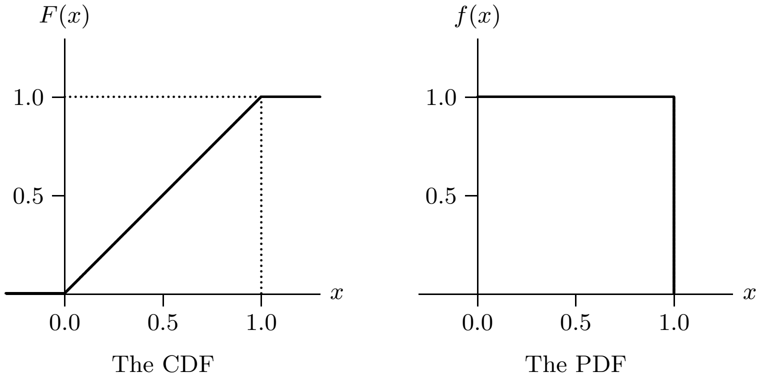

Figure 1.4 The CDF and PDF of the uniform distribution on $[0, 1]$

has probability 1, in order to keep the total probability equal to 1. Event A can be realized only if the realized point is in the intersection $A \cap B$, since the set of all points of A outside this intersection has zero probability. The probability of A, conditional on knowing that B has been realized, is thus the ratio of the area of $A \cap B$ to that of B. This construction leads directly to equation (1.14).

There are many ways to associate a random variable X with the rectangle shown in Figure 1.3. Such a random variable could be any function of the two coordinates that define a point in the rectangle. For example, it could be the horizontal coordinate of the point measured from the origin at the lower left-hand corner of the rectangle, or its vertical coordinate, or the Euclidean distance of the point from the origin. The realization of X is the value of the function it corresponds to at the realized point in the rectangle.

For concreteness, let us assume that the function is simply the horizontal coordinate, and let the width of the rectangle be equal to 1. Then, since all values of the horizontal coordinate between 0 and 1 are equally probable, the random variable X has what is called the **uniform distribution** on the interval $[0, 1]$. The CDF of this distribution is

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1. \end{cases}$$

Because $F(x)$ is not differentiable at $x = 0$ and $x = 1$, the PDF of the uniform distribution does not exist at those points. Elsewhere, the derivative of $F(x)$ is 0 outside $[0, 1]$ and 1 inside. The CDF and PDF are illustrated in Figure 1.4. This special case of the uniform distribution is often denoted the $\mathrm{U}(0, 1)$ distribution.

If the information were available that B had been realized, then the distribution of X conditional on this information would be very different from the

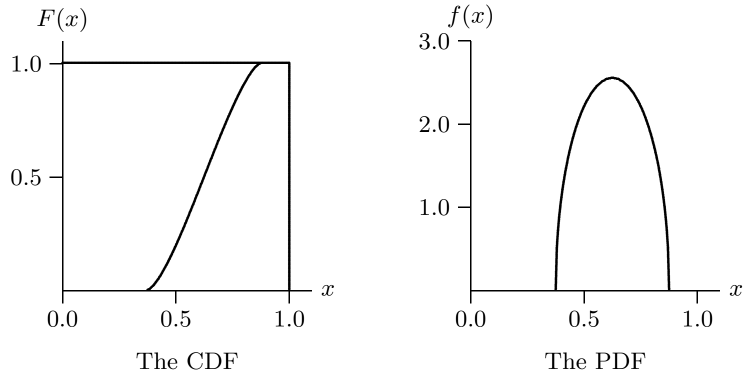

Figure 1.5 The CDF and PDF conditional on event B

U(0, 1) distribution. Now only values between the extreme horizontal limits of the circle of B are allowed. If one computes the area of the part of the circle to the left of a given vertical line, then for each event $a \equiv (X \leq x)$ the probability of this event conditional on B can be worked out. The result is just the CDF of X conditional on the event B. Its derivative is the PDF of X conditional on B. These are shown in Figure 1.5.

The concept of conditional probability can be extended beyond probability conditional on an event to probability conditional on a random variable. Suppose that X_1 is a r.v. and X_2 is a discrete r.v. with permitted values $z_1, \ldots, z_m$. For each $i = 1, \ldots, m$, the CDF of X_1, and, if X_1 is continuous, its PDF, can be computed conditional on the event $(X_2 = z_i)$. If X_2 is also a continuous r.v., then things are a little more complicated, because events like $(X_2 = x_2)$ for some real x_2 have zero probability, and so cannot be conditioned on in the manner of (1.14).

On the other hand, it makes perfect intuitive sense to think of the distribution of X_1 conditional on some specific realized value of X_2. This conditional distribution gives us the probabilities of events concerning X_1 when we know that the realization of X_2 was actually x_2. We therefore make use of the **conditional density** of X_1 for a given value x_2 of X_2. This conditional density, or **conditional PDF**, is defined as

$$f(x_1 \,|\, x_2) = \frac{f(x_1, x_2)}{f(x_2)}. \tag{1.15}$$

Thus, for a given value x_2 of X_2, the conditional density is proportional to the joint density of X_1 and X_2. Of course, (1.15) is well defined only if $f(x_2) > 0$. In some cases, more sophisticated definitions can be found that would allow $f(x_1 \,|\, x_2)$ to be defined for all x_2 even if $f(x_2) = 0$, but we will not need these in this book. See, among others, Billingsley (1979).

Conditional Expectations

Whenever we can describe the distribution of a random variable, X_1, conditional on another, X_2, either by a conditional CDF or a conditional PDF, we can consider the **conditional expectation** or **conditional mean** of X_1. If it exists, this conditional expectation is just the ordinary expectation computed using the conditional distribution. If x_2 is a possible value for X_2, then this conditional expectation is written as $\mathrm{E}(X_1 \,|\, x_2)$.

For a given value x_2, the conditional expectation $\mathrm{E}(X_1 \,|\, x_2)$ is, like any other ordinary expectation, a deterministic, that is, nonrandom, quantity. But we can consider the expectation of X_1 conditional on *every* possible realization of X_2. In this way, we can construct a new random variable, which we denote by $\mathrm{E}(X_1 \,|\, X_2)$, the realization of which is $\mathrm{E}(X_1 \,|\, x_2)$ when the realization of X_2 is x_2. We can call $\mathrm{E}(X_1 \,|\, X_2)$ a deterministic function of the random variable X_2, because the realization of $\mathrm{E}(X_1 \,|\, X_2)$ is unambiguously determined by the realization of X_2.

Conditional expectations defined as random variables in this way have a number of interesting and useful properties. The first, called the **Law of Iterated Expectations**, can be expressed as follows:

$$\mathrm{E}\big(\mathrm{E}(X_1 \,|\, X_2)\big) = \mathrm{E}(X_1). \tag{1.16}$$

If a conditional expectation of X_1 can be treated as a random variable, then the conditional expectation itself may have an expectation. According to (1.16), this expectation is just the ordinary expectation of X_1.

Another property of conditional expectations is that any deterministic function of a conditioning variable X_2 is its own conditional expectation. Thus, for example, $\mathrm{E}(X_2 \,|\, X_2) = X_2$, and $\mathrm{E}(X_2^2 \,|\, X_2) = X_2^2$. Similarly, conditional on X_2, the expectation of a product of another random variable X_1 and a deterministic function of X_2 is the product of that deterministic function and the expectation of X_1 conditional on X_2:

$$\mathrm{E}\big(X_1 h(X_2) \,|\, X_2\big) = h(X_2)\mathrm{E}(X_1 \,|\, X_2), \tag{1.17}$$

for any deterministic function $h(\cdot)$. An important special case of this, which we will make use of in Section 1.5, arises when $\mathrm{E}(X_1 \,|\, X_2) = 0$. In that case, for any function $h(\cdot)$, $\mathrm{E}(X_1 h(X_2)) = 0$, because

$$\begin{aligned}
\mathrm{E}\big(X_1 h(X_2)\big) &= \mathrm{E}\big(\mathrm{E}(X_1 h(X_2) \,|\, X_2)\big) \\
&= \mathrm{E}\big(h(X_2)\mathrm{E}(X_1 \,|\, X_2)\big) \\
&= \mathrm{E}(0) = 0.
\end{aligned}$$

The first equality here follows from the Law of Iterated Expectations, (1.16). The second follows from (1.17). Since $\mathrm{E}(X_1 \,|\, X_2) = 0$, the third line then follows immediately. We will present other properties of conditional expectations as the need arises.