3

General Random Variables

Contents

Random variables with a continuous range of possible experimental values are quite common – the velocity of a vehicle traveling along the highway could be one example. If such a velocity is measured by a digital speedometer, the speedometer's reading is a discrete random variable. But if we also wish to model the exact velocity, a continuous random variable is called for. Models involving continuous random variables can be useful for several reasons. Besides being finer-grained and possibly more accurate, they allow the use of powerful tools from calculus and often admit an insightful analysis that would not be possible under a discrete model.

All of the concepts and methods introduced in Chapter 2, such as expectation, PMFs, and conditioning, have continuous counterparts. Developing and interpreting these counterparts is the subject of this chapter.

3.1 CONTINUOUS RANDOM VARIABLES AND PDFS

A random variable X is called **continuous** if its probability law can be described in terms of a nonnegative function f_X, called the **probability density function of** X, or PDF for short, which satisfies

$$\mathbf{P}(X \in B) = \int_B f_X(x)\,dx,$$

for every subset B of the real line.† In particular, the probability that the value of X falls within an interval is

$$\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)\,dx,$$

and can be interpreted as the area under the graph of the PDF (see Fig. 3.1). For any single value a, we have $\mathbf{P}(X = a) = \int_a^a f_X(x)\,dx = 0$. For this reason, including or excluding the endpoints of an interval has no effect on its probability:

$$\mathbf{P}(a \le X \le b) = \mathbf{P}(a < X < b) = \mathbf{P}(a \le X < b) = \mathbf{P}(a < X \le b).$$

Note that to qualify as a PDF, a function f_X must be nonnegative, i.e., $f_X(x) \ge 0$ for every x, and must also satisfy the normalization equation

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \mathbf{P}(-\infty < X < \infty) = 1.$$

† The integral $\int_B f_X(x)\,dx$ is to be interpreted in the usual calculus/Riemann sense and we implicitly assume that it is well-defined. For highly unusual functions and sets, this integral can be harder – or even impossible – to define, but such issues belong to a more advanced treatment of the subject. In any case, it is comforting to know that mathematical subtleties of this type do not arise if f_X is a piecewise continuous function with a finite number of points of discontinuity, and B is the union of a finite or countable number of intervals.

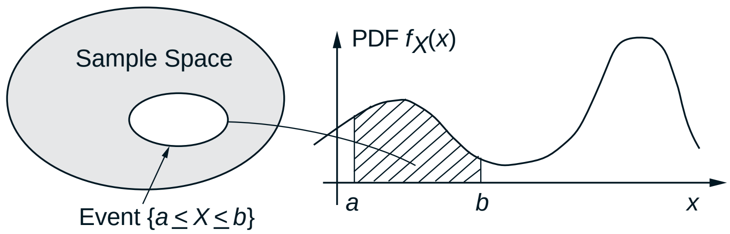

Figure 3.1: Illustration of a PDF. The probability that X takes value in an interval $[a,b]$ is $\int_a^b f_X(x)\,dx$, which is the shaded area in the figure.

Graphically, this means that the entire area under the graph of the PDF must be equal to 1.

To interpret the PDF, note that for an interval $[x, x+\delta]$ with very small length δ, we have

$$\mathbf{P}\big([x, x+\delta]\big) = \int_x^{x+\delta} f_X(t)\,dt \approx f_X(x)\cdot\delta,$$

so we can view $f_X(x)$ as the "probability mass per unit length" near x (cf. Fig. 3.2). It is important to realize that even though a PDF is used to calculate event probabilities, $f_X(x)$ is not the probability of any particular event. In particular, it is not restricted to be less than or equal to one.

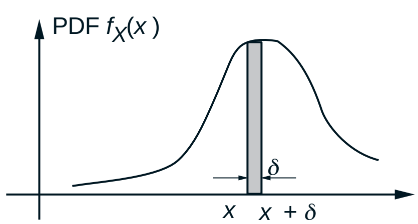

Figure 3.2: Interpretation of the PDF $f_X(x)$ as "probability mass per unit length" around x. If δ is very small, the probability that X takes value in the interval $[x, x+\delta]$ is the shaded area in the figure, which is approximately equal to $f_X(x)\cdot\delta$.

Example 3.1. Continuous Uniform Random Variable. A gambler spins a wheel of fortune, continuously calibrated between 0 and 1, and observes the resulting number. Assuming that all subintervals of [0,1] of the same length are equally likely, this experiment can be modeled in terms a random variable X with PDF

$$f_X(x) = \begin{cases} c & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

for some constant c. This constant can be determined by using the normalization property

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 c\,dx = c\int_0^1 dx = c$$

so that $c = 1$.

More generally, we can consider a random variable X that takes values in an interval $[a, b]$, and again assume that all subintervals of the same length are equally likely. We refer to this type of random variable as **uniform** or **uniformly distributed**. Its PDF has the form

$$f_X(x) = \begin{cases} c & \text{if } a \le x \le b, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. This is the continuous analog of the discrete uniform random variable discussed in Chapter 2. For f_X to satisfy the normalization property, we must have (cf. Fig. 3.3)

$$1 = \int_a^b c\,dx = c\int_a^b dx = c(b-a),$$

so that

$$c = \frac{1}{b-a}.$$

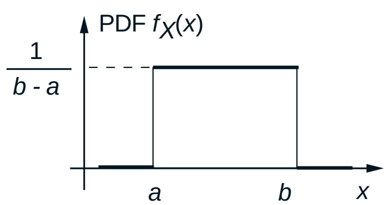

Figure 3.3: The PDF of a uniform random variable.

Note that the probability $\mathbf{P}(X \in I)$ that X takes value in a set I is

$$\mathbf{P}(X \in I) = \int_{[a,b]\cap I} \frac{1}{b-a}\,dx = \frac{1}{b-a}\int_{[a,b]\cap I} dx = \frac{\text{length of } [a,b]\cap I}{\text{length of } [a,b]}.$$

The uniform random variable bears a relation to the discrete uniform law, which involves a sample space with a finite number of equally likely outcomes. The difference is that to obtain the probability of various events, we must now calculate the "length" of various subsets of the real line instead of counting the number of outcomes contained in various events.

Example 3.2. Piecewise Constant PDF. Alvin's driving time to work is between 15 and 20 minutes if the day is sunny, and between 20 and 25 minutes if the day is rainy, with all times being equally likely in each case. Assume that a day is sunny with probability 2/3 and rainy with probability 1/3. What is the PDF of the driving time, viewed as a random variable X?

We interpret the statement that "all times are equally likely" in the sunny and the rainy cases, to mean that the PDF of X is constant in each of the intervals $[15, 20]$ and $[20, 25]$. Furthermore, since these two intervals contain all possible driving times, the PDF should be zero everywhere else:

$$f_X(x) = \begin{cases} c_1 & \text{if } 15 \le x < 20, \\ c_2 & \text{if } 20 \le x \le 25, \\ 0 & \text{otherwise,} \end{cases}$$

where c_1 and c_2 are some constants. We can determine these constants by using the given probabilities of a sunny and of a rainy day:

$$\frac{2}{3} = \mathbf{P}(\text{sunny day}) = \int_{15}^{20} f_X(x)\,dx = \int_{15}^{20} c_1\,dx = 5c_1,$$

$$\frac{1}{3} = \mathbf{P}(\text{rainy day}) = \int_{20}^{25} f_X(x)\,dx = \int_{20}^{25} c_2\,dx = 5c_2,$$

so that

$$c_1 = \frac{2}{15}, \qquad c_2 = \frac{1}{15}.$$

Generalizing this example, consider a random variable X whose PDF has the piecewise constant form

$$f_X(x) = \begin{cases} c_i & \text{if } a_i \le x < a_{i+1}, \quad i = 1, 2, \ldots, n-1, \\ 0 & \text{otherwise,} \end{cases}$$

where $a_1, a_2, \ldots, a_n$ are some scalars with $a_i < a_{i+1}$ for all i, and $c_1, c_2, \ldots, c_n$ are some nonnegative constants (cf. Fig. 3.4). The constants c_i may be determined by additional problem data, as in the case of the preceding driving context. Generally, the c_i must be such that the normalization property holds:

$$1 = \int_{a_1}^{a_n} f_X(x)\,dx = \sum_{i=1}^{n-1} \int_{a_i}^{a_{i+1}} c_i\,dx = \sum_{i=1}^{n-1} c_i(a_{i+1} - a_i).$$

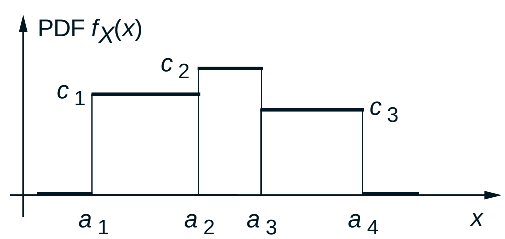

Figure 3.4: A piecewise constant PDF involving three intervals.

Example 3.3. A PDF can be arbitrarily large. Consider a random variable X with PDF

$$f_X(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} & \text{if } 0 < x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Even though $f_X(x)$ becomes infinitely large as x approaches zero, this is still a valid PDF, because

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 \frac{1}{2\sqrt{x}}\,dx = \sqrt{x}\Big|_0^1 = 1.$$

Summary of PDF Properties

Let X be a continuous random variable with PDF f_X.

- $f_X(x) \geq 0$ for all x.
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.
- If δ is very small, then $\mathbf{P}\big([x, x+\delta]\big) \approx f_X(x) \cdot \delta$.
- For any subset B of the real line,

$$\mathbf{P}(X \in B) = \int_B f_X(x)\,dx.$$

Expectation

The **expected value** or **mean** of a continuous random variable X is defined by†

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx.$$

† One has to deal with the possibility that the integral $\int_{-\infty}^{\infty} x f_X(x)\,dx$ is infinite or undefined. More concretely, we will say that the expectation is well-defined if $\int_{-\infty}^{\infty} |x| f_X(x)\,dx < \infty$. In that case, it is known that the integral $\int_{-\infty}^{\infty} x f_X(x)\,dx$ takes a finite and unambiguous value.

For an example where the expectation is not well-defined, consider a random variable X with PDF $f_X(x) = c/(1+x^2)$, where c is a constant chosen to enforce the normalization condition. The expression $|x| f_X(x)$ is approximately the same as $1/|x|$ when $|x|$ is large. Using the fact $\int_1^{\infty} (1/x)\,dx = \infty$, one can show that $\int_{-\infty}^{\infty} |x| f_X(x)\,dx = \infty$. Thus, $\mathbf{E}[X]$ is left undefined, despite the symmetry of the PDF around zero.

Throughout this book, in lack of an indication to the contrary, we implicitly assume that the expected value of the random variables of interest is well-defined.

This is similar to the discrete case except that the PMF is replaced by the PDF, and summation is replaced by integration. As in Chapter 2, $\mathbf{E}[X]$ can be interpreted as the "center of gravity" of the probability law and, also, as the anticipated average value of X in a large number of independent repetitions of the experiment. Its mathematical properties are similar to the discrete case – after all, an integral is just a limiting form of a sum.

If X is a continuous random variable with given PDF, any real-valued function $Y = g(X)$ of X is also a random variable. Note that Y can be a continuous random variable: for example, consider the trivial case where $Y = g(X) = X$. But Y can also turn out to be discrete. For example, suppose that $g(x) = 1$ for $x > 0$, and $g(x) = 0$, otherwise. Then $Y = g(X)$ is a discrete random variable. In either case, the mean of $g(X)$ satisfies the **expected value rule**

$$\mathbf{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx,$$

in complete analogy with the discrete case.

The n**th moment** of a continuous random variable X is defined as $\mathbf{E}[X^n]$, the expected value of the random variable X^n. The **variance**, denoted by $\mathrm{var}(X)$, is defined as the expected value of the random variable $\big(X - \mathbf{E}[X]\big)^2$.

We now summarize this discussion and list a number of additional facts that are practically identical to their discrete counterparts.

Expectation of a Continuous Random Variable and its Properties

Let X be a continuous random variable with PDF f_X.

- The expectation of X is defined by

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

- The expected value rule for a function $g(X)$ has the form

$$\mathbf{E}\big[g(X)\big] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx.$$

- The variance of X is defined by

$$\mathrm{var}(X) = \mathbf{E}\Big[\big(X - \mathbf{E}[X]\big)^2\Big] = \int_{-\infty}^{\infty} \big(x - \mathbf{E}[X]\big)^2 f_X(x)\, dx.$$

- We have
$$0 \le \mathrm{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2.$$
- If $Y = aX + b$, where a and b are given scalars, then
$$\mathbf{E}[Y] = a\mathbf{E}[X] + b, \qquad \mathrm{var}(Y) = a^2\mathrm{var}(X).$$

Example 3.4. Mean and Variance of the Uniform Random Variable. Consider the case of a uniform PDF over an interval $[a, b]$, as in Example 3.1. We have

$$\begin{aligned}
\mathbf{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\,dx \\
&= \int_a^b x \cdot \frac{1}{b-a}\,dx \\
&= \frac{1}{b-a} \cdot \frac{1}{2}x^2\Big|_a^b \\
&= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} \\
&= \frac{a+b}{2},
\end{aligned}$$

as one expects based on the symmetry of the PDF around $(a+b)/2$.

To obtain the variance, we first calculate the second moment. We have

$$\begin{aligned}
\mathbf{E}[X^2] &= \int_a^b \frac{x^2}{b-a}\,dx \\
&= \frac{1}{b-a}\int_a^b x^2\,dx \\
&= \frac{1}{b-a} \cdot \frac{1}{3}x^3\Big|_a^b \\
&= \frac{b^3 - a^3}{3(b-a)} \\
&= \frac{a^2 + ab + b^2}{3}.
\end{aligned}$$

Thus, the variance is obtained as

$$\mathrm{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12},$$

after some calculation.

Suppose now that $[a,b]=[0,1]$, and consider the function $g(x)=1$ if $x \le 1/3$, and $g(x)=2$ if $x>1/3$. The random variable $Y=g(X)$ is a discrete one with PMF $p_Y(1)=\mathbf{P}(X\le 1/3)=1/3$, $p_Y(2)=1-p_Y(1)=2/3$. Thus,

$$\mathbf{E}[Y]=\frac{1}{3}\cdot 1+\frac{2}{3}\cdot 2=\frac{5}{3}.$$

The same result could be obtained using the expected value rule:

$$\mathbf{E}[Y]=\int_0^1 g(x)f_X(x)\,dx=\int_0^{1/3} dx+\int_{1/3}^1 2\,dx=\frac{5}{3}.$$

Exponential Random Variable

An **exponential** random variable has a PDF of the form

$$f_X(x)=\begin{cases}\lambda e^{-\lambda x} & \text{if } x\ge 0,\\ 0 & \text{otherwise,}\end{cases}$$

where λ is a positive parameter characterizing the PDF (see Fig. 3.5). This is a legitimate PDF because

$$\int_{-\infty}^{\infty} f_X(x)\,dx=\int_0^{\infty}\lambda e^{-\lambda x}\,dx=-e^{-\lambda x}\Big|_0^{\infty}=1.$$

Note that the probability that X exceeds a certain value falls exponentially. Indeed, for any $a\ge 0$, we have

$$\mathbf{P}(X\ge a)=\int_a^{\infty}\lambda e^{-\lambda x}\,dx=-e^{-\lambda x}\Big|_a^{\infty}=e^{-\lambda a}.$$

An exponential random variable can be a very good model for the amount of time until a piece of equipment breaks down, until a light bulb burns out, or until an accident occurs. It will play a major role in our study of random processes in Chapter 5, but for the time being we will simply view it as an example of a random variable that is fairly tractable analytically.

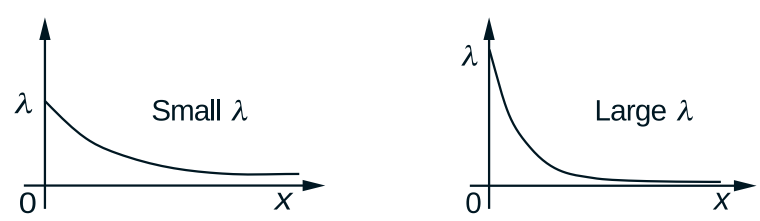

Figure 3.5: The PDF $\lambda e^{-\lambda x}$ of an exponential random variable.

The mean and the variance can be calculated to be

$$\mathbf{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{var}(X) = \frac{1}{\lambda^2}.$$

These formulas can be verified by straightforward calculation, as we now show. We have, using integration by parts,

$$\begin{aligned}
\mathbf{E}[X] &= \int_0^\infty x\lambda e^{-\lambda x}\,dx \\
&= \left(-xe^{-\lambda x}\right)\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx \\
&= 0 - \frac{e^{-\lambda x}}{\lambda}\Bigg|_0^\infty \\
&= \frac{1}{\lambda}.
\end{aligned}$$

Using again integration by parts, the second moment is

$$\begin{aligned}
\mathbf{E}[X^2] &= \int_0^\infty x^2\lambda e^{-\lambda x}\,dx \\
&= \left(-x^2e^{-\lambda x}\right)\Big|_0^\infty + \int_0^\infty 2xe^{-\lambda x}\,dx \\
&= 0 + \frac{2}{\lambda}\mathbf{E}[X] \\
&= \frac{2}{\lambda^2}.
\end{aligned}$$

Finally, using the formula $\mathrm{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2$, we obtain

$$\mathrm{var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Example 3.5. The time until a small meteorite first lands anywhere in the Sahara desert is modeled as an exponential random variable with a mean of 10 days. The time is currently midnight. What is the probability that a meteorite first lands some time between 6am and 6pm of the first day?

Let X be the time elapsed until the event of interest, measured in days. Then, X is exponential, with mean $1/\lambda = 10$, which yields $\lambda = 1/10$. The desired probability is

$$\mathbf{P}(1/4 \le X \le 3/4) = \mathbf{P}(X \ge 1/4) - \mathbf{P}(X > 3/4) = e^{-1/40} - e^{-3/40} = 0.0476,$$

where we have used the formula $\mathbf{P}(X \ge a) = \mathbf{P}(X > a) = e^{-\lambda a}$.

Let us also derive an expression for the probability that the time when a meteorite first lands will be between 6am and 6pm of some day. For the kth day, this set of times corresponds to the event $k - (3/4) \leq X \leq k - (1/4)$. Since these events are disjoint, the probability of interest is

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbf{P}\Big(k - \frac{3}{4} \leq X \leq k - \frac{1}{4}\Big) &= \sum_{k=1}^{\infty} \Big(\mathbf{P}\Big(X \geq k - \frac{3}{4}\Big) - \mathbf{P}\Big(X > k - \frac{1}{4}\Big)\Big) \\ &= \sum_{k=1}^{\infty} \big(e^{-(4k-3)/40} - e^{-(4k-1)/40}\big).\end{aligned}$$

We omit the remainder of the calculation, which involves using the geometric series formula.

3.2 CUMULATIVE DISTRIBUTION FUNCTIONS

We have been dealing with discrete and continuous random variables in a somewhat different manner, using PMFs and PDFs, respectively. It would be desirable to describe all kinds of random variables with a single mathematical concept. This is accomplished by the **cumulative distribution function**, or CDF for short. The CDF of a random variable X is denoted by F_X and provides the probability $\mathbf{P}(X \leq x)$. In particular, for every x we have

$$F_X(x) = \mathbf{P}(X \leq x) = \begin{cases} \displaystyle\sum_{k \leq x} p_X(k) & X\text{: discrete,} \\ \displaystyle\int_{-\infty}^{x} f_X(t)\,dt & X\text{: continuous.} \end{cases}$$

Loosely speaking, the CDF $F_X(x)$ "accumulates" probability "up to" the value x.

Any random variable associated with a given probability model has a CDF, regardless of whether it is discrete, continuous, or other. This is because $\{X \leq x\}$ is always an event and therefore has a well-defined probability. Figures 3.6 and 3.7 illustrate the CDFs of various discrete and continuous random variables. From these figures, as well as from the definition, some general properties of the CDF can be observed.

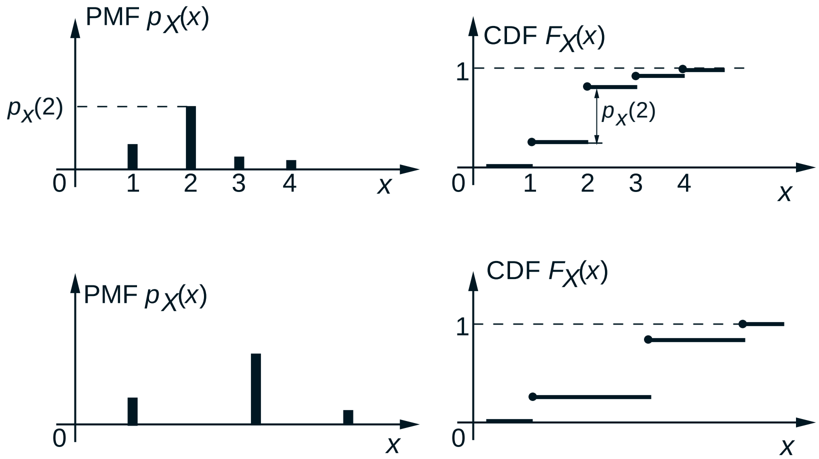

Figure 3.6: CDFs of some discrete random variables. The CDF is related to the PMF through the formula

$$F_X(x) = \mathbf{P}(X \leq x) = \sum_{k \leq x} p_X(k),$$

and has a staircase form, with jumps occurring at the values of positive probability mass. Note that at the points where a jump occurs, the value of F_X is the larger of the two corresponding values (i.e., F_X is continuous from the right).

Properties of a CDF

The CDF F_X of a random variable X is defined by

$$F_X(x) = \mathbf{P}(X \leq x), \qquad \text{for all } x,$$

and has the following properties.

- F_X is monotonically nondecreasing:

$$\text{if } x \leq y, \text{ then } F_X(x) \leq F_X(y).$$

- $F_X(x)$ tends to 0 as $x \to -\infty$, and to 1 as $x \to \infty$.
- If X is discrete, then F_X has a piecewise constant and staircase-like form.
- If X is continuous, then F_X has a continuously varying form.

- If X is discrete and takes integer values, the PMF and the CDF can be obtained from each other by summing or differencing:

$$F_X(k) = \sum_{i=-\infty}^{k} p_X(i),$$

$$p_X(k) = \mathbf{P}(X \le k) - \mathbf{P}(X \le k-1) = F_X(k) - F_X(k-1),$$

for all integers k.

- If X is continuous, the PDF and the CDF can be obtained from each other by integration or differentiation:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt,$$
$$f_X(x) = \frac{dF_X}{dx}(x).$$

(The latter relation is valid for those x for which the CDF has a derivative.)

Because the CDF is defined for any type of random variable, it provides a convenient means for exploring the relations between continuous and discrete random variables. This is illustrated in the following example, which shows that there is a close relation between the geometric and the exponential random variables.

Example 3.6. The Geometric and Exponential CDFs. Let X be a geometric random variable with parameter p; that is, X is the number of trials to obtain the first success in a sequence of independent Bernoulli trials, where the probability of success is p. Thus, for $k = 1, 2, \ldots$, we have $\mathbf{P}(X = k) = p(1-p)^{k-1}$ and the CDF is given by

$$F^{\text{geo}}(n) = \sum_{k=1}^{n} p(1-p)^{k-1} = p\frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n, \qquad \text{for } n = 1, 2, \ldots$$

Suppose now that X is an exponential random variable with parameter $\lambda > 0$. Its CDF is given by

$$F^{exp}(x) = \mathbf{P}(X \le x) = 0, \qquad \text{for } x \le 0,$$

and

$$F^{\text{exp}}(x) = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}, \qquad \text{for } x > 0.$$

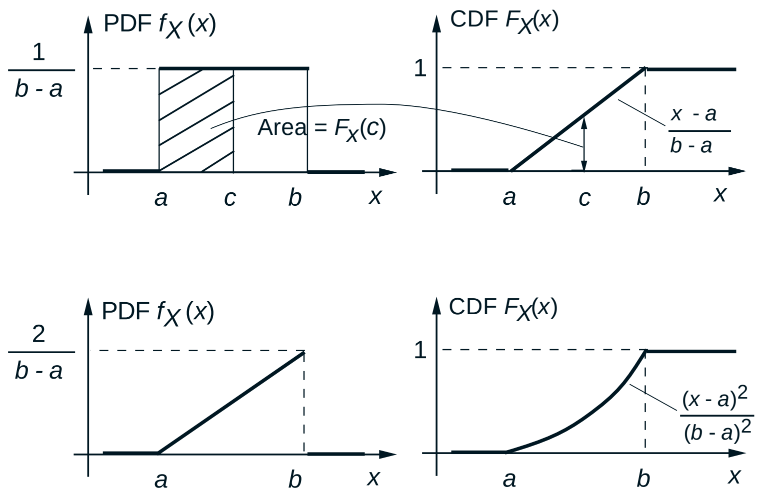

Figure 3.7: CDFs of some continuous random variables. The CDF is related to the PDF through the formula

$$F_X(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

Thus, the PDF f_X can be obtained from the CDF by differentiation:

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

For a continuous random variable, the CDF has no jumps, i.e., it is continuous.

To compare the two CDFs above, let $\delta = -\ln(1-p)/\lambda$, so that

$$e^{-\lambda\delta} = 1 - p.$$

Then we see that the values of the exponential and the geometric CDFs are equal for all $x = n\delta$, where $n = 1, 2, \ldots$, i.e.,

$$F^{\text{exp}}(n\delta) = F^{\text{geo}}(n), \qquad n = 1, 2, \ldots,$$

as illustrated in Fig. 3.8.

If δ is very small, there is close proximity of the exponential and the geometric CDFs, provided that we scale the values taken by the geometric random variable by δ. This relation is best interpreted by viewing X as time, either continuous, in the case of the exponential, or δ-discretized, in the case of the geometric. In particular, suppose that δ is a small number, and that every δ seconds, we flip a coin with the probability of heads being a small number p. Then, the time of the first occurrence of heads is well approximated by an exponential random variable. The parameter

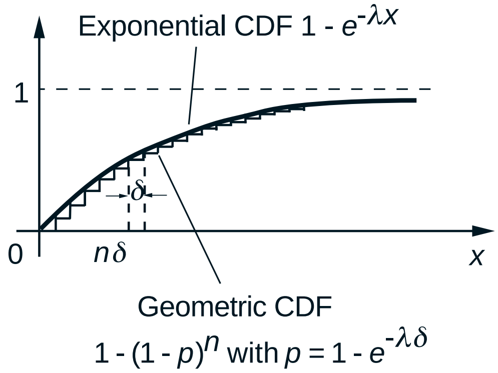

Figure 3.8: Relation of the geometric and the exponential CDFs. We have

$$F^{\exp}(n\delta) = F^{\text{geo}}(n), \qquad n = 1, 2, \ldots,$$

if the interval δ is such that $e^{-\lambda\delta} = 1 - p$. As δ approaches 0, the exponential random variable can be interpreted as the "limit" of the geometric.

λ of this exponential is such that $e^{-\lambda\delta} = 1 - p$ or $\lambda = -\ln(1-p)/\delta$. This relation between the geometric and the exponential random variables will play an important role in the theory of the Bernoulli and Poisson stochastic processes in Chapter 5.

Sometimes, in order to calculate the PMF or PDF of a discrete or continuous random variable, respectively, it is more convenient to first calculate the CDF and then use the preceding relations. The systematic use of this approach for the case of a continuous random variable will be discussed in Section 3.6. The following is a discrete example.

Example 3.7. The Maximum of Several Random Variables. You are allowed to take a certain test three times, and your final score will be the maximum of the test scores. Thus,

$$X = \max\{X_1, X_2, X_3\},$$

where X_1, X_2, X_3 are the three test scores and X is the final score. Assume that your score in each test takes one of the values from 1 to 10 with equal probability 1/10, independently of the scores in other tests. What is the PMF p_X of the final score?

We calculate the PMF indirectly. We first compute the CDF $F_X(k)$ and then obtain the PMF as

$$p_X(k) = F_X(k) - F_X(k-1), \qquad k = 1, \ldots, 10.$$

We have

$$\begin{aligned} F_X(k) &= \mathbf{P}(X \le k) \\ &= \mathbf{P}(X_1 \le k,\ X_2 \le k,\ X_3 \le k) \\ &= \mathbf{P}(X_1 \le k)\mathbf{P}(X_2 \le k)\mathbf{P}(X_3 \le k) \\ &= \left(\frac{k}{10}\right)^3, \end{aligned}$$

where the third equality follows from the independence of the events $\{X_1 \le k\}$, $\{X_2 \le k\}$, $\{X_3 \le k\}$. Thus the PMF is given by

$$p_X(k) = \left(\frac{k}{10}\right)^3 - \left(\frac{k-1}{10}\right)^3, \qquad k = 1, \ldots, 10.$$

3.3 NORMAL RANDOM VARIABLES

A continuous random variable X is said to be **normal** or **Gaussian** if it has a PDF of the form (see Fig. 3.9)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2},$$

where μ and σ are two scalar parameters characterizing the PDF, with σ assumed nonnegative. It can be verified that the normalization property

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\, dx = 1$$

holds (see the theoretical problems).

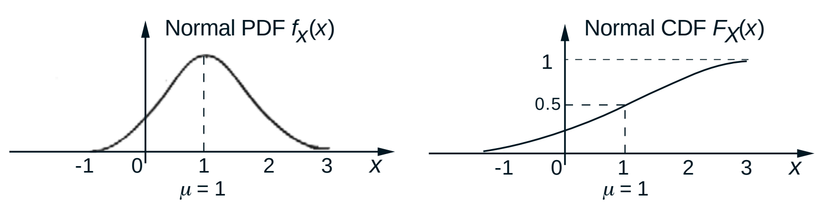

Figure 3.9: A normal PDF and CDF, with $\mu = 1$ and $\sigma^2 = 1$. We observe that the PDF is symmetric around its mean μ, and has a characteristic bell-shape. As x gets further from μ, the term $e^{-(x-\mu)^2/2\sigma^2}$ decreases very rapidly. In this figure, the PDF is very close to zero outside the interval $[-1, 3]$.

The mean and the variance can be calculated to be

$$\mathbf{E}[X] = \mu, \qquad \mathrm{var}(X) = \sigma^2.$$

To see this, note that the PDF is symmetric around μ, so its mean must be μ. Furthermore, the variance is given by

$$\mathrm{var}(X) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}\, dx.$$

Using the change of variables $y = (x-\mu)/\sigma$ and integration by parts, we have

$$\begin{aligned}
\mathrm{var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2}\, dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left(-y e^{-y^2/2}\right)\Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy \\
&= \sigma^2.
\end{aligned}$$

The last equality above is obtained by using the fact

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy = 1,$$

which is just the normalization property of the normal PDF for the case where $\mu = 0$ and $\sigma = 1$.

The normal random variable has several special properties. The following one is particularly important and will be justified in Section 3.6.

Normality is Preserved by Linear Transformations

If X is a normal random variable with mean μ and variance σ^2, and if a, b are scalars, then the random variable

$$Y = aX + b$$

is also normal, with mean and variance

$$\mathbf{E}[Y] = a\mu + b, \qquad \mathrm{var}(Y) = a^2\sigma^2.$$

The Standard Normal Random Variable

A normal random variable Y with zero mean and unit variance is said to be a **standard normal**. Its CDF is denoted by Φ,

$$\Phi(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(Y < y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2}\, dt.$$

It is recorded in a table (given in the next page), and is a very useful tool for calculating various probabilities involving normal random variables; see also Fig. 3.10.

Note that the table only provides the values of $\Phi(y)$ for $y \geq 0$, because the omitted values can be found using the symmetry of the PDF. For example, if Y is a standard normal random variable, we have

$$\begin{aligned}\Phi(-0.5) = \mathbf{P}(Y \leq -0.5) &= \mathbf{P}(Y \geq 0.5) = 1 - \mathbf{P}(Y < 0.5) \\ &= 1 - \Phi(0.5) = 1 - .6915 = 0.3085.\end{aligned}$$

Let X be a normal random variable with mean μ and variance σ^2. We "standardize" X by defining a new random variable Y given by

$$Y = \frac{X - \mu}{\sigma}.$$

Since Y is a linear transformation of X, it is normal. Furthermore,

$$\mathbf{E}[Y] = \frac{\mathbf{E}[X] - \mu}{\sigma} = 0, \qquad \text{var}(Y) = \frac{\text{var}(X)}{\sigma^2} = 1.$$

Thus, Y is a standard normal random variable. This fact allows us to calculate the probability of any event defined in terms of X: we redefine the event in terms of Y, and then use the standard normal table.

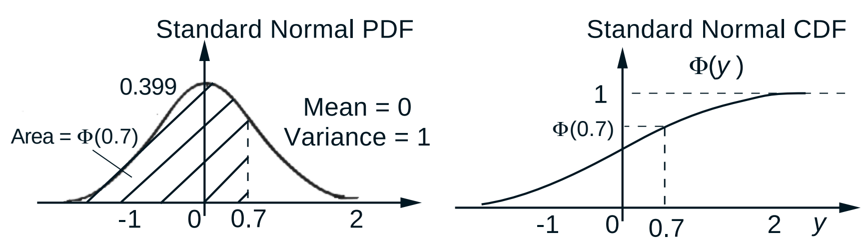

Figure 3.10: The PDF

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

of the standard normal random variable. Its corresponding CDF, which is denoted by $\Phi(y)$, is recorded in a table.

Example 3.8. Using the Normal Table. The annual snowfall at a particular geographic location is modeled as a normal random variable with a mean of $\mu = 60$ inches, and a standard deviation of $\sigma = 20$. What is the probability that this year's snowfall will be at least 80 inches?

Let X be the snow accumulation, viewed as a normal random variable, and let

$$Y = \frac{X - \mu}{\sigma} = \frac{X - 60}{20},$$

be the corresponding standard normal random variable. We want to find

$$\mathbf{P}(X \geq 80) = \mathbf{P}\left(\frac{X - 60}{20} \geq \frac{80 - 60}{20}\right) = \mathbf{P}\left(Y \geq \frac{80 - 60}{20}\right) = \mathbf{P}(Y \geq 1) = 1 - \Phi(1),$$

where Φ is the CDF of the standard normal. We read the value $\Phi(1)$ from the table:

$$\Phi(1) = 0.8413,$$

so that

$$\mathbf{P}(X \geq 80) = 1 - \Phi(1) = 0.1587.$$

Generalizing the approach in the preceding example, we have the following procedure.

CDF Calculation of the Normal Random Variable

The CDF of a normal random variable X with mean μ and variance σ^2 is obtained using the standard normal table as

$$\mathbf{P}(X \leq x) = \mathbf{P}\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \mathbf{P}\left(Y \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

where Y is a standard normal random variable.

The normal random variable is often used in signal processing and communications engineering to model noise and unpredictable distortions of signals. The following is a typical example.

Example 3.9. Signal Detection. A binary message is transmitted as a signal that is either -1 or $+1$. The communication channel corrupts the transmission with additive normal noise with mean $\mu = 0$ and variance σ^2. The receiver concludes that the signal -1 (or $+1$) was transmitted if the value received is < 0 (or ≥ 0, respectively); see Fig. 3.11. What is the probability of error?

An error occurs whenever -1 is transmitted and the noise N is at least 1 so that $N + S = N - 1 \geq 0$, or whenever $+1$ is transmitted and the noise N is smaller

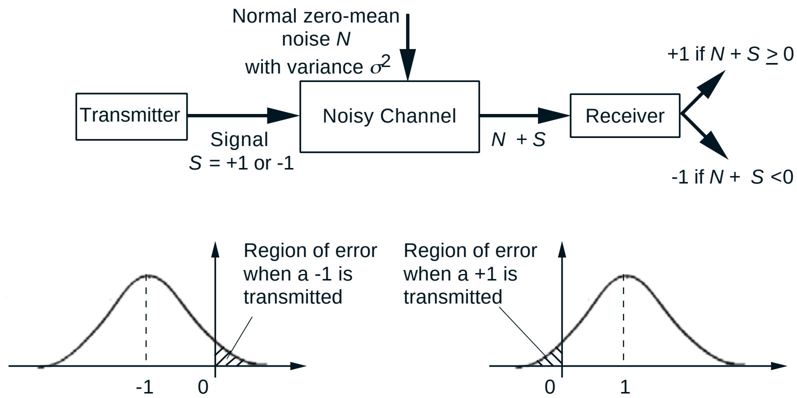

Figure 3.11: The signal detection scheme of Example 3.9. The area of the shaded region gives the probability of error in the two cases where -1 and $+1$ is transmitted.

than -1 so that $N+S=N+1<0$. In the former case, the probability of error is

$$\mathbf{P}(N \geq 1) = 1 - \mathbf{P}(N < 1) = 1 - \mathbf{P}\left(\frac{N-\mu}{\sigma} < \frac{1-\mu}{\sigma}\right)$$
$$= 1 - \Phi\left(\frac{1-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{1}{\sigma}\right).$$

In the latter case, the probability of error is the same, by symmetry. The value of $\Phi(1/\sigma)$ can be obtained from the normal table. For $\sigma = 1$, we have $\Phi(1/\sigma) = \Phi(1) = 0.8413$, and the probability of the error is 0.1587.

The normal random variable plays an important role in a broad range of probabilistic models. The main reason is that, generally speaking, it models well the additive effect of many independent factors, in a variety of engineering, physical, and statistical contexts. Mathematically, the key fact is that *the sum of a large number of independent and identically distributed (not necessarily normal) random variables has an approximately normal CDF, regardless of the CDF of the individual random variables.* This property is captured in the celebrated *central limit theorem*, which will be discussed in Chapter 7.

3.4 CONDITIONING ON AN EVENT

The **conditional PDF** of a continuous random variable X, conditioned on a particular event A with $\mathbf{P}(A) > 0$, is a function $f_{X|A}$ that satisfies

$$\mathbf{P}(X \in B \mid A) = \int_B f_{X|A}(x)\, dx,$$

for any subset B of the real line. It is the same as an ordinary PDF, except that it now refers to a new universe in which the event A is known to have occurred.

An important special case arises when we condition on X belonging to a subset A of the real line, with $\mathbf{P}(X \in A) > 0$. We then have

$$\mathbf{P}(X \in B \,|\, X \in A) = \frac{\mathbf{P}(X \in B \text{ and } X \in A)}{\mathbf{P}(X \in A)} = \frac{\displaystyle\int_{A \cap B} f_X(x)\,dx}{\mathbf{P}(X \in A)}.$$

This formula must agree with the earlier one, and therefore,†

$$f_{X|A}(x \,|\, A) = \begin{cases} \dfrac{f_X(x)}{\mathbf{P}(X \in A)} & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

As in the discrete case, the conditional PDF is zero outside the conditioning set. Within the conditioning set, the conditional PDF has exactly the same shape as the unconditional one, except that it is scaled by the constant factor $1/\mathbf{P}(X \in A)$. This normalization ensures that $f_{X|A}$ integrates to 1, which makes it a legitimate PDF; see Fig. 3.13.

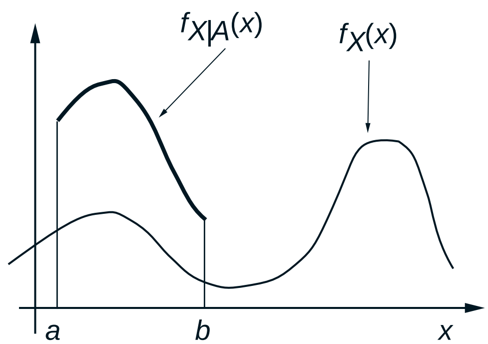

Figure 3.13: The unconditional PDF f_X and the conditional PDF $f_{X|A}$, where A is the interval $[a, b]$. Note that within the conditioning event A, $f_{X|A}$ retains the same shape as f_X, except that it is scaled along the vertical axis.

Example 3.10. The exponential random variable is memoryless. Alvin goes to a bus stop where the time T between two successive buses has an exponential PDF with parameter λ. Suppose that Alvin arrives t secs after the preceding bus arrival and let us express this fact with the event $A = \{T > t\}$. Let X be the time that Alvin has to wait for the next bus to arrive. What is the conditional CDF $F_{X|A}(x \,|\, A)$?

† We are using here the simpler notation $f_{X|A}(x)$ in place of $f_{X|X \in A}$, which is more accurate.

We have

$$\begin{aligned}\mathbf{P}(X > x \,|\, A) &= \mathbf{P}(T > t + x \,|\, T > t)\\ &= \frac{\mathbf{P}(T > t + x \text{ and } T > t)}{\mathbf{P}(T > t)}\\ &= \frac{\mathbf{P}(T > t + x)}{\mathbf{P}(T > t)}\\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}}\\ &= e^{-\lambda x},\end{aligned}$$

where we have used the expression for the CDF of an exponential random variable derived in Example 3.6.

Thus, the conditional CDF of X is exponential with parameter λ, regardless the time t that elapsed between the preceding bus arrival and Alvin's arrival. This is known as the *memorylessness property* of the exponential. Generally, if we model the time to complete a certain operation by an exponential random variable X, this property implies that as long as the operation has not been completed, the remaining time up to completion has the same exponential CDF, no matter when the operation started.

For a continuous random variable, the conditional expectation is defined similar to the unconditional case, except that we now need to use the conditional PDF. We summarize the discussion so far, together with some additional properties in the table that follows.

Conditional PDF and Expectation Given an Event

- The conditional PDF $f_{X|A}$ of a continuous random variable X given an event A with $\mathbf{P}(A) > 0$, satisfies

$$\mathbf{P}(X \in B \,|\, A) = \int_B f_{X|A}(x)\,dx.$$

- If A be a subset of the real line with $\mathbf{P}(X \in A) > 0$, then

$$f_{X|A}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbf{P}(X \in A)} & \text{if } x \in A,\\ 0 & \text{otherwise,}\end{cases}$$

  and

$$\mathbf{P}(X \in B \,|\, X \in A) = \int_B f_{X|A}(x)\,dx,$$

  for any set B.

- The corresponding conditional expectation is defined by

$$\mathbf{E}[X \mid A] = \int_{-\infty}^{\infty} x f_{X|A}(x)\, dx.$$

- The expected value rule remains valid:

$$\mathbf{E}\big[g(X) \mid A\big] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x)\, dx.$$

- If $A_1, A_2, \ldots, A_n$ are disjoint events with $\mathbf{P}(A_i) > 0$ for each i, that form a partition of the sample space, then

$$f_X(x) = \sum_{i=1}^{n} \mathbf{P}(A_i) f_{X|A_i}(x)$$

(a version of the total probability theorem), and

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{P}(A_i)\mathbf{E}[X \mid A_i]$$

(the total expectation theorem). Similarly,

$$\mathbf{E}\big[g(X)\big] = \sum_{i=1}^{n} \mathbf{P}(A_i)\mathbf{E}\big[g(X) \mid A_i\big].$$

To justify the above version of the total probability theorem, we use the total probability theorem from Chapter 1, to obtain

$$\mathbf{P}(X \le x) = \sum_{i=1}^{n} \mathbf{P}(A_i)\mathbf{P}(X \le x \mid A_i).$$

This formula can be rewritten as

$$\int_{-\infty}^{x} f_X(t)\, dt = \sum_{i=1}^{n} \mathbf{P}(A_i) \int_{-\infty}^{x} f_{X|A_i}(t)\, dt.$$

We take the derivative of both sides, with respect to x, and obtain the desired relation

$$f_X(x) = \sum_{i=1}^{n} \mathbf{P}(A_i) f_{X|A_i}(x).$$

If we now multiply both sides by x and then integrate from $-\infty$ to ∞, we obtain the total expectation theorem for continuous random variables.

The total expectation theorem can often facilitate the calculation of the mean, variance, and other moments of a random variable, using a divide-and-conquer approach.

Example 3.11. Mean and Variance of a Piecewise Constant PDF. Suppose that the random variable X has the piecewise constant PDF

$$f_X(x) = \begin{cases} 1/3 & \text{if } 0 \le x \le 1, \\ 2/3 & \text{if } 1 < x \le 2, \\ 0 & \text{otherwise,} \end{cases}$$

(see Fig. 3.14). Consider the events

$$A_1 = \big\{X \text{ lies in the first interval } [0,1]\big\},$$
$$A_2 = \big\{X \text{ lies in the second interval } (1,2]\big\}.$$

We have from the given PDF,

$$\mathbf{P}(A_1) = \int_0^1 f_X(x)\,dx = \frac{1}{3}, \qquad \mathbf{P}(A_2) = \int_1^2 f_X(x)\,dx = \frac{2}{3}.$$

Furthermore, the conditional mean and second moment of X, conditioned on A_1 and A_2, are easily calculated since the corresponding conditional PDFs $f_{X|A_1}$ and $f_{X|A_2}$ are uniform. We recall from Example 3.4 that the mean of a uniform random variable on an interval $[a,b]$ is $(a+b)/2$ and its second moment is $(a^2+ab+b^2)/3$. Thus,

$$\mathbf{E}[X \mid A_1] = \frac{1}{2}, \qquad \mathbf{E}[X \mid A_2] = \frac{3}{2},$$
$$\mathbf{E}\big[X^2 \mid A_1\big] = \frac{1}{3}, \qquad \mathbf{E}\big[X^2 \mid A_2\big] = \frac{7}{3}.$$

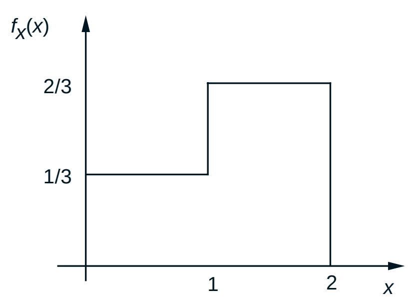

Figure 3.14: Piecewise constant PDF for Example 3.11.

We now use the total expectation theorem to obtain

$$\mathbf{E}[X] = \mathbf{P}(A_1)\mathbf{E}[X \mid A_1] + \mathbf{P}(A_2)\mathbf{E}[X \mid A_2] = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{3}{2} = \frac{7}{6},$$

$$\mathbf{E}[X^2] = \mathbf{P}(A_1)\mathbf{E}[X^2 \mid A_1] + \mathbf{P}(A_2)\mathbf{E}[X^2 \mid A_2] = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{7}{3} = \frac{15}{9}.$$

The variance is given by

$$\text{var}(X) = \mathbf{E}[X^2] - \big(\mathbf{E}[X]\big)^2 = \frac{15}{9} - \frac{49}{36} = \frac{11}{36}.$$

Note that this approach to the mean and variance calculation is easily generalized to piecewise constant PDFs with more than two pieces.

The next example illustrates a divide-and-conquer approach that uses the total probability theorem to calculate a PDF.

Example 3.12. The metro train arrives at the station near your home every quarter hour starting at 6:00 AM. You walk into the station every morning between 7:10 and 7:30 AM, with the time in this interval being a uniform random variable. What is the PDF of the time you have to wait for the first train to arrive?

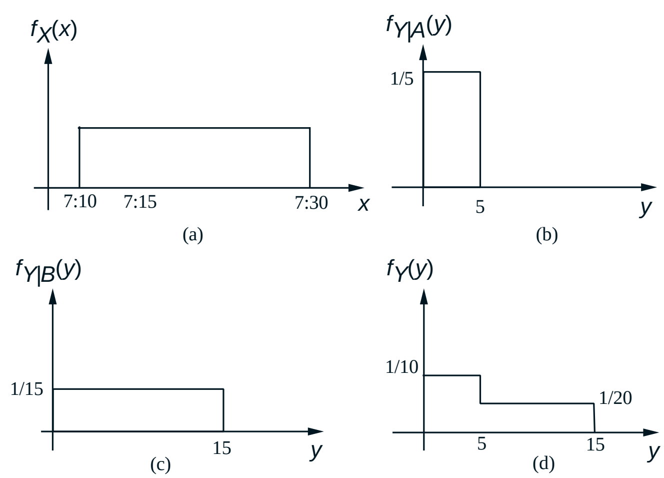

Figure 3.15: The PDFs f_X, $f_{Y|A}$, $f_{Y|B}$, and f_Y in Example 3.12.

The time of your arrival, denoted by X, is a uniform random variable on the interval from 7:10 to 7:30; see Fig. 3.15(a). Let Y be the waiting time. We calculate the PDF f_Y using a divide-and-conquer strategy. Let A and B be the events

$$A = \{7{:}10 \le X \le 7{:}15\} = \{\text{you board the 7:15 train}\},$$

$$B = \{7{:}15 < X \le 7{:}30\} = \{\text{you board the 7:30 train}\}.$$

Conditioned on the event A, your arrival time is uniform on the interval from 7:10 to 7:15. In that case, the waiting time Y is also uniform and takes values between 0 and 5 minutes; see Fig. 3.15(b). Similarly, conditioned on B, Y is uniform and takes values between 0 and 15 minutes; see Fig. 3.15(c). The PDF of Y is obtained using the total probability theorem,

$$f_Y(y) = \mathbf{P}(A) f_{Y|A}(y) + \mathbf{P}(B) f_{Y|B}(y),$$

and is shown in Fig. 3.15(d). In particular,

$$f_Y(y) = \frac{1}{4} \cdot \frac{1}{5} + \frac{3}{4} \cdot \frac{1}{15} = \frac{1}{10}, \quad \text{for } 0 \le y \le 5,$$

and

$$f_Y(y) = \frac{1}{4} \cdot 0 + \frac{3}{4} \cdot \frac{1}{15} = \frac{1}{20}, \quad \text{for } 5 < y \le 15.$$

3.5 MULTIPLE CONTINUOUS RANDOM VARIABLES

We will now extend the notion of a PDF to the case of multiple random variables. In complete analogy with discrete random variables, we introduce joint, marginal, and conditional PDFs. Their intuitive interpretation as well as their main properties parallel the discrete case.

We say that two continuous random variables associated with a common experiment are **jointly continuous** and can be described in terms of a **joint PDF** $f_{X,Y}$, if $f_{X,Y}$ is a nonnegative function that satisfies

$$\mathbf{P}\big((X, Y) \in B\big) = \int\!\!\int_{(x,y)\in B} f_{X,Y}(x, y)\, dx\, dy,$$

for every subset B of the two-dimensional plane. The notation above means that the integration is carried over the set B. In the particular case where B is a rectangle of the form $B = [a, b] \times [c, d]$, we have

$$\mathbf{P}(a \le X \le b,\, c \le Y \le d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dx\, dy.$$

Furthermore, by letting B be the entire two-dimensional plane, we obtain the normalization property

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1.$$

To interpret the PDF, we let δ be very small and consider the probability of a small rectangle. We have

$$\mathbf{P}(a \le X \le a+\delta,\, c \le Y \le c+\delta) = \int_c^{c+\delta} \int_a^{a+\delta} f_{X,Y}(x,y)\,dx\,dy \approx f_{X,Y}(a,c)\cdot\delta^2,$$

so we can view $f_{X,Y}(a,c)$ as the "probability per unit area" in the vicinity of (a,c).

The joint PDF contains all conceivable probabilistic information on the random variables X and Y, as well as their dependencies. It allows us to calculate the probability of any event that can be defined in terms of these two random variables. As a special case, it can be used to calculate the probability of an event involving only one of them. For example, let A be a subset of the real line and consider the event $\{X \in A\}$. We have

$$\mathbf{P}(X \in A) = \mathbf{P}\big(X \in A \text{ and } Y \in (-\infty,\infty)\big) = \int_A \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx.$$

Comparing with the formula

$$\mathbf{P}(X \in A) = \int_A f_X(x)\,dx,$$

we see that the **marginal** PDF f_X of X is given by

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

Similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

Example 3.13. Two-Dimensional Uniform PDF. Romeo and Juliet have a date at a given time, and each will arrive at the meeting place with a delay between 0 and 1 hour (recall the example given in Section 1.2). Let X and Y denote the delays of Romeo and Juliet, respectively. Assuming that no pairs (x,y) in the square $[0,1]\times[0,1]$ are more likely than others, a natural model involves a joint PDF of the form

$$f_{X,Y}(x,y) = \begin{cases} c & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. For this PDF to satisfy the normalization property

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_0^1 c\,dx\,dy = 1,$$

we must have

$$c = 1.$$

This is an example of a uniform PDF on the unit square. More generally, let us fix some subset S of the two-dimensional plane. The corresponding uniform joint PDF on S is defined to be

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{area of } S} & \text{if } (x,y) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For any set $A \subset S$, the probability that the experimental value of (X,Y) lies in A is

$$\mathbf{P}\big((X,Y) \in A\big) = \int\!\!\int_{(x,y)\in A} f_{X,Y}(x,y)\,dx\,dy = \frac{1}{\text{area of } S} \int\!\!\int_{(x,y)\in A\cap S} dx\,dy = \frac{\text{area of } A \cap S}{\text{area of } S}.$$

Example 3.14. We are told that the joint PDF of the random variables X and Y is a constant c on the set S shown in Fig. 3.16 and is zero outside. Find the value of c and the marginal PDFs of X and Y.

The area of the set S is equal to 4 and, therefore, $f_{X,Y}(x,y) = c = 1/4$, for $(x,y) \in S$. To find the marginal PDF $f_X(x)$ for some particular x, we integrate (with respect to y) the joint PDF over the vertical line corresponding to that x. The resulting PDF is shown in the figure. We can compute f_Y similarly.

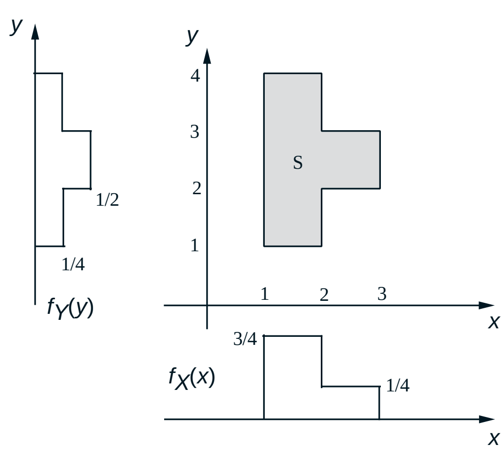

Figure 3.16: The joint PDF in Example 3.14 and the resulting marginal PDFs.

Example 3.15. Buffon's Needle. This is a famous example, which marks the origin of the subject of geometrical probability, that is, the analysis of the geometrical configuration of randomly placed objects.

A surface is ruled with parallel lines, which are at distance d from each other (see Fig. 3.17). Suppose that we throw a needle of length l on the surface at random. What is the probability that the needle will intersect one of the lines?

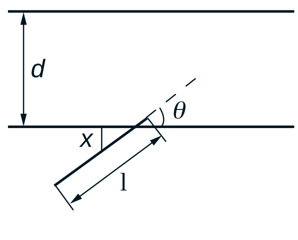

Figure 3.17: Buffon's needle. The length of the line segment between the midpoint of the needle and the point of intersection of the axis of the needle with the closest parallel line is $x/\sin\theta$. The needle will intersect the closest parallel line if and only if this length is less than $l/2$.

We assume here that $l < d$ so that the needle cannot intersect two lines simultaneously. Let X be the distance from the midpoint of the needle to the nearest of the parallel lines, and let Θ be the acute angle formed by the axis of the needle and the parallel lines (see Fig. 3.17). We model the pair of random variables (X, Θ) with a uniform joint PDF over the rectangle $[0, d/2] \times [0, \pi/2]$, so that

$$f_{X,\Theta}(x,\theta) = \begin{cases} 4/(\pi d) & \text{if } x \in [0, d/2] \text{ and } \theta \in [0, \pi/2], \\ 0 & \text{otherwise.} \end{cases}$$

As can be seen from Fig. 3.17, the needle will intersect one of the lines if and only if

$$X \le \frac{l}{2}\sin\Theta,$$

so the probability of intersection is

$$\begin{aligned} \mathbf{P}\big(X \le (l/2)\sin\Theta\big) &= \int\!\!\int_{x \le (l/2)\sin\theta} f_{X,\Theta}(x,\theta)\,dx\,d\theta \\ &= \frac{4}{\pi d}\int_0^{\pi/2}\int_0^{(l/2)\sin\theta} dx\,d\theta \\ &= \frac{4}{\pi d}\int_0^{\pi/2}\frac{l}{2}\sin\theta\,d\theta \\ &= \frac{2l}{\pi d}(-\cos\theta)\Big|_0^{\pi/2} \\ &= \frac{2l}{\pi d}. \end{aligned}$$

The probability of intersection can be empirically estimated, by repeating the experiment a large number of times. Since it is equal to $2l/\pi d$, this provides us with a method for the experimental evaluation of π.

Expectation

If X and Y are jointly continuous random variables, and g is some function, then $Z = g(X,Y)$ is also a random variable. We will see in Section 3.6 methods for computing the PDF of Z, if it has one. For now, let us note that the expected value rule is still applicable and

$$\mathbf{E}\big[g(X,Y)\big] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy.$$

As an important special case, for any scalars a, b, we have

$$\mathbf{E}[aX+bY] = a\mathbf{E}[X] + b\mathbf{E}[Y].$$

Conditioning One Random Variable on Another

Let X and Y be continuous random variables with joint PDF $f_{X,Y}$. For any fixed y with $f_Y(y) > 0$, the conditional PDF of X given that $Y = y$, is defined by

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

This definition is analogous to the formula $p_{X|Y} = p_{X,Y}/p_Y$ for the discrete case.

When thinking about the conditional PDF, it is best to view y as a fixed number and consider $f_{X|Y}(x\,|\,y)$ as a function of the single variable x. As a function of x, the conditional PDF $f_{X|Y}(x\,|\,y)$ has the same shape as the joint PDF $f_{X,Y}(x,y)$, because the normalizing factor $f_Y(y)$ does not depend on x; see Fig. 3.18. Note that the normalization ensures that

$$\int_{-\infty}^{\infty} f_{X|Y}(x\,|\,y)\,dx = 1,$$

so *for any fixed* y, $f_{X|Y}(x\,|\,y)$ is a legitimate PDF.

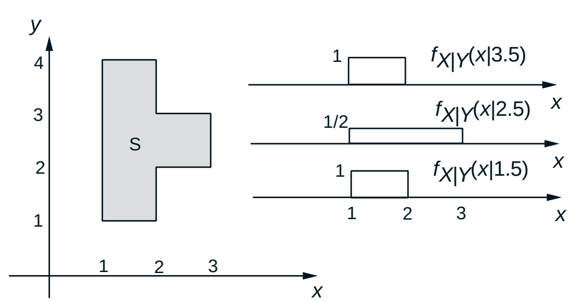

Figure 3.18: Visualization of the conditional PDF $f_{X|Y}(x\,|\,y)$. Let X, Y have a joint PDF which is uniform on the set S. For each fixed y, we consider the joint PDF along the slice $Y = y$ and normalize it so that it integrates to 1.

Example 3.16. Circular Uniform PDF. John throws a dart at a circular target of radius r (see Fig. 3.19). We assume that he always hits the target, and that all points of impact (x, y) are equally likely, so that the joint PDF of the random variables X and Y is uniform. Following Example 3.13, and since the area of the circle is πr^2, we have

$$
\begin{aligned}
f_{X,Y}(x,y) &= \begin{cases} \dfrac{1}{\text{area of the circle}} & \text{if } (x,y) \text{ is in the circle,} \\ 0 & \text{otherwise,} \end{cases} \\
&= \begin{cases} \dfrac{1}{\pi r^2} & \text{if } x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

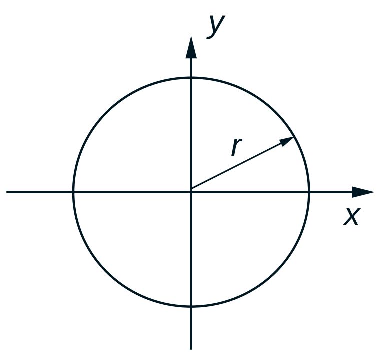

Figure 3.19: Circular target for Example 3.16.

To calculate the conditional PDF $f_{X|Y}(x \mid y)$, let us first calculate the marginal PDF $f_Y(y)$. For $|y| > r$, it is zero. For $|y| \leq r$, it can be calculated as follows:

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx \\
&= \frac{1}{\pi r^2} \int_{x^2+y^2 \leq r^2} dx \\
&= \frac{1}{\pi r^2} \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx \\
&= \frac{2}{\pi r^2} \sqrt{r^2 - y^2}.
\end{aligned}
$$

Note that the marginal $f_Y(y)$ is not a uniform PDF.

The conditional PDF is

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\dfrac{1}{\pi r^2}}{\dfrac{2}{\pi r^2}\sqrt{r^2-y^2}} = \frac{1}{2\sqrt{r^2-y^2}}.$$

Thus, for a fixed value of y, the conditional PDF $f_{X|Y}$ is uniform.

To interpret the conditional PDF, let us fix some small positive numbers δ_1 and δ_2, and condition on the event $B = \{y \le Y \le y+\delta_2\}$. We have

$$\mathbf{P}(x \le X \le x+\delta_1 \,|\, y \le Y \le y+\delta_2) = \frac{\mathbf{P}(x \le X \le x+\delta_1 \text{ and } y \le Y \le y+\delta_2)}{\mathbf{P}(y \le Y \le y+\delta_2)} \approx \frac{f_{X,Y}(x,y)\delta_1\delta_2}{f_Y(y)\delta_2} = f_{X|Y}(x\,|\,y)\delta_1.$$

In words, $f_{X|Y}(x\,|\,y)\delta_1$ provides us with the probability that X belongs in a small interval $[x, x+\delta_1]$, given that Y belongs in a small interval $[y, y+\delta_2]$. Since $f_{X|Y}(x\,|\,y)\delta_1$ does not depend on δ_2, we can think of the limiting case where δ_2 decreases to zero and write

$$\mathbf{P}(x \le X \le x+\delta_1 \,|\, Y=y) \approx f_{X|Y}(x\,|\,y)\delta_1, \qquad (\delta_1 \text{ small}),$$

and, more generally,

$$\mathbf{P}(X \in A \,|\, Y=y) = \int_A f_{X|Y}(x\,|\,y)\,dx.$$

Conditional probabilities, given the zero probability event $\{Y = y\}$, were left undefined in Chapter 1. But the above formula provides a natural way of defining such conditional probabilities in the present context. In addition, it allows us to view the conditional PDF $f_{X|Y}(x\,|\,y)$ (as a function of x) as a description of the probability law of X, given that the event $\{Y = y\}$ has occurred.

As in the discrete case, the conditional PDF $f_{X|Y}$, together with the marginal PDF f_Y are sometimes used to calculate the joint PDF. Furthermore, this approach can be also used for modeling: instead of directly specifying $f_{X,Y}$, it is often natural to provide a probability law for Y, in terms of a PDF f_Y, and then provide a conditional probability law $f_{X|Y}(x,y)$ for X, given any possible value y of Y.

Example 3.17. Let X be exponentially distributed with mean 1. Once we observe the experimental value x of X, we generate a normal random variable Y with zero mean and variance $x+1$. What is the joint PDF of X and Y?

We have $f_X(x) = e^{-x}$, for $x \geq 0$, and

$$f_{Y|X}(y\,|\,x) = \frac{1}{\sqrt{2\pi(x+1)}}e^{-y^2/2(x+1)}.$$

Thus,

$$f_{X,Y}(x,y) = f_X(x)f_{Y|X}(y\,|\,x) = e^{-x}\frac{1}{\sqrt{2\pi(x+1)}}e^{-y^2/2(x+1)},$$

for all $x \geq 0$ and all y.

Having defined a conditional probability law, we can also define a corresponding conditional expectation by letting

$$\mathbf{E}[X\,|\,Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x\,|\,y)\,dx.$$

The properties of (unconditional) expectation carry though, with the obvious modifications, to conditional expectation. For example the conditional version of the expected value rule

$$\mathbf{E}[g(X)\,|\,Y=y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x\,|\,y)\,dx$$

remains valid.

Summary of Facts About Multiple Continuous Random Variables

Let X and Y be jointly continuous random variables with joint PDF $f_{X,Y}$.

- The joint, marginal, and conditional PDFs are related to each other by the formulas

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

  The conditional PDF $f_{X|Y}(x\,|\,y)$ is defined only for those y for which $f_Y(y) > 0$.

- They can be used to calculate probabilities:

$$\mathbf{P}\big((X,Y)\in B\big) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy,$$
$$\mathbf{P}(X\in A) = \int_A f_X(x)\,dx,$$
$$\mathbf{P}(X\in A\,|\,Y=y) = \int_A f_{X|Y}(x\,|\,y)\,dx.$$

- They can also be used to calculate expectations:

$$\mathbf{E}[g(X)] = \int g(x)f_X(x)\,dx,$$
$$\mathbf{E}\big[g(X,Y)\big] = \int\int g(x,y)f_{X,Y}(x,y)\,dx\,dy,$$
$$\mathbf{E}\big[g(X)\,|\,Y=y\big] = \int g(x)f_{X|Y}(x\,|\,y)\,dx,$$
$$\mathbf{E}\big[g(X,Y)\,|\,Y=y\big] = \int g(x,y)f_{X|Y}(x\,|\,y)\,dx.$$

- We have the following versions of the total expectation theorem:

$$\mathbf{E}[X] = \int \mathbf{E}[X\,|\,Y=y]f_Y(y)\,dy,$$
$$\mathbf{E}\big[g(X)\big] = \int \mathbf{E}\big[g(X)\,|\,Y=y\big]f_Y(y)\,dy,$$
$$\mathbf{E}\big[g(X,Y)\big] = \int \mathbf{E}\big[g(X,Y)\,|\,Y=y\big]f_Y(y)\,dy.$$

To justify the first version of the total expectation theorem, we observe that

$$\begin{aligned}\int \mathbf{E}[X\,|\,Y=y]f_Y(y)\,dy &= \int\left[\int xf_{X|Y}(x\,|\,y)\,dx\right]f_Y(y)\,dy\\ &= \int\int xf_{X|Y}(x\,|\,y)f_Y(y)\,dx\,dy\\ &= \int\int xf_{X,Y}(x,y)\,dx\,dy\end{aligned}$$

$$
\begin{aligned}
&= \int x \left[\int f_{X,Y}(x, y)\, dy \right] dx \\
&= \int x f_X(x)\, dx \\
&= \mathbf{E}[X].
\end{aligned}
$$

The other two versions are justified similarly.

Inference and the Continuous Bayes' Rule

In many situations, we have a model of an underlying but unobserved phenomenon, represented by a random variable X with PDF f_X, and we make noisy measurements Y. The measurements are supposed to provide information about X and are modeled in terms of a conditional PDF $f_{Y|X}$. For example, if Y is the same as X, but corrupted by zero-mean normally distributed noise, one would let the conditional PDF $f_{Y|X}(y\,|\,x)$ of Y, given that $X = x$, be normal with mean equal to x. Once the experimental value of Y is measured, what information does this provide on the unknown value of X?

This setting is similar to that encountered in Section 1.4, when we introduced the Bayes rule and used it to solve inference problems. The only difference is that we are now dealing with continuous random variables.

Note that the information provided by the event $\{Y = y\}$ is described by the conditional PDF $f_{X|Y}(x\,|\,y)$. It thus suffices to evaluate the latter PDF. A calculation analogous to the original derivation of the Bayes' rule, based on the formulas $f_X f_{Y|X} = f_{X,Y} = f_Y f_{X|Y}$, yields

$$
f_{X|Y}(x\,|\,y) = \frac{f_X(x) f_{Y|X}(y\,|\,x)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y\,|\,x)}{\displaystyle\int f_X(t) f_{Y|X}(y\,|\,t) dt},
$$

which is the desired formula.

Example 3.18. A lightbulb produced by the General Illumination Company is known to have an exponentially distributed lifetime Y. However, the company has been experiencing quality control problems. On any given day, the parameter λ of the PDF of Y is actually a random variable, uniformly distributed in the interval $[0, 1/2]$. We test a lightbulb and record the experimental value y of its lifetime. What can we say about the underlying parameter λ?

We model the parameter λ as a random variable X, with a uniform distribution. All available information about X is contained in the conditional PDF $f_{X|X}(x\,|\,y)$. We view y as a constant (equal to the observed value of Y) and concentrate on the dependence of the PDF on x. Note that $f_X(x) = 2$, for $0 \le x \le 1/2$. By the continuous Bayes rule, we have

$$
f_{X|Y}(x\,|\,y) = \frac{2xe^{-xy}}{\int_0^{1/2} 2te^{-ty} dt}, \qquad \text{for } 0 \le x \le \frac{1}{2}.
$$

In some cases, the unobserved phenomenon is inherently discrete. For example, if a binary signal is observed in the presence of noise with a normal distribution. Or if a medical diagnosis is to be made on the basis of continuous measurements like temperature and blood counts. In such cases, a somewhat different version of Bayes' rule applies.

Let X be a discrete random variable that takes values in a finite set $\{1, \ldots, n\}$ and which represents the different discrete possibilities for the unobserved phenomenon of interest. The PMF p_X of X is assumed to be known. Let Y be a continuous random variable which, for any given value x, is described by a conditional PDF $f_{Y\,|\,X}(y\,|\,x)$. We are interested in the conditional PMF of X given the experimental value y of Y.

Instead of working with conditioning event $\{Y = y\}$ which has zero probability, let us instead condition on the event $\{y \le Y \le y + \delta\}$, where δ is a small positive number, and then take the limit as δ tends to zero. We have, using the Bayes rule

$$\begin{aligned}
\mathbf{P}(X = x\,|\,Y = y) &\approx \mathbf{P}(X = x\,|\,y \le Y \le y + \delta) \\
&= \frac{p_X(x)\mathbf{P}(y \le Y \le y + \delta\,|\,X = x)}{\mathbf{P}(y \le Y \le y + \delta)} \\
&\approx \frac{p_X(x) f_{Y|X}(y\,|\,x)\delta}{f_Y(y)\delta} \\
&= \frac{p_X(x) f_{Y|X}(y\,|\,x)}{f_Y(y)}.
\end{aligned}$$

The denominator can be evaluated using a version of the total probability theorem introduced in Section 3.4. We have

$$f_Y(y) = \sum_{i=1}^{n} p_X(i) f_{Y|X}(y\,|\,i).$$

Example 3.19. Let us revisit the signal detection problem considered in 3.9. A signal S is transmitted and we are given that $\mathbf{P}(S = 1) = p$ and $\mathbf{P}(S = -1) = 1-p$. The received signal is $Y = N+S$, where N is zero mean normal noise, with variance σ^2, independent of S. What is the probability that $S = 1$, as a function of the observed value y of Y?

Conditioned on $S = s$, the random variable Y has a normal distribution with mean s and variance σ^2. Applying the formula developed above, we obtain

$$\mathbf{P}(S = 1\,|\,Y = y) = \frac{p_S(1) f_{Y|S}(y\,|\,1)}{f_Y(y)} = \frac{\frac{p}{\sqrt{2\pi}\,\sigma} e^{-(y-1)^2/2\sigma^2}}{\frac{p}{\sqrt{2\pi}\,\sigma} e^{-(y-1)^2/2\sigma^2} + \frac{1-p}{\sqrt{2\pi}\,\sigma} e^{-(y+1)^2/2\sigma^2}}.$$

Independence

In full analogy with the discrete case, we say that two continuous random variables X and Y are **independent** if their joint PDF is the product of the marginal PDFs:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \qquad \text{for all } x, y.$$

Comparing with the formula $f_{X,Y}(x,y) = f_{X|Y}(x \,|\, y) f_Y(y)$, we see that independence is the same as the condition

$$f_{X|Y}(x \,|\, y) = f_X(x), \qquad \text{for all } x \text{ and all } y \text{ with } f_Y(y) > 0,$$

or, symmetrically,

$$f_{Y|X}(y \,|\, x) = f_Y(y), \qquad \text{for all } y \text{ and all } x \text{ with } f_X(x) > 0.$$

If X and Y are independent, then any two events of the form $\{X \in A\}$ and $\{Y \in B\}$ are independent. Indeed,

$$\begin{aligned}
\mathbf{P}(X \in A \text{ and } Y \in B) &= \int_{x \in A} \int_{y \in B} f_{X,Y}(x,y)\, dy\, dx \\
&= \int_{x \in A} \int_{y \in B} f_X(x) f_Y(y)\, dy\, dx \\
&= \int_{x \in A} f_X(x)\, dx \int_{y \in B} f_Y(y)\, dy \\
&= \mathbf{P}(X \in A)\mathbf{P}(Y \in B).
\end{aligned}$$

A converse statement is also true; see the theoretical problems.

A calculation similar to the discrete case shows that if X and Y are independent, then

$$\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)],$$

for any two functions g and h. Finally, the variance of the sum of *independent* random variables is again equal to the sum of the variances.

Independence of Continuous Random Variables

Suppose that X and Y are independent, that is,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \qquad \text{for all } x, y.$$

We then have the following properties.

- The random variables $g(X)$ and $h(Y)$ are independent, for any functions g and h.
- We have
$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y],$$
and, more generally,
$$\mathbf{E}\big[g(X)h(Y)\big] = \mathbf{E}\big[g(X)\big]\mathbf{E}\big[h(Y)\big],$$
- We have
$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y).$$

Joint CDFs

If X and Y are two random variables associated with the same experiment, we define their joint CDF by

$$F_{X,Y}(x,y) = \mathbf{P}(X \le x,\, Y \le y).$$

As in the case of one random variable, the advantage of working with the CDF is that it applies equally well to discrete and continuous random variables. In particular, if X and Y are described by a joint PDF $f_{X,Y}$, then

$$F_{X,Y}(x,y) = \mathbf{P}(X \le x,\, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)\,ds\,dt.$$

Conversely, the PDF can be recovered from the PDF by differentiating:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y).$$

Example 3.20. Let X and Y be described by a uniform PDF on the unit square. The joint CDF is given by

$$F_{X,Y}(x,y) = \mathbf{P}(X \le x,\, Y \le y) = xy, \qquad \text{for } 0 \le x, y \le 1.$$

We then verify that

$$\frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = \frac{\partial^2 (xy)}{\partial x \partial y}(x,y) = 1 = f_{X,Y}(x,y),$$

for all (x,y) in the unit square.

More than Two Random Variables

The joint PDF of three random variables X, Y, and Z is defined in analogy with the above. For example, we have

$$\mathbf{P}\big((X,Y,Z)\in B\big) = \int\int\int_{(x,y,z)\in B} f_{X,Y,Z}(x,y,z)\,dx\,dy\,dz,$$

for any set B. We also have relations such as

$$f_{X,Y}(x,y) = \int f_{X,Y,Z}(x,y,z)\,dz,$$

and

$$f_X(x) = \int\int f_{X,Y,Z}(x,y,z)\,dy\,dz.$$

One can also define conditional PDFs by formulas such as

$$f_{X,Y|Z}(x,y\,|\,z) = \frac{f_{X,Y,Z}(x,y,z)}{f_Z(z)}, \qquad \text{for } f_Z(z) > 0,$$

$$f_{X|Y,Z}(x\,|\,y,z) = \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}, \qquad \text{for } f_{Y,Z}(y,z) > 0.$$

There is an analog of the multiplication rule:

$$f_{X,Y,Z}(x,y,z) = f_{X|Y,Z}(x\,|\,y,z) f_{Y|Z}(y\,|\,z) f_Z(z).$$

Finally, we say that the three random variables X, Y, and Z are independent if

$$f_{X,Y,Z}(x,y,z) = f_X(x) f_Y(y) f_Z(z), \qquad \text{for all } x, y, z.$$

The expected value rule for functions takes the form

$$\mathbf{E}\big[g(X,Y,Z)\big] = \int\int\int g(x,y,z) f_{X,Y,Z}(x,y,z)\,dx\,dy\,dz,$$

and if g is linear and of the form $aX + bY + cZ$, then

$$\mathbf{E}[aX + bY + cZ] = a\mathbf{E}[X] + b\mathbf{E}[Y] + c\mathbf{E}[Z].$$

Furthermore, there are obvious generalizations of the above to the case of more than three random variables. For example, for any random variables $X_1, X_2, \ldots, X_n$ and any scalars $a_1, a_2, \ldots, a_n$, we have

$$\mathbf{E}[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1\mathbf{E}[X_1] + a_2\mathbf{E}[X_2] + \cdots + a_n\mathbf{E}[X_n].$$

3.6 DERIVED DISTRIBUTIONS

We have seen that the mean of a function $Y = g(X)$ of a continuous random variable X, can be calculated using the expected value rule

$$\mathbf{E}[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx,$$

without first finding the PDF f_Y of Y. Still, in some cases, we may be interested in an explicit formula for f_Y. Then, the following two-step approach can be used.

Calculation of the PDF of a Function $Y = g(X)$ of a Continuous Random Variable X

1. Calculate the CDF F_Y of Y using the formula

$$F_Y(y) = \mathbf{P}\big(g(X) \leq y\big) = \int_{\{x \mid g(x) \leq y\}} f_X(x)\, dx.$$

2. Differentiate to obtain the PDF of Y:

$$f_Y(y) = \frac{dF_Y}{dy}(y).$$

Example 3.21. Let X be uniform on $[0, 1]$. Find the PDF of $Y = \sqrt{X}$. Note that Y takes values between 0 and 1. For every $y \in [0, 1]$, we have

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(\sqrt{X} \leq y) = \mathbf{P}(X \leq y^2) = y^2, \quad 0 \leq y \leq 1.$$

We then differentiate and obtain

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{d(y^2)}{dy} = 2y, \qquad 0 \leq y \leq 1.$$

Outside the range $[0, 1]$, the CDF $F_Y(y)$ is constant, with $F_Y(y) = 0$ for $y \leq 0$, and $F_Y(y) = 1$ for $y \geq 1$. By differentiating, we see that $f_Y(y) = 0$ for y outside $[0, 1]$.

Example 3.22. John Slow is driving from Boston to the New York area, a distance of 180 miles. His average speed is uniformly distributed between 30 and 60 miles per hour. What is the PDF of the duration of the trip?

Let X be the speed and let $Y = g(X)$ be the trip duration:

$$g(X) = \frac{180}{X}.$$

To find the CDF of Y, we must calculate

$$\mathbf{P}(Y \le y) = \mathbf{P}\left(\frac{180}{X} \le y\right) = \mathbf{P}\left(\frac{180}{y} \le X\right).$$

We use the given uniform PDF of X, which is

$$f_X(x) = \begin{cases} 1/30 & \text{if } 30 \le x \le 60, \\ 0 & \text{otherwise,} \end{cases}$$

and the corresponding CDF, which is

$$F_X(x) = \begin{cases} 0 & \text{if } x \le 30, \\ (x-30)/30 & \text{if } 30 \le x \le 60, \\ 1 & \text{if } 60 \le x. \end{cases}$$

Thus,

$$\begin{aligned} F_Y(y) &= \mathbf{P}\left(\frac{180}{y} \le X\right) \\ &= 1 - F_X\left(\frac{180}{y}\right) \\ &= \begin{cases} 0 & \text{if } y \le 180/60, \\ 1 - \dfrac{\dfrac{180}{y} - 30}{30} & \text{if } 180/60 \le y \le 180/30, \\ 1 & \text{if } 180/30 \le y, \end{cases} \\ &= \begin{cases} 0 & \text{if } y \le 3, \\ 2 - (6/y) & \text{if } 3 \le y \le 6, \\ 1 & \text{if } 6 \le y, \end{cases} \end{aligned}$$

(see Fig. 3.20). Differentiating this expression, we obtain the PDF of Y:

$$f_Y(y) = \begin{cases} 0 & \text{if } y \le 3, \\ 6/y^2 & \text{if } 3 \le y \le 6, \\ 0 & \text{if } 6 \le y. \end{cases}$$

Example 3.23. Let $Y = g(X) = X^2$, where X is a random variable with known PDF. For any $y \ge 0$, we have

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \le y) \\ &= \mathbf{P}(X^2 \le y) \\ &= \mathbf{P}(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \end{aligned}$$

and therefore, by differentiating and using the chain rule,

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}), \qquad y \ge 0.$$

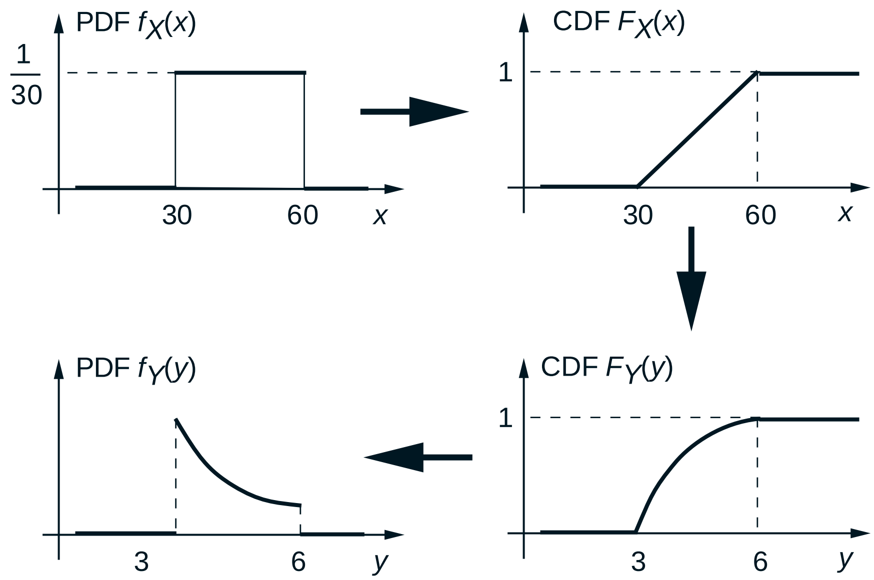

Figure 3.20: The calculation of the PDF of $Y = 180/X$ in Example 3.22. The arrows indicate the flow of the calculation.

The Linear Case

An important case arises when Y is a linear function of X. See Fig. 3.21 for a graphical interpretation.

> **The PDF of a Linear Function of a Random Variable**
>
> Let X be a continuous random variable with PDF f_X, and let
>
> $$Y = aX + b,$$
>
> for some scalars $a \neq 0$ and b. Then,
>
> $$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

To verify this formula, we use the two-step procedure. We only show the

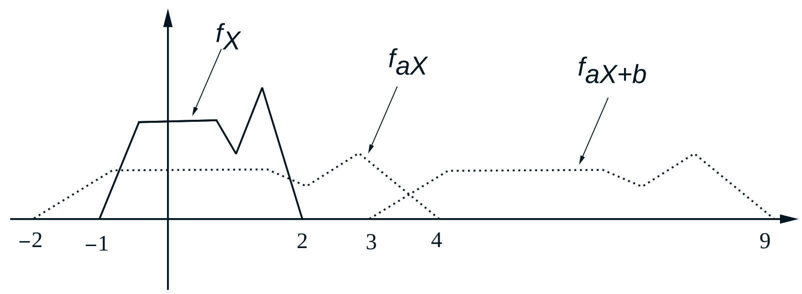

Figure 3.21: The PDF of $aX + b$ in terms of the PDF of X. In this figure, $a = 2$ and $b = 5$. As a first step, we obtain the PDF of aX. The range of Y is wider than the range of X, by a factor of a. Thus, the PDF f_X must be stretched (scaled horizontally) by this factor. But in order to keep the total area under the PDF equal to 1, we need to scale the PDF (vertically) by the same factor a. The random variable $aX + b$ is the same as aX except that its values are shifted by b. Accordingly, we take the PDF of aX and shift it (horizontally) by b. The end result of these operations is the PDF of $Y = aX + b$ and is given mathematically by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

If a were negative, the procedure would be the same except that the PDF of X would first need to be reflected around the vertical axis ("flipped") yielding f_{-X}. Then a horizontal and vertical scaling (by a factor of $|a|$ and $1/|a|$, respectively) yields the PDF of $-|a|X = aX$. Finally, a horizontal shift of b would again yield the PDF of $aX + b$.

steps for the case where $a > 0$; the case $a < 0$ is similar. We have

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \leq y) \\ &= \mathbf{P}(aX + b \leq y) \\ &= \mathbf{P}\left(X \leq \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

We now differentiate this equality and use the chain rule, to obtain

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{1}{a} \cdot \frac{dF_X}{dx}\left(\frac{y-b}{a}\right) = \frac{1}{a} \cdot f_X\left(\frac{y-b}{a}\right).$$

Example 3.24. A linear function of an exponential random variable. Suppose that X is an exponential random variable with PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

where λ is a positive parameter. Let $Y = aX + b$. Then,

$$f_Y(y) = \begin{cases} \dfrac{\lambda}{|a|} e^{-\lambda(y-b)/a} & \text{if } (y-b)/a \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that if $b = 0$ and $a > 0$, then Y is an exponential random variable with parameter λ/a. In general, however, Y need not be exponential. For example, if $a < 0$ and $b = 0$, then the range of Y is the negative real axis.

Example 3.25. A linear function of a normal random variable is normal. Suppose that X is a normal random variable with mean μ and variance σ^2, and let $Y = aX + b$, where a and b are some scalars. We have

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

Therefore,

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-((y-b)/a)-\mu)^2/2\sigma^2} \\ &= \frac{1}{\sqrt{2\pi}\,|a|\sigma} e^{-(y-b-a\mu)^2/2a^2\sigma^2}. \end{aligned}$$

We recognize this as a normal PDF with mean $a\mu + b$ and variance $a^2\sigma^2$. In particular, Y is a normal random variable.

The Monotonic Case

The calculation and the formula for the linear case can be generalized to the case where g is a monotonic function. Let X be a continuous random variable and suppose that its range is contained in a certain interval I, in the sense that $f_X(x) = 0$ for $x \notin I$. We consider the random variable $Y = g(X)$, and assume that g is **strictly monotonic** over the interval I. That is, either

(a) $g(x) < g(x')$ for all $x, x' \in I$ satisfying $x < x'$ (monotonically increasing case), or

(b) $g(x) > g(x')$ for all $x, x' \in I$ satisfying $x < x'$ (monotonically decreasing case).

Furthermore, we assume that the function g is differentiable. Its derivative will necessarily be nonnegative in the increasing case and nonpositive in the decreasing case.

An important fact is that a monotonic function can be "inverted" in the sense that there is some function h, called the inverse of g, such that for all $x \in I$, we have $y = g(x)$ if and only if $x = h(y)$. For example, the inverse of the function $g(x) = 180/x$ considered in Example 3.22 is $h(y) = 180/y$, because we have $y = 180/x$ if and only if $x = 180/y$. Other such examples of pairs of inverse functions include

$$g(x) = ax + b, \qquad h(y) = \frac{y - b}{a},$$

where a and b are scalars with $a \neq 0$ (see Fig. 3.22), and

$$g(x) = e^{ax}, \qquad h(y) = \frac{\ln y}{a},$$

where a is a nonzero scalar.

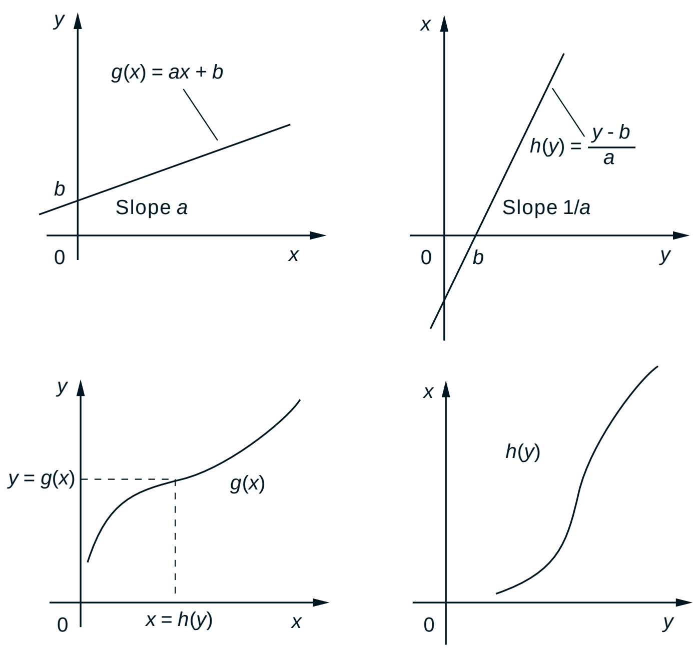

Figure 3.22: A monotonically increasing function g (on the left) and its inverse (on the right). Note that the graph of h has the same shape as the graph of g, except that it is rotated by 90 degrees and then reflected (this is the same as interchanging the x and y axes).

For monotonic functions g, the following is a convenient analytical formula for the PDF of the function $Y = g(X)$.

PDF Formula for a Monotonic Function of a Continuous Random Variable

Suppose that g is monotonic and that for some function h and all x in the range I of X we have

$$y = g(x) \qquad \text{if and only if} \qquad x = h(y).$$

Assume that h has first derivative $(dh/dy)(y)$. Then the PDF of Y in the region where $f_Y(y) > 0$ is given by

$$f_Y(y) = f_X\big(h(y)\big)\left|\frac{dh}{dy}(y)\right|.$$

For a verification of the above formula, assume first that g is monotonically increasing. Then, we have

$$F_Y(y) = \mathbf{P}\big(g(X) \le y\big) = \mathbf{P}\big(X \le h(y)\big) = F_X\big(h(y)\big),$$

where the second equality can be justified using the monotonically increasing property of g (see Fig. 3.23). By differentiating this relation, using also the chain rule, we obtain

$$f_Y(y) = \frac{dF_Y}{dy}(y) = f_X\big(h(y)\big)\frac{dh}{dy}(y).$$

Because g is monotonically increasing, h is also monotonically increasing, so its derivative is positive:

$$\frac{dh}{dy}(y) = \left|\frac{dh}{dy}(y)\right|.$$

This justifies the PDF formula for a monotonically increasing function g. The justification for the case of monotonically decreasing function is similar: we differentiate instead the relation

$$F_Y(y) = \mathbf{P}\big(g(X) \le y\big) = \mathbf{P}\big(X \ge h(y)\big) = 1 - F_X\big(h(y)\big),$$

and use the chain rule.

There is a similar formula involving the derivative of g, rather than the derivative of h. To see this, differentiate the equality $g\big(h(y)\big) = y$, and use the chain rule to obtain

$$\frac{dg}{dh}\big(h(y)\big) \cdot \frac{dh}{dy}(y) = 1.$$

Let us fix some x and y that are related by $g(x) = y$, which is the same as $h(y) = x$. Then,

$$\frac{dg}{dx}(x) \cdot \frac{dh}{dy}(y) = 1,$$

which leads to

$$f_Y(y) = f_X(x) \Big/ \left| \frac{dg}{dx}(x) \right|.$$

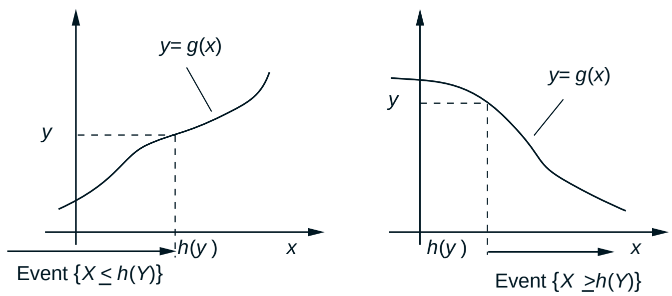

Figure 3.23: Calculating the probability $\mathbf{P}\big(g(X) \leq y\big)$. When g is monotonically increasing (left figure), the event $\{g(X) \leq y\}$ is the same as the event $\{X \leq h(y)\}$. When g is monotonically decreasing (right figure), the event $\{g(X) \leq y\}$ is the same as the event $\{X \geq h(y)\}$.

Example 3.22. (Continued) To check the PDF formula, let us apply it to the problem of Example 3.22. In the region of interest, $x \in [30, 60]$, we have $h(y) = 180/y$, and

$$\frac{dF_X}{dh}\big(h(y)\big) = \frac{1}{30}, \qquad \left| \frac{dh}{dy}(y) \right| = \frac{180}{y^2}.$$

Thus, in the region of interest $y \in [3, 6]$, the PDF formula yields

$$f_Y(y) = f_X\big(h(y)\big) \left| \frac{dh}{dy}(y) \right| = \frac{1}{30} \cdot \frac{180}{y^2} = \frac{6}{y^2},$$

consistently with the expression obtained earlier.

Example 3.26. Let $Y = g(X) = X^2$, where X is a continuous uniform random variable in the interval $(0, 1]$. Within this interval, g is monotonic, and its inverse is $h(y) = \sqrt{y}$. Thus, for any $y \in (0, 1]$, we have

$$\left| \frac{dh}{dy}(y) \right| = \frac{1}{2\sqrt{y}}, \qquad f_X(\sqrt{y}) = 1,$$

and

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} & \text{if } y \in (0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

We finally note that if we interpret PDFs in terms of probabilities of small intervals, the content of our formulas becomes pretty intuitive; see Fig. 3.24.

Functions of Two Random Variables

The two-step procedure that first calculates the CDF and then differentiates to obtain the PDF also applies to functions of more than one random variable.

Example 3.27. Two archers shoot at a target. The distance of each shot from the center of the target is uniformly distributed from 0 to 1, independently of the other shot. What is the PDF of the distance of the losing shot from the center?

Let X and Y be the distances from the center of the first and second shots, respectively. Let also Z be the distance of the losing shot:

$$Z = \max\{X, Y\}.$$

We know that X and Y are uniformly distributed over $[0, 1]$, so that for all $z \in [0, 1]$, we have

$$\mathbf{P}(X \le z) = \mathbf{P}(Y \le z) = z.$$

Thus, using the independence of X and Y, we have for all $z \in [0, 1]$,

$$\begin{aligned} F_Z(z) &= \mathbf{P}\big(\max\{X, Y\} \le z\big) \\ &= \mathbf{P}(X \le z, Y \le z) \\ &= \mathbf{P}(X \le z)\mathbf{P}(Y \le z) \\ &= z^2. \end{aligned}$$

Differentiating, we obtain

$$f_Z(z) = \begin{cases} 2z & \text{if } 0 \le z \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Example 3.28. Let X and Y be independent random variables that are uniformly distributed on the interval $[0, 1]$. What is the PDF of the random variable $Z = Y/X$?

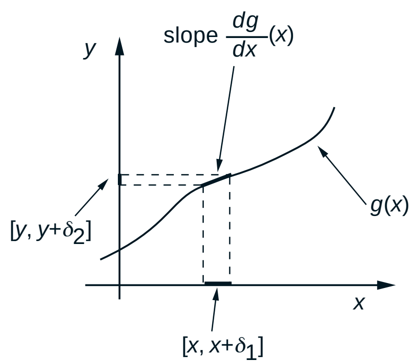

Figure 3.24: Illustration of the PDF formula for a monotonically increasing function g. Consider an interval $[x, x+\delta_1]$, where δ_1 is a small number. Under the mapping g, the image of this interval is another interval $[y, y+\delta_2]$. Since $(dg/dx)(x)$ is the slope of g, we have

$$\frac{\delta_2}{\delta_1} \approx \frac{dg}{dx}(x),$$

or in terms of the inverse function,

$$\frac{\delta_1}{\delta_2} \approx \frac{dh}{dy}(y),$$

We now note that the event $\{x \le X \le x+\delta_1\}$ is the same as the event $\{y \le Y \le y+\delta_2\}$. Thus,

$$\begin{aligned} f_Y(y)\delta_2 &\approx \mathbf{P}(y \le Y \le y+\delta_2) \\ &= \mathbf{P}(x \le X \le x+\delta_1) \\ &\approx f_X(x)\delta_1. \end{aligned}$$

We move δ_1 to the left-hand side and use our earlier formula for the ratio δ_2/δ_1, to obtain

$$f_Y(y)\frac{dg}{dx}(x) = f_X(x).$$

Alternatively, if we move δ_2 to the right-hand side and use the formula for δ_1/δ_2, we obtain

$$f_Y(y) = f_X\big(h(y)\big) \cdot \frac{dh}{dy}(y).$$

We will find the PDF of Z by first finding its CDF and then differentiating. We consider separately the cases $0 \le z \le 1$ and $z > 1$. As shown in Fig. 3.25, we have

$$F_Z(z) = \mathbf{P}\left(\frac{Y}{X} \le z\right) = \begin{cases} z/2 & \text{if } 0 \le z \le 1, \\ 1 - 1/(2z) & \text{if } z > 1, \\ 0 & \text{otherwise.} \end{cases}$$

By differentiating, we obtain

$$f_Z(z) = \begin{cases} 1/2 & \text{if } 0 \le z \le 1, \\ 1/(2z^2) & \text{if } z > 1, \\ 0 & \text{otherwise.} \end{cases}$$

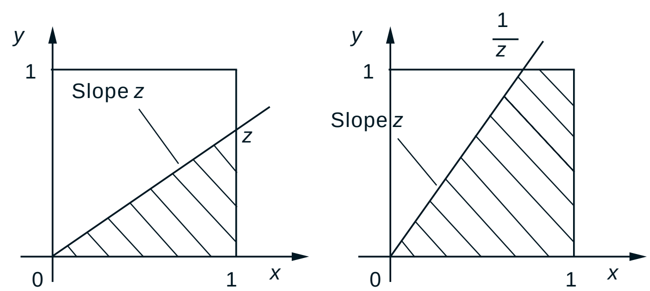

Figure 3.25: The calculation of the CDF of $Z = Y/X$ in Example 3.28. The value $\mathbf{P}(Y/X \le z)$ is equal to the shaded subarea of the unit square. The figure on the left deals with the case where $0 \le z \le 1$ and the figure on the right refers to the case where $z > 1$.

Example 3.29. Romeo and Juliet have a date at a given time, and each, independently, will be late by an amount of time that is exponentially distributed with parameter λ. What is the PDF of the difference between their times of arrival?

Let us denote by X and Y the amounts by which Romeo and Juliet are late, respectively. We want to find the PDF of $Z = X - Y$, assuming that X and Y are independent and exponentially distributed with parameter λ. We will first calculate the CDF $F_Z(z)$ by considering separately the cases $z \ge 0$ and $z < 0$ (see Fig. 3.26).

For $z \ge 0$, we have (see the left side of Fig. 3.26)

$$\begin{aligned} F_Z(z) &= \mathbf{P}(X - Y \le z) \\ &= 1 - \mathbf{P}(X - Y > z) \\ &= 1 - \int_0^\infty \left(\int_{z+y}^\infty f_{X,Y}(x, y)\, dx \right) dy \\ &= 1 - \int_0^\infty \lambda e^{-\lambda y} \left(\int_{z+y}^\infty \lambda e^{-\lambda x}\, dx \right) dy \\ &= 1 - \int_0^\infty \lambda e^{-\lambda y} e^{-\lambda(z+y)}\, dy \\ &= 1 - e^{-\lambda z} \int_0^\infty \lambda e^{-2\lambda y}\, dy \\ &= 1 - \frac{1}{2} e^{-\lambda z}. \end{aligned}$$

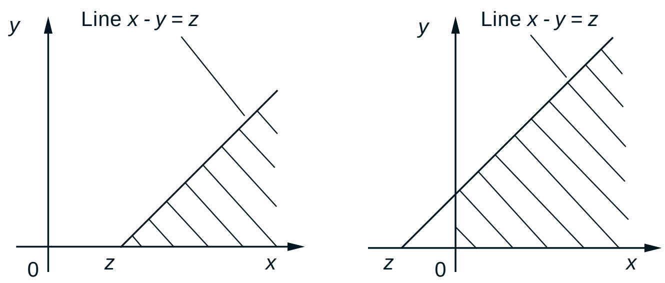

Figure 3.26: The calculation of the CDF of $Z = X - Y$ in Example 3.29. To obtain the value $\mathbf{P}(X - Y > z)$ we must integrate the joint PDF $f_{X,Y}(x, y)$ over the shaded area in the above figures, which correspond to $z \geq 0$ (left side) and $z < 0$ (right side).

For the case $z < 0$, we can use a similar calculation, but we can also argue using symmetry. Indeed, the symmetry of the situation implies that the random variables $Z = X - Y$ and $-Z = Y - X$ have the same distribution. We have

$$F_Z(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(-Z \geq -z) = \mathbf{P}(Z \geq -z) = 1 - F_Z(-z).$$

With $z < 0$, we have $-z \geq 0$ and using the formula derived earlier,

$$F_Z(z) = 1 - F_Z(-z) = 1 - \left(1 - \frac{1}{2}e^{-\lambda(-z)}\right) = \frac{1}{2}e^{\lambda z}.$$

Combining the two cases $z \geq 0$ and $z < 0$, we obtain

$$F_Z(z) = \begin{cases} 1 - \frac{1}{2}e^{-\lambda z} & \text{if } z \geq 0, \\ \frac{1}{2}e^{\lambda z} & \text{if } z < 0, \end{cases}$$

We now calculate the PDF of Z by differentiating its CDF. We obtain

$$f_Z(z) = \begin{cases} \frac{\lambda}{2}e^{-\lambda z} & \text{if } z \geq 0, \\ \\ \frac{\lambda}{2}e^{\lambda z} & \text{if } z < 0, \end{cases}$$

or

$$f_Z(z) = \frac{\lambda}{2}e^{-\lambda|z|}.$$

This is known as a **two-sided exponential PDF**, also known as the **Laplace PDF**.

3.7 SUMMARY AND DISCUSSION

Continuous random variables are characterized by PDFs and arise in many applications. PDFs are used to calculate event probabilities. This is similar to the use of PMFs for the discrete case, except that now we need to integrate instead of adding. Joint PDFs are similar to joint PMFs and are used to determine the probability of events that are defined in terms of multiple random variables. Finally, conditional PDFs are similar to conditional PMFs and are used to calculate conditional probabilities, given the value of the conditioning random variable.

We have also introduced a few important continuous probability laws and derived their mean and variance. A summary is provided in the table that follows.

Summary of Results for Special Random Variables

Continuous Uniform Over $[a, b]$**:**

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbf{E}[X] = \frac{a+b}{2}, \qquad \text{var}(X) = \frac{(b-a)^2}{12}.$$

Exponential with Parameter λ**:**

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \ge 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbf{E}[X] = \frac{1}{\lambda}, \qquad \text{var}(X) = \frac{1}{\lambda^2}.$$

Normal with Parameters μ **and** σ^2**:**

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},$$

$$\mathbf{E}[X] = \mu, \qquad \text{var}(X) = \sigma^2.$$