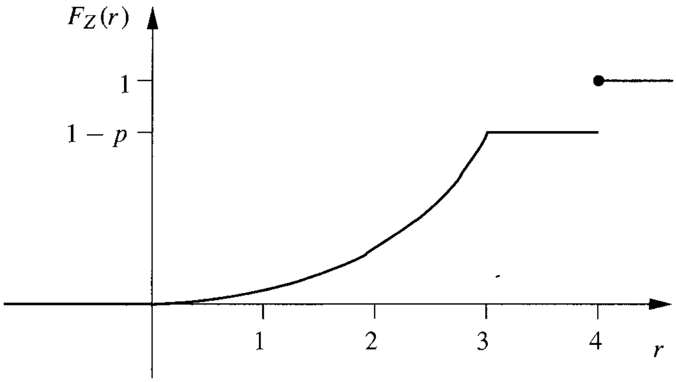

Figure 2.6. The distribution function F_Z of Z in Example (4).

This distribution function is sketched in Figure 2.5. ●

(4) Example. Continuation of (1). Now suppose that the player fails to hit the target with fixed probability p; if he is successful then we suppose that the distribution of the hitting point is described by equation (2). His score is specified as follows. If he hits the target then he scores an amount equal to the distance between the hitting point and the centre; if he misses then he scores 4. What is the distribution function of his score Z?

Solution. Clearly Z takes values in the interval $[0, 4]$. Use Lemma (1.4.4) to see that

$$\begin{aligned}
F_Z(r) &= \mathbb{P}(Z \le r) \\
&= \mathbb{P}(Z \le r \mid \text{hits target})\mathbb{P}(\text{hits target}) + \mathbb{P}(Z \le r \mid \text{misses target})\mathbb{P}(\text{misses target}) \\
&= \begin{cases} 0 & \text{if } r < 0, \\ (1-p)F_Y(r) & \text{if } 0 \le r < 4, \\ 1 & \text{if } r \ge 4, \end{cases}
\end{aligned}$$

where F_Y is given in Example (3) (see Figure 2.6 for a sketch of F_Z). ●

Exercises for Section 2.4

1. Let X be a random variable with a continuous distribution function F. Find expressions for the distribution functions of the following random variables:

(a) X^2, (b) $\sqrt{X}$,
(c) $\sin X$, (d) $G^{-1}(X)$,
(e) $F(X)$, (f) $G^{-1}(F(X))$,

where G is a continuous and strictly increasing function.

2. Truncation. Let X be a random variable with distribution function F, and let $a < b$. Sketch the distribution functions of the 'truncated' random variables Y and Z given by

$$Y = \begin{cases} a & \text{if } X < a, \\ X & \text{if } a \le X \le b, \\ b & \text{if } X > b, \end{cases} \qquad Z = \begin{cases} X & \text{if } |X| \le b, \\ 0 & \text{if } |X| > b. \end{cases}$$

Indicate how these distribution functions behave as $a \to -\infty$, $b \to \infty$.

2.5 Random vectors

Suppose that X and Y are random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Their distribution functions, F_X and F_Y, contain information about their associated probabilities. But how may we encapsulate information about their properties *relative to each other*? The key is to think of X and Y as being the components of a 'random vector' (X, Y) taking values in $\mathbb{R}^2$, rather than being unrelated random variables each taking values in $\mathbb{R}$.

(1) Example. Tontine is a scheme wherein subscribers to a common fund each receive an annuity from the fund during his or her lifetime, this annuity increasing as the other subscribers die. When all the subscribers are dead, the fund passes to the French government (this was the case in the first such scheme designed by Lorenzo Tonti around 1653). The performance of the fund depends on the lifetimes $L_1, L_2, \dots, L_n$ of the subscribers (as well as on their wealths), and we may record these as a vector $(L_1, L_2, \dots, L_n)$ of random variables. ●

(2) Example. Darts. A dart is flung at a conventional dartboard. The point of striking determines a distance R from the centre, an angle Θ with the upward vertical (measured clockwise, say), and a score S. With this experiment we may associate the random vector (R, Θ, S), and we note that S is a function of the pair (R, Θ). ●

(3) Example. Coin tossing. Suppose that we toss a coin n times, and set X_i equal to 0 or 1 depending on whether the ith toss results in a tail or a head. We think of the vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ as describing the result of this composite experiment. The total number of heads is the sum of the entries in $\mathbf{X}$. ●

An individual random variable X has a distribution function F_X defined by $F_X(x) = \mathbb{P}(X \le x)$ for $x \in \mathbb{R}$. The corresponding 'joint' distribution function of a random vector $(X_1, X_2, \dots, X_n)$ is the quantity $\mathbb{P}(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$, a function of n real variables $x_1, x_2, \dots, x_n$. In order to aid the notation, we introduce an ordering of vectors of real numbers: for vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$ we write $\mathbf{x} \le \mathbf{y}$ if $x_i \le y_i$ for each $i = 1, 2, \dots, n$.

(4) Definition. The **joint distribution function** of a random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the function $F_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]$ given by $F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \le \mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^n$.

As before, the expression $\{\mathbf{X} \le \mathbf{x}\}$ is an abbreviation for the event $\{\omega \in \Omega : \mathbf{X}(\omega) \le \mathbf{x}\}$. Joint distribution functions have properties similar to those of ordinary distribution functions. For example, Lemma (2.1.6) becomes the following.

(5) Lemma. *The joint distribution function $F_{X,Y}$ of the random vector (X, Y) has the following properties*:

(a) $\lim_{x,y\to-\infty} F_{X,Y}(x, y) = 0$, $\lim_{x,y\to\infty} F_{X,Y}(x, y) = 1$,

(b) *if* $(x_1, y_1) \le (x_2, y_2)$ *then* $F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$,

(c) $F_{X,Y}$ *is continuous from above, in that*

$$F_{X,Y}(x + u, y + v) \to F_{X,Y}(x, y) \quad \textit{as} \quad u, v \downarrow 0.$$

We state this lemma for a random vector with only two components X and Y, but the corresponding result for n components is valid also. The proof of the lemma is left as an *exercise*. Rather more is true. It may be seen without great difficulty that

(6) $$\lim_{y\to\infty} F_{X,Y}(x, y) = F_X(x) \; (= \mathbb{P}(X \le x))$$

and similarly

(7) $$\lim_{x\to\infty} F_{X,Y}(x, y) = F_Y(y) \; (= \mathbb{P}(Y \le y)).$$

This more refined version of part (a) of the lemma tells us that we may recapture the individual distribution functions of X and Y from a knowledge of their joint distribution function. The converse is false: it is not generally possible to calculate $F_{X,Y}$ from a knowledge of F_X and F_Y alone. The functions F_X and F_Y are called the 'marginal' distribution functions of $F_{X,Y}$.

(8) Example. A schoolteacher asks each member of his or her class to flip a fair coin twice and to record the outcomes. The diligent pupil D does this and records a pair (X_D, Y_D) of outcomes. The lazy pupil L flips the coin only once and writes down the result twice, recording thus a pair (X_L, Y_L) where $X_L = Y_L$. Clearly X_D, Y_D, X_L, and Y_L are random variables with the same distribution functions. However, the pairs (X_D, Y_D) and (X_L, Y_L) have different *joint* distribution functions. In particular, $\mathbb{P}(X_D = Y_D = \text{heads}) = \frac{1}{4}$ since only one of the four possible pairs of outcomes contains heads only, whereas $\mathbb{P}(X_L = Y_L = \text{heads}) = \frac{1}{2}$. ●

Once again there are two classes of random vectors which are particularly interesting: the 'discrete' and the 'continuous'.

(9) Definition. The random variables X and Y on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are called **(jointly) discrete** if the vector (X, Y) takes values in some countable subset of $\mathbb{R}^2$ only. The jointly discrete random variables X, Y have **joint (probability) mass function** $f : \mathbb{R}^2 \to [0, 1]$ given by $f(x, y) = \mathbb{P}(X = x, Y = y)$.

(10) Definition. The random variables X and Y on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are called **(jointly) continuous** if their joint distribution function can be expressed as

$$F_{X,Y}(x, y) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{y} f(u, v)\, du\, dv \qquad x, y \in \mathbb{R},$$

for some integrable function $f : \mathbb{R}^2 \to [0, \infty)$ called the **joint (probability) density function** of the pair (X, Y).

We shall return to such questions in later chapters. Meanwhile here are two concrete examples.

(11) Example. Three-sided coin. We are provided with a special three-sided coin, each toss of which results in one of the possibilities H (heads), T (tails), E (edge), each having probability $\frac{1}{3}$. Let H_n, T_n, and E_n be the numbers of such outcomes in n tosses of the coin. The vector (H_n, T_n, E_n) is a vector of random variables satisfying $H_n + T_n + E_n = n$. If the outcomes of different tosses have no influence on each other, it is not difficult to see that

$$\mathbb{P}\big((H_n, T_n, E_n) = (h, t, e)\big) = \frac{n!}{h!\,t!\,e!}\left(\frac{1}{3}\right)^n$$

for any triple (h, t, e) of non-negative integers with sum n. The random variables H_n, T_n, E_n are (jointly) discrete and are said to have (jointly) the *trinomial* distribution. ●

(12) Example. Darts. Returning to the flung dart of Example (2), let us assume that no region of the dartboard is preferred unduly over any other region of equal area. It may then be shown (see Example (2.4.3)) that

$$\mathbb{P}(R \le r) = \frac{r^2}{\rho^2}, \quad \mathbb{P}(\Theta \le \theta) = \frac{\theta}{2\pi}, \quad \text{for} \quad 0 \le r \le \rho,\ 0 \le \theta \le 2\pi,$$

where ρ is the radius of the board, and furthermore

$$\mathbb{P}(R \le r,\ \Theta \le \theta) = \mathbb{P}(R \le r)\mathbb{P}(\Theta \le \theta).$$

It follows that

$$F_{R,\Theta}(r, \theta) = \int_{u=0}^{r} \int_{v=0}^{\theta} f(u, v)\, du\, dv$$

where

$$f(u, v) = \frac{u}{\pi \rho^2}, \quad 0 \le u \le \rho,\ 0 \le v \le 2\pi.$$

The pair (R, Θ) is (jointly) continuous. ●

Exercises for Section 2.5

1. A fair coin is tossed twice. Let X be the number of heads, and let W be the indicator function of the event $\{X = 2\}$. Find $\mathbb{P}(X = x, W = w)$ for all appropriate values of x and w.

2. Let X be a Bernoulli random variable, so that $\mathbb{P}(X = 0) = 1 - p$, $\mathbb{P}(X = 1) = p$. Let $Y = 1 - X$ and $Z = XY$. Find $\mathbb{P}(X = x,\ Y = y)$ and $\mathbb{P}(X = x,\ Z = z)$ for $x, y, z \in \{0, 1\}$.

3. The random variables X and Y have joint distribution function

$$F_{X,Y}(x, y) = \begin{cases} 0 & \text{if } x < 0, \\ (1 - e^{-x})\left(\dfrac{1}{2} + \dfrac{1}{\pi}\tan^{-1} y\right) & \text{if } x \geq 0. \end{cases}$$

Show that X and Y are (jointly) continuously distributed.

4. Let X and Y have joint distribution function F. Show that

$$\mathbb{P}(a < X \leq b,\ c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

whenever $a < b$ and $c < d$.

5. Let X, Y be discrete random variables taking values in the integers, with joint mass function f. Show that, for integers x, y,

$$\begin{aligned} f(x, y) = \mathbb{P}(X \geq x,\ Y \leq y) - \mathbb{P}(X \geq x + 1,\ Y \leq y) \\ - \mathbb{P}(X \geq x,\ Y \leq y - 1) + \mathbb{P}(X \geq x + 1,\ Y \leq y - 1). \end{aligned}$$

Hence find the joint mass function of the smallest and largest numbers shown in r rolls of a fair die.

6. Is the function $F(x, y) = 1 - e^{-xy}$, $0 \leq x, y < \infty$, the joint distribution function of some pair of random variables?

2.6 Monte Carlo simulation

It is presumably the case that the physical shape of a coin is one of the major factors relevant to whether or not it will fall with heads uppermost. In principle, the shape of the coin may be determined by direct examination, and hence we may arrive at an estimate for the chance of heads. Unfortunately, such a calculation would be rather complicated, and it is easier to estimate this chance by simulation, which is to say that we may toss the coin many times and record the proportion of successes. Similarly, roulette players are well advised to observe the behaviour of the wheel with care in advance of placing large bets, in order to discern its peculiarities (unfortunately, casinos are now wary of such observation, and change their wheels at regular intervals).

Here is a related question. Suppose that we know that our coin is fair (so that the chance of heads is $\frac{1}{2}$ on each toss), and we wish to know the chance that a sequence of 50 tosses contains a run of outcomes of the form HTHHT. In principle, this probability may be calculated explicitly and exactly. If we require only an estimate of its value, then another possibility is to simulate the experiment: toss the coin $50N$ times for some N, divide the result into N runs of 50, and find the proportion of such runs which contain HTHHT.

It is not unusual in real life for a specific calculation to be possible in principle but extremely difficult in practice, often owing to limitations on the operating speed or the size of the memory of a computer. Simulation can provide a way around such a problem. Here are some examples.

(1) Example. Gambler's ruin revisited. The gambler of Example (1.7.4) eventually won his Jaguar after a long period devoted to tossing coins, and he has now decided to save up for a yacht. His bank manager has suggested that, in order to speed things up, the stake on each gamble should not remain constant but should vary as a certain prescribed function of the gambler's current fortune. The gambler would like to calculate the chance of winning the yacht in advance of embarking on the project, but he finds himself incapable of doing so.

Fortunately, he has kept a record of the extremely long sequence of heads and tails encountered in his successful play for the Jaguar. He calculates his sequence of hypothetical fortunes based on this information, until the point when this fortune reaches either zero or the price of the yacht. He then starts again, and continues to repeat the procedure until he has completed it a total of N times, say. He estimates the probability that he will actually win the yacht by the proportion of the N calculations which result in success.

Can you see why this method will make him overconfident? He might do better to retoss the coins. ●

(2) Example. A dam. It is proposed to build a dam in order to regulate the water supply, and in particular to prevent seasonal flooding downstream. How high should the dam be? Dams are expensive to construct, and some compromise between cost and risk is necessary. It is decided to build a dam which is just high enough to ensure that the chance of a flood of some given extent within ten years is less than 10^{-2}, say. No one knows exactly how high such a dam need be, and a young probabilist proposes the following scheme. Through examination of existing records of rainfall and water demand we may arrive at an acceptable model for the pattern of supply and demand. This model includes, for example, estimates for the distributions of rainfall on successive days over long periods. With the aid of a computer, the 'real world' situation is simulated many times in order to study the likely consequences of building dams of various heights. In this way we may arrive at an accurate estimate of the height required. ●

(3) Example. Integration. Let $g : [0, 1] \to [0, 1]$ be a continuous but nowhere differentiable function. How may we calculate its integral $I = \int_0^1 g(x)\,dx$? The following experimental technique is known as the 'hit or miss Monte Carlo technique'.

Let (X, Y) be a random vector having the uniform distribution on the unit square. That is, we assume that $\mathbb{P}\big((X, Y) \in A\big) = |A|$, the area of A, for any nice subset A of the unit square $[0, 1]^2$; we leave the assumption of niceness somewhat up in the air for the moment, and shall return to such matters in Chapter 4. We declare (X, Y) to be 'successful' if $Y \le g(X)$. The chance that (X, Y) is successful equals I, the area under the curve $y = g(x)$. We now repeat this experiment a large number N of times, and calculate the proportion of times that the experiment is successful. Following the law of averages, Theorem (2.2.1), we may use this value as an estimate of I.

Clearly it is desirable to know the accuracy of this estimate. This is a harder problem to which we shall return later. ●

Simulation is a dangerous game, and great caution is required in interpreting the results. There are two major reasons for this. First, a computer simulation is limited by the degree to which its so-called 'pseudo-random number generator' may be trusted. It has been said for example that the summon-according-to-birthday principle of conscription to the United States armed forces may have been marred by a pseudo-random number generator with a bias for some numbers over others. Secondly, in estimating a given quantity, one may in some circumstances have little or no idea how many repetitions are necessary in order to achieve an estimate within a specified accuracy.

We have made no remark about the methods by which computers calculate 'pseudo-random numbers'. Needless to say they do not flip coins, but rely instead on operations of sufficient numerical complexity that the outcome, although deterministic, is apparently unpredictable except by an exact repetition of the calculation.

These techniques were named in honour of Monte Carlo by Metropolis, von Neumann, and Ulam, while they were involved in the process of building bombs at Los Alamos in the 1940s.

2.7 Problems

1. Each toss of a coin results in a head with probability p. The coin is tossed until the first head appears. Let X be the total number of tosses. What is $\mathbb{P}(X > m)$? Find the distribution function of the random variable X.

2. (a) Show that any discrete random variable may be written as a linear combination of indicator variables.

(b) Show that any random variable may be expressed as the limit of an increasing sequence of discrete random variables.

(c) Show that the limit of any increasing convergent sequence of random variables is a random variable.

3. (a) Show that, if X and Y are random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then so are $X + Y$, XY, and $\min\{X, Y\}$.

(b) Show that the set of all random variables on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ constitutes a vector space over the reals. If Ω is finite, write down a basis for this space.

4. Let X have distribution function

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

and let $Y = X^2$. Find

(a) $\mathbb{P}(\frac{1}{2} \le X \le \frac{3}{2})$, (b) $\mathbb{P}(1 \le X < 2)$,
(c) $\mathbb{P}(Y \le X)$, (d) $\mathbb{P}(X \le 2Y)$,
(e) $\mathbb{P}(X + Y \le \frac{3}{4})$, (f) the distribution function of $Z = \sqrt{X}$.

5. Let X have distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < -1, \\ 1 - p & \text{if } -1 \le x < 0, \\ 1 - p + \frac{1}{2}xp & \text{if } 0 \le x \le 2, \\ 1 & \text{if } x > 2. \end{cases}$$

Sketch this function, and find: (a) $\mathbb{P}(X = -1)$, (b) $\mathbb{P}(X = 0)$, (c) $\mathbb{P}(X \ge 1)$.

6. Buses arrive at ten minute intervals starting at noon. A man arrives at the bus stop a random number X minutes after noon, where X has distribution function

$$\mathbb{P}(X \le x) = \begin{cases} 0 & \text{if } x < 0, \\ x/60 & \text{if } 0 \le x \le 60, \\ 1 & \text{if } x > 60. \end{cases}$$

What is the probability that he waits less than five minutes for a bus?

7. Airlines find that each passenger who reserves a seat fails to turn up with probability $\frac{1}{10}$ independently of the other passengers. So Teeny Weeny Airlines always sell 10 tickets for their 9 seat aeroplane while Blockbuster Airways always sell 20 tickets for their 18 seat aeroplane. Which is more often over-booked?

8. A fairground performer claims the power of telekinesis. The crowd throws coins and he wills them to fall heads up. He succeeds five times out of six. What chance would he have of doing at least as well if he had no supernatural powers?

9. Express the distribution functions of

$$X^+ = \max\{0, X\}, \quad X^- = -\min\{0, X\}, \quad |X| = X^+ + X^-, \quad -X,$$

in terms of the distribution function F of the random variable X.

10. Show that $F_X(x)$ is continuous at $x = x_0$ if and only if $\mathbb{P}(X = x_0) = 0$.

11. The real number m is called a *median* of the distribution function F whenever $\lim_{y \uparrow m} F(y) \le \frac{1}{2} \le F(m)$. Show that every distribution function F has at least one median, and that the set of medians of F is a closed interval of $\mathbb{R}$.

12. Show that it is not possible to weight two dice in such a way that the sum of the two numbers shown by these loaded dice is equally likely to take any value between 2 and 12 (inclusive).

13. A function $d : S \times S \to \mathbb{R}$ is called a *metric* on S if:

(i) $d(s,t) = d(t,s) \ge 0$ for all $s, t \in S$,

(ii) $d(s,t) = 0$ if and only if $s = t$, and

(iii) $d(s,t) \le d(s,u) + d(u,t)$ for all $s, t, u \in S$.

(a) **Lévy metric.** Let F and G be distribution functions and define the *Lévy metric*

$$d_L(F, G) = \inf\Big\{\epsilon > 0 : G(x - \epsilon) - \epsilon \le F(x) \le G(x + \epsilon) + \epsilon \text{ for all } x\Big\}.$$

Show that d_L is indeed a metric on the space of distribution functions.

(b) **Total variation distance.** Let X and Y be integer-valued random variables, and let

$$d_{TV}(X, Y) = \sum_k \big|\mathbb{P}(X = k) - \mathbb{P}(Y = k)\big|.$$

Show that d_{TV} satisfies (i) and (iii) with S the space of integer-valued random variables, and that $d_{TV}(X, Y) = 0$ if and only if $\mathbb{P}(X = Y) = 1$. Thus d_{TV} is a metric on the space of equivalence classes of S with equivalence relation given by $X \sim Y$ if $\mathbb{P}(X = Y) = 1$. We call d_{TV} the *total variation distance*.

Show that

$$d_{TV}(X, Y) = 2 \sup_{A \subseteq \mathbb{Z}} \big|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\big|.$$

14. Ascertain in the following cases whether or not F is the joint distribution function of some pair (X, Y) of random variables. If your conclusion is affirmative, find the distribution functions of X and Y separately.

(a) $$F(x, y) = \begin{cases} 1 - e^{-x-y} & \text{if } x, y \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

(b) $$F(x, y) = \begin{cases} 1 - e^{-x} - xe^{-y} & \text{if } 0 \le x \le y, \\ 1 - e^{-y} - ye^{-y} & \text{if } 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases}$$

15. It is required to place in order n books $B_1, B_2, \ldots, B_n$ on a library shelf in such a way that readers searching from left to right waste as little time as possible on average. Assuming that each reader requires book B_i with probability p_i, find the ordering of the books which minimizes $\mathbb{P}(T \geq k)$ for all k, where T is the (random) number of titles examined by a reader before discovery of the required book.

16. Transitive coins. Three coins each show heads with probability $\frac{3}{5}$ and tails otherwise. The first counts 10 points for a head and 2 for a tail, the second counts 4 points for both head and tail, and the third counts 3 points for a head and 20 for a tail.

You and your opponent each choose a coin; you cannot choose the same coin. Each of you tosses your coin and the person with the larger score wins £10^{10}. Would you prefer to be the first to pick a coin or the second?

17. Before the development of radar and inertial navigation, flying to isolated islands (for example, from Los Angeles to Hawaii) was somewhat 'hit or miss'. In heavy cloud or at night it was necessary to fly by dead reckoning, and then to search the surface. With the aid of a radio, the pilot had a good idea of the correct great circle along which to search, but could not be sure which of the two directions along this great circle was correct (since a strong tailwind could have carried the plane over its target). When you are the pilot, you calculate that you can make n searches before your plane will run out of fuel. On each search you will discover the island with probability p (if it is indeed in the direction of the search) independently of the results of other searches; you estimate initially that there is probability α that the island is ahead of you. What policy should you adopt in deciding the directions of your various searches in order to maximize the probability of locating the island?

18. Eight pawns are placed randomly on a chessboard, no more than one to a square. What is the probability that:

(a) they are in a straight line (do not forget the diagonals)?

(b) no two are in the same row or column?

19. Which of the following are distribution functions? For those that are, give the corresponding density function f.

(a) $F(x) = \begin{cases} 1 - e^{-x^2} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$

(b) $F(x) = \begin{cases} e^{-1/x} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$

(c) $F(x) = e^x/(e^x + e^{-x})$, $x \in \mathbb{R}$.

(d) $F(x) = e^{-x^2} + e^x/(e^x + e^{-x})$, $x \in \mathbb{R}$.

20. (a) If U and V are jointly continuous, show that $\mathbb{P}(U = V) = 0$.

(b) Let X be uniformly distributed on $(0, 1)$, and let $Y = X$. Then X and Y are continuous, and $\mathbb{P}(X = Y) = 1$. Is there a contradiction here?