

## 2.6 Random Variables

The concept of a random variable enables us to associate numerical quantities with experiments. This concept is essential in almost all problems of engineering and science. The calculation of averages, variances, probability distributions and so on all depend upon the use of random variables.

Concepts that are powerful and useful in a broad range of problems must have abstract expressions. In this chapter we will concentrate on the development of those expressions with only a few examples of their use. Later chapters will then apply them to problems in imaging science.

We have seen that the outcomes of experiments are not necessarily numerical. We associated numerical values to the faces of a die, but we could just as well have used distinguishable colors. When the experiment is a horse race, the winner is a horse and jockey, not a number. However, if you made a bet, the amount that you won or lost is a number. The amount that you would have won or lost is a function of which horse came in first. The function that relates the amount you might win with which horse comes in first is an example of a random variable.

This simple example illustrates two important facts about random variables. First, one can associate numerical values with the outcomes of just about any experiment. Second, the association is not in any way unique. One can easily define many such functions for any experiment. When we build a model of some real physical system we will find that the step of constructing the random variable is distinct from constructing the probability model. They are different parts that work together in a modeling system.

**Definition 2.6.1** *Let  $\mathbf{E}$  be an experiment whose outcomes are a sample space  $\mathcal{U}$  for which the probability  $P(e)$  is defined for each outcome  $e \in \mathcal{U}$ . A random variable is a function  $X(e)$  that associates a number with each outcome  $e \in \mathcal{U}$ .*

If  $\mathbf{E}$  is the horse race and  $\mathcal{U}$  is the set of horses, then  $X$  may be the amount you win.

It is useful to express random variables in terms of set operations. A random variable is simply a mapping from the space  $\mathcal{U}$  onto the real numbers,  $\mathcal{R}$ . If we say that  $x = X(e)$  we mean that  $x \in \mathcal{R}$  is the number that is associated with the outcome  $e \in \mathcal{U}$ . This is illustrated in Figure 2.3. In a sense, the values  $x \in \mathcal{R}$  can be considered as outcomes in a new experiment.

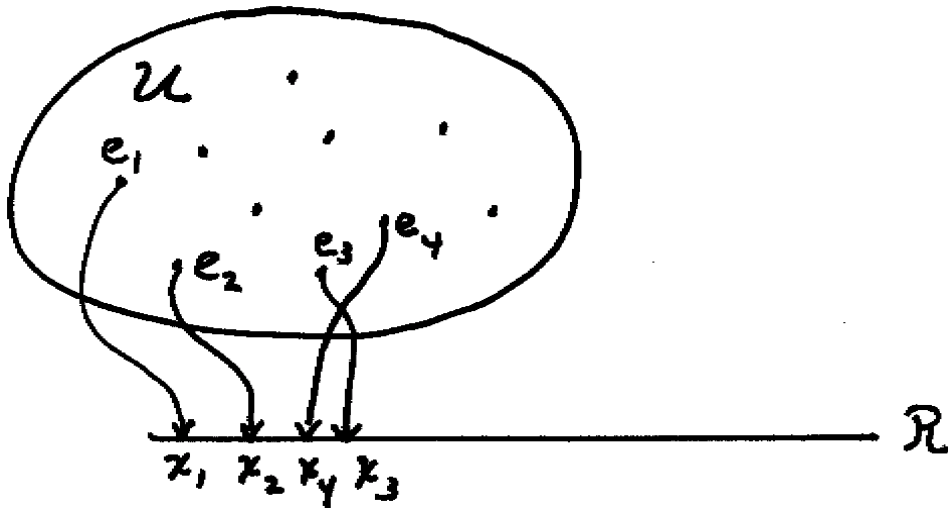


Figure 2.3: A random variable is a function that maps outcomes to numbers on the real line. Here the mapping is shown for a few of the outcomes in the sample space.

We have the option of describing an experiment that produces outcomes in some non-numerical space  $\mathcal{U}$  and then mapping it onto  $\mathcal{R}$  or describing an experiment that directly produces results in  $\mathcal{R}$ . Further, there are situations where outcomes are naturally in numerical form but we need to apply various functions to them to express values of interest. When an outcome is in numerical form then it is trivially a random variable as well. The option of applying transformations to both numerical and non-numerical outcomes is of great power and utility in systems analysis.

### 2.6.1 Events

Every interval on  $\mathcal{R}$  corresponds to a set of outcomes in  $\mathcal{U}$ . Let  $I \in \mathcal{R}$  be an interval and let  $A = \{e \in \mathcal{U} : X(e) = x, x \in I\}$ . Then  $A$  is the event that is associated with  $I$ , and  $P(I) = P(A)$ . Every interval of  $\mathcal{R}$  is an event whose probability can be calculated. This implies that there is an inverse operation from  $\mathcal{R}$  to  $\mathcal{U}$ , which we may write as  $A = X^{-1}(I)$ . Note that it is perfectly possible that there may be no points in  $\mathcal{U}$  that map to some selected interval  $I$ . In that case  $X^{-1}(I) = \phi$ , the empty set, and  $P(I) = P(\phi) = 0$ . It is

also possible that an interval may be selected such that  $A$  contains all of the points in  $\mathcal{U}$ . In that case  $X^{-1}(I) = \mathcal{U}$ , and  $P(I) = P(\mathcal{U}) = 1$ .

Not only is every interval of  $\mathcal{R}$  an event, but so is every collection of intervals. This gives us the tool to construct a probability measure on the real number line. Later it will be extended to any geometric space.

Consider the set of points  $R = \{x : X(e) = x \text{ for some } e \in \mathcal{U}\}$ . This set of points is the of the random variable. It is the entire set of the points that can be assumed by the random variable  $X$ .

## 2.6.2 Probability Distribution Function

Let  $X$  be a random variable and let  $x$  be a point on the real number line. Let  $I_x = \{s : s \leq x\}$  be the interval to the left of  $x$ . The probability  $P(I_x)$  is defined for every  $x$  by the mapping from  $\mathcal{U}$  to  $\mathcal{R}$  defined by  $X$ . As  $x$  is increased the interval includes more of  $\mathcal{R}$  so that  $P(I_x)$  must either increase or remain constant. Let

$$F_X(x) = P(I_x), \quad I_x = \{s : s \leq x\} \quad (2.17)$$

A common way to write the same expression is

$$F_X(x) = P(X \leq x) \quad (2.18)$$

In this expression  $x$  is a numerical value and  $X$  is a random variable. The distribution function is the probability that the random variable has a value less than or equal to the number  $x$  on a trial of the experiment.

The distribution function can be used to calculate the probability that a random variable  $X$  falls in any interval. The probability that the random variable  $X$  assumes a value  $x$  in the interval  $a < x \leq b$  for any pair of numbers  $a$  and  $b$  with  $a < b$  is

$$P(a < x \leq b) = F_X(b) - F_X(a) \quad (2.19)$$

Since a probability cannot be negative, this means that  $F_X(b) \geq F_X(a)$  so that  $F_X(x)$  is a monotonic non-decreasing function.

From (2.18) it is evident that  $F_X(x)$  has the limits

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad (2.20)$$

$$\lim_{x \rightarrow \infty} F_X(x) = 1 \quad (2.21)$$

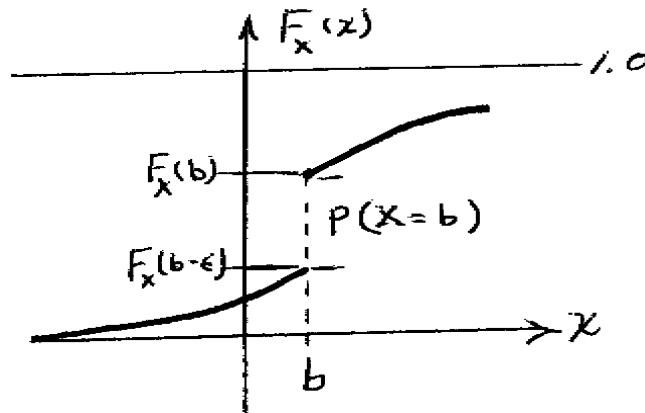


Figure 2.4: A probability distribution with a jump discontinuity at  $x = b$ . Distribution functions are continuous except at points such as  $b$  where  $P(X = b) > 0$ .

$F_X(x)$  is a nice function in the sense that it is either continuous or has step discontinuities. To see this, let  $a = b - \varepsilon$  and rewrite (2.19) as  $F_X(b) = F_X(b - \varepsilon) + P(b - \varepsilon < X \leq b)$ . The last term is just the probability that  $X$  falls in the interval  $(b - \varepsilon, b]$ , which is equal to  $P(X = b)$ . Hence,  $F_X(x)$  is either continuous at  $x = b$  or has a step of size  $P(X = b)$ . This is shown in Figure 2.4. Distribution functions come in three types, continuous, discrete and mixed. We will look at examples of each.

### Discrete Distribution

A discrete random variable can assume only discrete values. The number of values must be either finite or at least countable. The range of a discrete random variable consists of isolated points.

Let the values that can be assumed by  $X$  be  $x_k$ ,  $k = 0, 1, 2, \dots$ . Then the distribution function will have the staircase appearance shown in Figure 2.5. The steps occur at each  $x_k$  and have size  $P(X = x_k)$ .

The discrete values are the values that can be observed on a trial of the experiment.

**Example 2.6.1** *Suppose that we do an experiment that consists of the tossing three coins, a penny, a nickel and a dime. Let  $X$  be the number of*

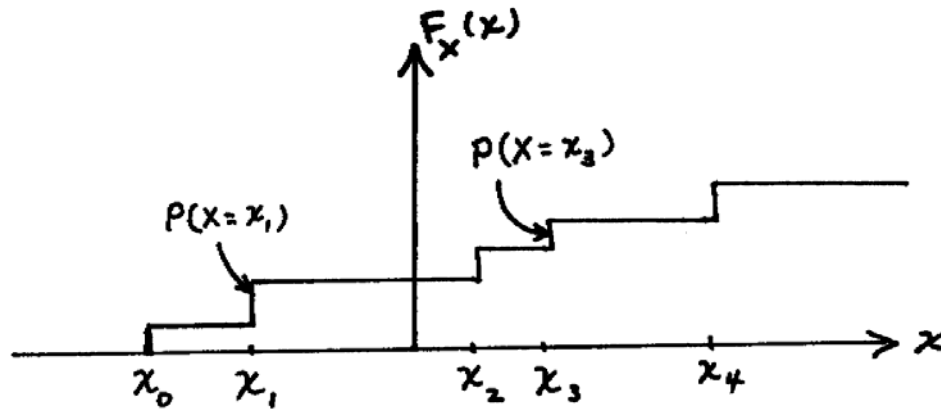


Figure 2.5: A discrete distribution function has a finite number of discontinuities. The random variable has a nonzero probability only at the points of discontinuity.

heads that appear. The possible values for the random variable are  $x_0 = 0$ ,  $x_1 = 1$ ,  $x_2 = 2$ ,  $x_3 = 3$  and the associated probabilities are  $P(X = 0) = 1/8$ ,  $P(X = 1) = 3/8$ ,  $P(X = 2) = 3/8$ ,  $P(X = 3) = 1/8$ . These probabilities can be computed by enumeration of the possibilities or use of the binomial formula that is developed in Exercise 5. The distribution function is shown in Figure 2.6

### Continuous Distribution

Suppose that  $F_X(x)$  is continuous for all  $x$ . Then  $\lim_{\varepsilon \rightarrow 0} F_X(x) - F_X(x - \varepsilon) = 0$  so that  $P(X = x) = 0$  for all  $x$ . That is, the probability that the random variable equals any chosen value of  $x$  is zero. This is like throwing a dart at a dartboard and asking the probability of exactly hitting any mathematical point. The total probability of hitting somewhere on the board is one, assuming it is early in the evening, but the probability of hitting any particular point is zero.

No isolated points exist in the range of the continuous random variable. If  $X$  can take on a value  $x$  then it can also take on other points in a neighborhood of size  $\varepsilon$  around  $x$ , no matter how small  $\varepsilon$ . The probability associated with any individual point is zero.

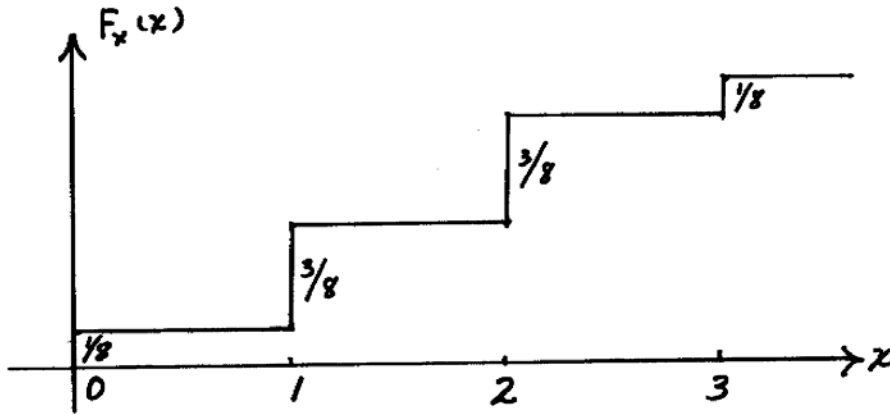


Figure 2.6: The probability distribution function for the number of heads in an experiment of tossing three fair coins.

We can compute the derivative of the continuous distribution by

$$\frac{dF_X(x)}{dx} = \lim_{\varepsilon \rightarrow 0} \frac{F_X(x) - F_X(x - \varepsilon)}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{P(x - \varepsilon < X \leq x)}{\varepsilon} \quad (2.22)$$

The slope of the probability distribution function is equivalent to the *density* of probability. We define this density by

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (2.23)$$

This function provides the probability that  $X$  falls in a small interval near  $x$ :  $P(x - \varepsilon < X \leq x) = f_X(x)\varepsilon$ . This is much like finding a mass by multiplying a density by a volume. The continuous distribution and its associated probability density function is shown in Figure 2.7. The relationship in (2.23) can be inverted, so that for any  $x$

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (2.24)$$

The probability that  $X$  is in the range  $a < X \leq b$  is therefore

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(u) du \quad (2.25)$$

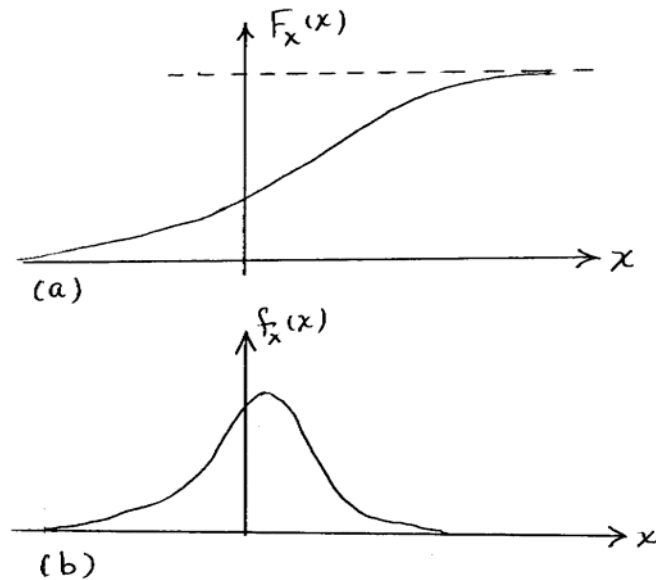


Figure 2.7: (a) The distribution function for a continuous random variable and (b) its probability density function. Note that the probability density function is highest where the slope of the distribution function is greatest.

This is the relationship that is illustrated in Figure 2.8. If  $|b - a| < \varepsilon$  and  $\varepsilon$  is of infinitesimal size then  $P(a < X \leq b) = f_X(x)\varepsilon$  for  $x$  in the interval  $(a, b)$ .

**Example 2.6.2** *The brightness of a pixel in a grayscale image is a random variable which may be called  $X$ . Consider scanning the image with a device that counts the fraction of pixels whose brightness is less than a level  $x$ . Let us call the function that is produced  $C(x)$ . This function will approximate the continuous probability distribution function,  $F_X(x)$ . You are asked to explore this experiment in Exercise 4.*

### Mixed Distribution

The range of a mixed distribution contains isolated points and points in a continuum. The distribution function is a smooth curve except at one or more points where there are finite steps. The derivative of the distribution

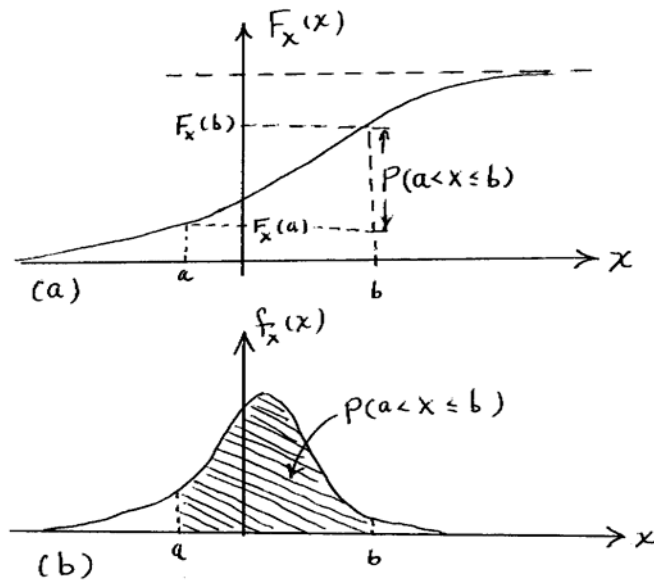


Figure 2.8: The probability  $P(a < X \leq b)$  is related to the change in height of the distribution and to the area shown in the probability density function.

function does not exist at the step points. However, it is common to write the distribution as the sum of the derivative of  $F_X(x)$  where the derivative is defined and delta functions where the steps occur. The weight of each delta function is equal to the height of the step at that point. Let  $x_k$ ,  $k = 0, 1, 2, \dots$  be the locations of finite jumps in  $F_X(x)$  and let  $c(x) = dF_X/dx$  at all other points. Then

$$f_X(x) = c(x) + \sum_k P(X = x_k)\delta(x - x_k) \quad (2.26)$$

The distribution and pdf for a mixed random variable is shown in Figure 2.9. In this case there are three finite steps in  $F_X(x)$  and three impulses in  $f_X(x)$ . The size of the step and the weight of the corresponding impulse are equal.

### 2.6.3 Multivariate Distributions

We have seen that a random variable is produced by mapping events with numbers on the real number line. The distribution of numbers along the line depends upon the probabilities of the associated events. This is a very

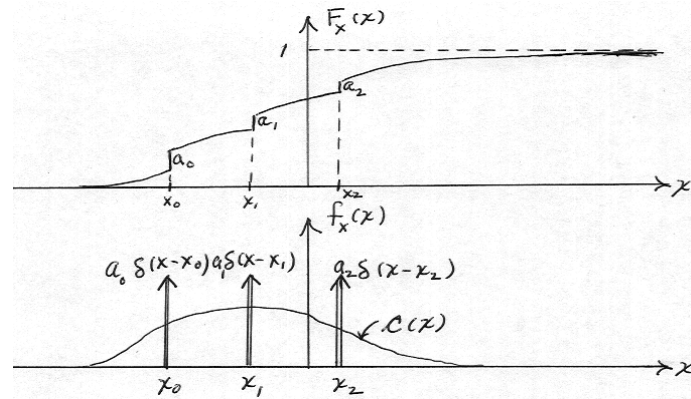


Figure 2.9: A mixed distribution exhibits steps in the distribution function and impulses in the density function. The height of each step is the weight of the corresponding impulse.

convenient way to model the production of certain number distributions by random experiments.

There is no reason in principle that an experiment should produce only one number. It is often useful to be able to model experiments in which the results are expressed in terms of several numbers. An example is the random selection of a person from a population. The person has a height, weight, girth and so on. These parameters may be considered by themselves as random variables or as a related set of random variables. In abstract mathematical terms, we can construct a vector  $\mathbf{X} = [X_1, X_2, \dots, X_r]$  to represent a set of  $r$  random variables that are produced by an experiment. The value of each component of  $\mathbf{X}$  is determined by the outcome of the experiment. We can define a random vector by extending definition 2.6.1.

**Definition 2.6.2** Let  $\mathbf{E}$  be an experiment whose outcomes are a sample space  $\mathcal{U}$  for which the probability  $P(e)$  is defined for each outcome  $e \in \mathcal{U}$ . A random vector is a function  $\mathbf{X}(e) = [X_1(e), X_2(e), \dots, X_r(e)]$  where  $X_i(e)$   $i = 1, 2, \dots, r$  are random variables defined over the space  $\mathcal{U}$ .

**Example 2.6.3** Let  $\mathbf{E}$  be the experiment of tossing a fair die. Let  $A$  be the event that an even face appears and  $B$  be the event that the number on the face is four or more. Let  $I(S)$  be the indicator function which takes on the value 1 if  $S$  is true and the value 0 if  $S$  is false. Then  $\mathbf{X} = [I(A), I(B)]$

is a random vector. The mapping of the outcomes of the experiment onto points in the plane are illustrated in Figure 2.10. The points that can be attained by  $\mathbf{X}$  are  $\{(0,0), (0,1), (1,0), (1,1)\}$ , which constitutes the range of  $\mathbf{X}$ . The probability of each point in the range can be found by summing the outcome probabilities in the related events. The event associated with point is a compound event,  $(0,0) \Leftrightarrow A^c \cap B^c = \{f_1, f_3\}$ ,  $(0,1) \Leftrightarrow A^c \cap B = \{f_5\}$ ,  $(1,0) \Leftrightarrow A \cap B^c = \{f_2\}$ ,  $(1,1) \Leftrightarrow A \cap B = \{f_4, f_6\}$ . The probabilities are  $P[\mathbf{X} = (0,0)] = 1/3$ ,  $P[\mathbf{X} = (0,1)] = 1/6$ ,  $P[\mathbf{X} = (1,0)] = 1/6$ ,  $P[\mathbf{X} = (1,1)] = 1/3$ .

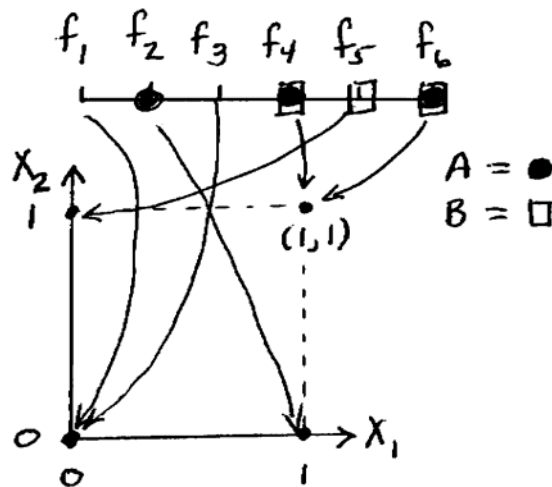


Figure 2.10: Mapping of the outcomes of the die tossing experiment onto points in a plane by a particular pair of random variables.

Random vectors can be used where we want to extend the number of variables to more than one. This is another tool that we can use in modeling random behavior. The extension to several random variables is the natural thing to do in many situations.

When there is more than one random variable associated with an experiment, it is natural to want to describe relationships between them. The most common relationships are joint probabilities, marginal probabilities and conditional probabilities. A variety of formulas for these relationships by using the event relationships in Section 2.3 and Section 2.4. The basic idea is quite

simple. We first notice that an expression such as  $X_1 = x$  defines an event. In Example 2.6.3 the expression  $X_1 = 0$  identifies the event  $\{f_1, f_3, f_5\}$ . Similarly, interval relationships such as  $a \leq X_1 < b$  can be used to define events in the outcome space of an experiment. This is also true of the random variable  $X_2$ . Thus, expressions such as  $P[(X_1 = x) \cap (X_2 = y)]$  make sense. So do expressions such as  $P[(a \leq X_1 < b) \cap (X_2 = y)]$ . These and many others can be constructed and computed by using the basic probability relationships that we already know. One of many is an expression like  $P[(a \leq X_1 < b) \cap (X_2 = y)] = P[a \leq X_1 < b]P[X_2 = y \mid a \leq X_1 < b]$ , which may be a convenient operation in some analysis.

Many other relationships are possible. The variety is too large to list. However, in each case it is possible to bring the analysis back to the basic ideas of event relationships. The exercises probe a number of the possibilities, and the applications that we will do throughout the course introduce many more.

### Joint-Probability Distribution Functions

The distribution function for a single random variable  $X$  was defined in (2.18) as the probability  $P[X \leq x]$ . This is the probability that the random variable falls in a particular interval. The joint distribution function for two random variables is defined in a similar manner. We define

$$F_{X_1 X_2}(x_1, x_2) = P[(X_1 \leq x_1) \cap (X_2 \leq x_2)] \quad (2.27)$$

This is a monotonically non-decreasing function of  $x_1$  and  $x_2$ . The following boundary relationships hold.

1.  $F_{X_1 X_2}(-\infty, -\infty) = 0$
2.  $F_{X_1 X_2}(-\infty, x_2) = 0$  for any  $x_2$
3.  $F_{X_1 X_2}(x_1, -\infty) = 0$  for any  $x_1$
4.  $F_{X_1 X_2}(+\infty, +\infty) = 1$
5.  $F_{X_1 X_2}(+\infty, x_2) = F_{X_2}(x_2)$  for any  $x_2$
6.  $F_{X_1 X_2}(x_1, +\infty) = F_{X_1}(x_1)$  for any  $x_1$

The last two relationships form a bridge between the joint distribution functions and the single-variate distribution functions. They are readily established by using (2.27).

The probability that an experiment produces a pair  $(X_1, X_2)$  that falls in a rectangular region with lower left corner  $(a, c)$  and upper right corner  $(b, d)$  is

$$P[(a < X_1 \leq b) \cap (c < X_2 \leq d)] = F_{X_1 X_2}(b, d) - F_{X_1 X_2}(a, d) - F_{X_1 X_2}(b, c) + F_{X_1 X_2}(a, c) \quad (2.28)$$

The rectangle is separated by subtracting the semi-infinite sections with corners at  $(a, d)$  and  $(b, c)$  from the semi-infinite section with corner  $(b, d)$ . Because the section with corner at  $(a, c)$  is subtracted twice, it needs to be added back. By extension, the distribution can be used to find the probability associated with any region that can be tiled by rectangles.

### Joint Probability Density Function

The probability density function is defined by computing the probability over a shrinking rectangle anchored at a point  $(x_1, x_2)$  divided by the area of the rectangle. This is simply the derivative at  $(x_1, x_2)$  provided the derivative exists. This can be expressed as the two-dimensional derivative

$$f_{X_1 X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1 X_2}(x_1, x_2)}{\partial x_1 \partial x_2} \quad (2.29)$$

It follows from the definition of the distribution function that the density function is non-negative. You are asked to carry out the details of showing this in Exercise 10.

The probability density function has a number of properties that are useful. Some of them are listed below. The notation  $U$  and  $V$  is used for the random variables and  $u$  and  $v$  for the values to reduce the number of subscripts. You are asked to compute these functions for a specific distribution

function in Exercise 11.

$$f_{U,V}(u, v) \geq 0 \quad (2.30)$$

$$F_{U,V}(u, v) = \int_{-\infty}^u \int_{-\infty}^v f_{U,V}(\xi, \eta) d\xi d\eta \quad (2.31)$$

$$\iint_{-\infty}^{\infty} f_{U,V}(\xi, \eta) d\xi d\eta = 1 \quad (2.32)$$

$$F_U(u) = \int_{-\infty}^u \int_{-\infty}^{\infty} f_{U,V}(\xi, \eta) d\xi d\eta \quad (2.33)$$

$$F_V(v) = \int_{-\infty}^{\infty} \int_{-\infty}^v f_{U,V}(\xi, \eta) d\xi d\eta \quad (2.34)$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u, \eta) d\eta \quad (2.35)$$

$$f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(\xi, v) d\xi \quad (2.36)$$

### Conditional Probability Distribution Functions

Let  $U$  and  $V$  be random variables and let  $A$  and  $B$  be sets of real numbers. Commonly these will be intervals, but that is not required. Then consider the event that an experiment with a probability density function  $f_{U,V}(u, v)$  has an outcome such that  $U \in A$  and  $V \in B$ . This can be expressed as<sup>2</sup>

$$P[U \in A, V \in B] = \int_{\xi \in A} \int_{\eta \in B} f_{U,V}(\xi, \eta) d\xi d\eta \quad (2.37)$$

The probability that  $V \in B$  without regard to the value of  $U$  is given by

$$P[V \in B] = \int_{\xi=-\infty}^{\infty} \int_{\eta \in B} f_{U,V}(\xi, \eta) d\xi d\eta \quad (2.38)$$

---

<sup>2</sup>We will use the notation  $P[U \in A, V \in B]$  rather than  $P[(U \in A) \cap (V \in B)]$  because it is more common in the literature and less cumbersome to write. The events in joint probability expressions are customarily separated by a comma when the operation is “and”.

We can now make use of the definition (2.12) to write the conditional probability computed from the joint density function.

$$\begin{aligned} P[U \in A|V \in B] &= \frac{P[U \in A, V \in B]}{P[V \in B]} \\ &= \frac{\int_{\xi \in A} \int_{\eta \in B} f_{U,V}(\xi, \eta) d\xi d\eta}{\int_{\xi=-\infty}^{\infty} \int_{\eta \in B} f_{U,V}(\xi, \eta) d\xi d\eta} \end{aligned} \quad (2.39)$$

provided  $P[V \in B] > 0$ . The probability that  $U$  falls in region  $A$  given that  $V$  falls in region  $B$  can be directly calculated from knowledge of the probability density functions. This construction is of interest in computing such relationships. But, it also is of interest in the calculation of the conditional probability distribution function.

**Definition 2.6.3** *The Conditional Probability Distribution Function is*

$$F_U(u|V \in B) = P[U \leq u|V \in B] \quad (2.40)$$

whenever  $P[V \in B] > 0$ .

The term on the right is computed from (2.39) with the region  $A = (-\infty, u]$ .

The conditional probability distribution function has all of the properties of an ordinary one-dimensional probability distribution function. That is, it is a nondecreasing function with  $F_U(-\infty|V \in B) = 0$  and  $F_U(\infty|V \in B) = 1$ . You should convince yourself of these properties by examining the definition. A conditional probability density function can be defined simply as the derivative, where it exists, of the conditional probability distribution function.

**Definition 2.6.4** *The conditional probability density function associated with  $F_U(u|V \in B)$  is*

$$f_U(u|V \in B) = \frac{dF_U(u|V \in B)}{du} \quad (2.41)$$

whenever the derivative exists.

### Statistically Independent Random Variables

We have noted that two events, say  $\mathcal{A}$  and  $\mathcal{B}$  are statistically independent when  $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$ . If we choose  $\mathcal{A}$  and  $\mathcal{B}$  to be the events  $U \in A$  and  $V \in B$ , where  $A$  and  $B$  are regions, then we have the basis for a definition of independent random variables.

**Definition 2.6.5** *Two random variables,  $U$  and  $V$  are statistically independent random variables if and only if*

$$P[U \in A, V \in B] = P[U \in A]P[V \in B] \quad (2.42)$$

for every possible choice of  $A$  and  $B$ .

If  $U$  and  $V$  are statistically independent random variables then

$$P[U \leq u, V \leq v] = P[U \leq u]P[V \leq v] \quad (2.43)$$

and this is equivalent to the following statement in terms of the distribution functions:

$$F_{U,V}(u, v) = F_U(u)F_V(v) \quad (2.44)$$

If the distribution functions are differentiable for almost all  $u$  and  $v$  then we also have the result

$$f_{U,V}(u, v) = f_U(u)f_V(v) \quad (2.45)$$

**Example 2.6.4** *A normal probability density function for two random variables  $U$  and  $V$  is given by*

$$f_{U,V}(u, v) = \frac{1}{2\pi} e^{-\left(\frac{u^2+v^2}{2}\right)} \quad (2.46)$$

Now,  $f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u, v) dv = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$ , which is easily established by using the fact that  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2/2} dv = 1$ . By a similar integration, we also find that  $f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$ . Therefore  $f_{U,V}(u, v) = f_U(u) f_V(v)$  and the random variables are statistically independent.

### 2.6.4 Functions of Random Variables

We often encounter a situation in which we know the distribution function (or density function) of a random variable, say  $U$ , and we want to find the probability description of another random variable that is constructed by a functional relationship, say  $V = g(U)$ . If the function  $g$  is any reasonable kind of transformation (it must be a Borel function, a condition that is satisfied by all functions we will encounter) then for any given value  $v$  we can find the set of values for  $U$  such that  $g(U) \leq v$ . We can then write the probability distribution function for  $V$  as

$$F_V(v) = P[u : g(u) \leq v] \quad (2.47)$$

If the inverse function  $u = g^{-1}(v)$  is known *and is single-valued* then the probability can be expressed as

$$F_V(v) = P[u \leq g^{-1}(v)] = \int_{-\infty}^{g^{-1}(v)} f_U(u) du \quad (2.48)$$

The probability density function  $f_V$  can be found by differentiation where the derivative exists.

$$f_V(v) = \frac{d}{dv} P[u : g(u) \leq v] \quad (2.49)$$

**Example 2.6.5** Suppose that  $U$  is uniformly distributed over the interval  $[-1, 1]$ , and that  $V = \sin \pi U/2$ . Then

$$f_U(u) = \begin{cases} 1/2, & -1 \leq u \leq 1 \\ 0, & \text{elsewhere} \end{cases} \quad (2.50)$$

The probability distribution function  $F_V(v)$  can be found by direct integration of (2.48) after substitution of this  $f_U(u)$  into ().

$$\begin{aligned} F_V(v) &= 0, & v < -1 \\ &= \int_{-1}^{\frac{2}{\pi} \text{Arcsin}(v)} f_U(u) du = \frac{1}{2} + \frac{1}{\pi} \text{Arcsin}(v), & -1 \leq v \leq 1 \\ &= 1, & v > 1 \end{aligned} \quad (2.51)$$

The probability density function can be found by differentiation of the above. This leads to

$$f_V(v) = \begin{cases} \frac{1}{\pi\sqrt{1-v^2}}, & -1 \leq v \leq 1 \\ 0, & \text{elsewhere} \end{cases} \quad (2.52)$$

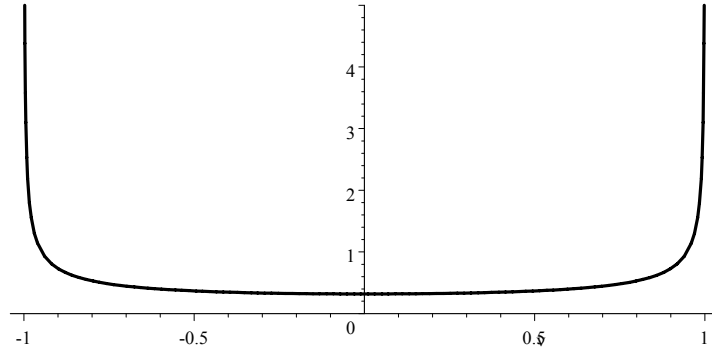


Figure 2.11: The probability density function  $f_V(v)$ .

*This function is plotted in Figure 2.11. Notice that its shape is substantially changed from that of a rectangular distribution for  $f_U(u)$  given by (2.6.5).*

### Sums of Random Variables

Let  $U$  and  $V$  be random variables, and let  $W = U + V$ . This is a situation that arises when we observe a signal in noise, for example. We need a means to compute the probability functions associated with  $W$  given that we know the distribution function  $F_{U,V}(u, v)$  and the pdf  $f_{U,V}(u, v)$ . The definition of the distribution function can be used to set this up. Since  $F_W(w) = P[W \leq w]$  and  $W = U + V$ , we have

$$F_W(w) = P[U + V \leq w] \quad (2.53)$$

This can be calculated by integrating  $f_{U,V}(u, v)$  over the region of the plane for which  $u + v \leq w$ . The expression for that calculation is

$$F_W(w) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{w-u} f_{U,V}(u, v) dv \right] du \quad (2.54)$$

The pdf can be obtained by differentiation<sup>3</sup>

$$\begin{aligned}
 f_W(w) &= \frac{d}{dw} \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{w-u} f_{U,V}(u, v) dv \right] du \\
 &= \int_{-\infty}^{\infty} \left[ \frac{d}{dw} \int_{-\infty}^{w-u} f_{U,V}(u, v) dv \right] du \\
 &= \int_{-\infty}^{\infty} f_{U,V}(u, w-u) du
 \end{aligned} \tag{2.55}$$

The computation of the integral in the last line is clear and straightforward, even if the details can be technically complicated. If necessary, it lends itself to numerical evaluation by computer.

A case that is particularly common and important arises when  $U$  and  $V$  are independent random variables. In that case we can apply (2.45) and write (2.55) as

$$f_W(w) = \int_{-\infty}^{\infty} f_U(u) f_V(w-u) du \tag{2.56}$$

If you look closely at the expression you will quickly recognize an old friend, namely the *convolution integral*! What a surprise to discover this old friend in a new setting!

---

<sup>3</sup>The differentiation makes use of Leibnitz' rule for the differentiation of an integral. Assuming that  $a(t)$ ,  $b(t)$  and  $r(s, t)$  are all differentiable with respect to  $t$  then  $\frac{d}{dt} \int_{a(t)}^{b(t)} r(s, t) ds = r[b(t), t]b'(t) - r[a(t), t]a'(t) + \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} r(s, t) ds$