Chapter 2

Discrete Probabilistic Models

When we talk about a random variable, it is helpful to think of an associated *random experiment* or *trial*. A random experiment or trial can be thought of as any activity that will result in one and only one of several well-defined outcomes, but one does not know in advance which one will occur. The set of all possible outcomes of a random experiment E, denoted by $S(E)$, is called the *sample space* of the random experiment E.

Suppose that the structural condition of a concrete structure (e.g., a bridge) can be classified into one of three categories: poor, fair, or good. An engineer examines one such structure to assess its condition. This is a random experiment and its sample space, $S(E) = \{poor, fair, good\}$, has three elements.

Definition 2.1 (Random variable) *A random variable can be defined as a real-valued function defined over a sample space of a random experiment. That is, the function assigns a real value to every element in the sample space of a random experiment. The set of all possible values of a random variable X, denoted by $S(X)$, is called the support or range of the random variable X.*

Example 2.1 (Concrete structure). In the previous concrete example, let X be $-1, 0$, or 1, depending on whether the structure is poor, fair, or good, respectively. Then X is a random variable with support $S(X) = \{-1, 0, 1\}$. The condition of the structure can also be assessed using a continuous scale, say, from 0 to 10, to measure the concrete quality, with 0 indicating the worst possible condition and 10 indicating the best. Let Y be the assessed condition of the structure. Then Y is a random variable with support $S(Y) = \{y : 0 \leq y \leq 10\}$. ■

We consistently use the customary notation of denoting random variables by uppercase letters such as X, Y, and Z or $X_1, X_2, \ldots, X_n$, where n is the number of random variables under consideration. Realizations of random variables (that is, the actual values they may take) are denoted by the corresponding lowercase letters such as x, y, and z or $x_1, x_2, \ldots, x_n$.

A random variable is said to be *discrete* if it can assume only a finite or countably infinite number of distinct values. Otherwise, it is said to be *continuous*. Thus, a continuous random variable can take an uncountable set of real values. The random variable X in Example 2.1 is discrete, whereas the random variable Y is continuous.

When we deal with a single random quantity, we have a *univariate* random variable. When we deal with two or more random quantities simultaneously, we have a *multivariate* random variable. Section 2.1 presents some probability functions of random variables that are common to all discrete random variables. Section 2.2 introduces examples of the commonly used discrete univariate random variables. Section 2.3 introduces discrete multivariate random variables. Commonly used continuous random variables are reviewed in Chapter 3.

2.1 Univariate Discrete Random Variables

To specify a random variable we need to know (a) its range or support, $S(X)$, which is the set of all possible values of the random variable, and (b) a tool by which we can obtain the probability associated with every subset in its support, $S(X)$. These tools are some functions such as the *probability mass function* (pmf), the *cumulative distribution function* (cdf), or the characteristic function. The pmf, cdf, and the so-called *moments* of random variables are described in this section.

2.1.1 Probability Mass Function

Every discrete random variable has a *probability mass function* (pmf). The pmf of a discrete random variable X is a function that assigns to each real value x the probability of X having the value x. That is, $P_X(x) = \Pr(X = x)$. For notational simplicity we sometimes use $P(x)$ instead of $P_X(x)$. Every pmf $P(x)$ must satisfy the following conditions:

$$P(x) > 0 \quad \text{for all } x \in S(X), \qquad \text{and} \qquad \sum_{x \in S(X)} P(x) = 1. \tag{2.1}$$

Example 2.2 (Concrete structure). Suppose in Example 2.1 that 20% of all concrete structures we are interested in are in poor condition, 30% are in fair condition, and the remaining 50% are in good condition, then if one such structure is selected at random, the probability that the selected structure is in poor condition is $P(-1) = 0.2$, the probability that it is in fair condition is $P(0) = 0.3$, and the probability that it is in good condition is $P(1) = 0.5$. ■

The pmf of a random variable X can be displayed in a table known as a *probability distribution table*. For example, Table 2.1 is the probability distribution table for the random variable X in Example 2.2. The first column in a

Table 2.1: The Probability Mass Function (pmf) of a Random Variable X.

x	$P(x)$
-1	0.2
0	0.3
1	0.5
Total	1.0

Table 2.2: The Probability Mass Function (pmf) and the Cumulative Distribution Function (cdf) of a Random Variable X.

x	$P(x)$	$F(x)$
-1	0.2	0.2
0	0.3	0.5
1	0.5	1.0

probability distribution table is a list of the values of $x \in S(X)$, that is, only the values of x for which $P(x) > 0$. The second column displays $P(x)$. It is understood that $P(x) = 0$ for every $x \notin S(X)$.

2.1.2 Cumulative Distribution Function

Every random variable also has a *cumulative distribution function* (cdf). The cdf of a random variable X, denoted by $F(x)$, is a function that assigns to each real value x the probability of X having values less than or equal to x, that is,

$$F(x) = \Pr(X \leq x) = \sum_{a \leq x} P(a).$$

Accordingly, the cdf can be obtained from the pmf and vice versa. For example, the cdf in the last column of Table 2.2 is computed from the pmf in Table 2.1 by accumulating $P(x)$ in the second column.

The pmf and cdf of any discrete random variable X can be displayed in probability distribution tables, such as Table 2.2, or they can be displayed graphically. For example, the graphs of the pmf and cdf in Table 2.2 are shown in Figure 2.1. In the graph of pmf, the height of a line on top of x is $P(x)$. The graph of the cdf for discrete random variable is a step function. The height of the step function is $F(x)$.

The cdf has the following properties as a direct consequence of the definitions of cdf and probability (see, for example, Fig. 2.1):

1. $F(\infty) = 1$ and $F(-\infty) = 0$.

2. $F(x)$ is nondecreasing and right continuous.

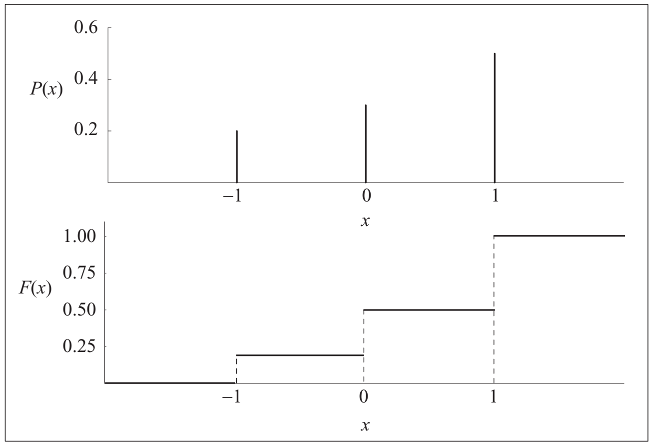

Figure 2.1: Graphs of the pmf and cdf of the random variable in Table 2.2.

3. $P(x)$ is the jump of the cdf at x.
4. $\Pr(a < X \leq b) = F(b) - F(a)$.

2.1.3 Moments

Let $g(X)$ be a function of a discrete random variable X. The *expected value* of $g(X)$ is defined by

$$E[g(X)] = \sum_{x \in S(X)} g(x)P(x). \tag{2.2}$$

For example, letting $g(X) = X^r$, we obtain the so-called rth *moment* of the discrete random variable X, with respect to the origin

$$E(X^r) = \sum_{x \in S(X)} x^r P(x). \tag{2.3}$$

When $r = 1$, we obtain the *mean*, μ, of the discrete random variable X,

$$\mu = E(X) = \sum_{x \in S(X)} xP(x). \tag{2.4}$$

Thus, the mean, μ, is the first moment of X with respect to the origin.

Letting $g(X) = (X - \mu)^r$, we obtain the rth *central moment*,

$$E[(X - \mu)^r] = \sum_{x \in S(X)} (x - \mu)^r P(x). \tag{2.5}$$

When $r = 1$, it can be shown that the *first* central moment of any random variable X is zero, that is,

$$E(X - \mu) = 0. \tag{2.6}$$

Table 2.3: Calculations of the Mean and Variance of the Random Variable X.

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
–1	0.2	–0.2	1	0.2
0	0.3	0.0	0	0.0
1	0.5	0.5	1	0.5
Total	1.0	0.3		0.7

When $r = 2$, we obtain the second central moment, better known as the *variance*, σ^2, of the discrete random variable X, that is,

$$\sigma^2 = E[(X - \mu)^2] = \sum_{x \in S(X)} (x - \mu)^2 P(x). \tag{2.7}$$

The *standard deviation*, σ, of the random variable X is the positive square root of its variance. The mean can be thought of as a measure of *center* and the standard deviation (or, equivalently, the variance) as a measure of *spread* or *variability*. It can be shown that the variance can also be expressed as

$$\sigma^2 = E(X^2) - \mu^2, \tag{2.8}$$

where

$$E(X^2) = \sum_{x \in S(X)} x^2 P(x)$$

is the second moment of X with respect to the origin. For example, the calculations of the mean and variance of the random variable X are shown in Table 2.3. Accordingly, the mean and variance of X are $\mu = 0.3$ and $\sigma^2 = 0.7 - 0.3^2 = 0.61$, respectively.

The expected value, defined in (2.2), can be thought of as an *operator*, which has the following properties:

1. $E(c) = c$, for any constant c, that is, the expected value of a constant (a degenerate random variable) is the constant.

2. $E[cg(X)] = cE[g(X)]$.

3. $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$, for any functions $g(X)$ and $h(X)$.

For example, $E(c + X) = E(c) + E(X) = c + \mu$. In other words, the mean of a constant plus a random variable is the constant plus the mean of the random variable. As another example,

$$\sigma^2 = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2.$$

This is actually the proof of the identity in (2.8).

2.2 Common Discrete Univariate Models

In this section we present several important discrete random variables that often arise in extreme value applications. For a more detailed description and additional random variables, see, for example, the books by Balakrishnan and Nevzorov (2003), Christensen (1984), Galambos (1995), Johnson et al. (1992), Ross (1992), and Wackerly et al. (2002).

2.2.1 Discrete Uniform Distribution

When a random variable X can have one on n possible values and they are all equally likely, then X is said to have a *discrete uniform distribution*. Since the possible values are equally likely, the probability for each one of them is equal to $1/n$. Without loss of generality, let us assume these values are $1, \ldots, n$. Then the pmf of X is

$$P(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n. \tag{2.9}$$

This discrete uniform distribution is denoted by $U(n)$. The mean and variance of $U(n)$ are

$$\mu = (n+1)/2 \quad \text{and} \quad \sigma^2 = (n+1)(4n^2 - n - 3)/12, \tag{2.10}$$

respectively.

Example 2.3 (Failure types). A computer system has four possible types of failure. let $X = i$ if the system results in a failure of type i, with $i = 1, 2, 3, 4$. If these failure types are equally likely to occur, then the distribution of X is $U(4)$ and the pmf is

$$P(x) = \frac{1}{4}, \quad x = 1, 2, 3, 4.$$

The mean and variance can be shown to be 2.5 and 23.75, respectively. ∎

2.2.2 Bernoulli Distribution

The *Bernoulli* random variable arises in situations where we have a random experiment, which has two possible mutually exclusive outcomes: *success* or *failure*. The probability of success is p and the probability of failure is $1 - p$. This random experiment is called a *Bernoulli trial* or *experiment*. Define a random variable X by

$$X = \begin{cases} 0, & \text{if a failure is observed,} \\ 1, & \text{if a success is observed.} \end{cases}$$

This is called a *Bernoulli random variable* and its distribution is called a *Bernoulli distribution*. The pmf of X is

$$P(x) = p^x (1-p)^{1-x}, \quad x = 0, 1, \tag{2.11}$$

and its cdf is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - p, & \text{if } 0 \le x < 1, \\ 1, & \text{if } x \ge 1. \end{cases} \tag{2.12}$$

The mean and variance of a Bernoulli random variable are

$$\mu = p \quad \text{and} \quad \sigma^2 = p(1-p), \tag{2.13}$$

respectively. Note that if $p = 1$, then X becomes a degenerate random variable (that is, a constant) and the pmf of X is

$$P(x) = \begin{cases} 1, & \text{if } x = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.14}$$

This is known as the *Dirac* function.

Example 2.4 (Concrete structure). Suppose we are interested in knowing whether or not a given concrete structure is in poor condition. Then, a random variable X can be defined as

$$X = \begin{cases} 1, & \text{if the condition is poor,} \\ 0, & \text{otherwise.} \end{cases}$$

This is a Bernoulli random variable. From Example 2.2, 20% of structures are in poor condition. Then the pmf is

$$P(x) = 0.2^x (1 - 0.2)^{1-x}, \quad x = 0, 1.$$

The mean and variance of X are $\mu = p = 0.2$ and $\sigma^2 = p(1-p) = 0.16$. ■

Bernoulli random variables arise frequently while handling extremes. Engineers are often interested in events that cause failure such as exceedances of a random variable over a threshold value.

Definition 2.2 (Exceedances) *Let X be a random variable and u a given threshold value. The event $\{X = x\}$ is said to be an exceedance at the level u if $x > u$.*

For example, waves can destroy a breakwater when their heights exceed a given value, say 9 m. Then it does not matter whether the height of a wave is 9.5, 10, or 12 m because the consequences of these events are the same.

Let X be a random variable representing heights of waves and Y_u be defined as

$$Y_u = \begin{cases} 0, & \text{if no exceedance occurred,} \\ 1, & \text{if an exceedance occurred.} \end{cases}$$

Then Y_u is a Bernoulli random variable with success probability $p_u = \Pr(X > u)$.

Bernoulli random variables arise in many important practical engineering situations.

Example 2.5 (Yearly maximum wave height). When designing a breakwater, civil engineers need to define the so-called design wave height, which is a wave height such that, when occurring, the breakwater will be able to withstand it without failure. Then, a natural design wave height would be the maximum wave height reaching the breakwater during its lifetime. However, this value is random and cannot be found. So, the only thing that an engineer can do is to choose this value with a small probability of being exceeded. In order to obtain this probability it is important to know the probability of exceedances of certain values during a year. Then, if we are concerned with whether the yearly maximum wave height exceeds a given threshold value h_0, we have a Bernoulli experiment. ∎

Example 2.6 (Tensile strength). Suspension bridges are supported by long cables. However, long cables are much weaker than short cables, the only ones tested in the laboratory. This is so because of the weakest link principle, which states that the strength of a long piece is the minimum strength of all its constituent pieces. Thus, the engineer has to extrapolate from lab results to real cables. The design of a suspension bridge requires the knowledge of the probability of the strength of the cable to fall below certain values. That is why values below a threshold are important. ∎

Example 2.7 (Nuclear power plant). When designing a nuclear power plant one has to consider the occurrence of earthquakes that can lead to disastrous consequences. Apart from the earthquake intensity, one of the main parameters to be considered is the distance from the earthquake epicenter to the location of the plant. Damage will be more severe for short distances than for long ones. Thus, engineers need to know whether this distance is below a given threshold value. ∎

Example 2.8 (Temperature). Temperatures have a great influence on engineering works and can cause problems either for large or small values. Then, values above or below given threshold values are important. ∎

Example 2.9 (Water flows). The water circulating through rivers also influences the life of humans. If the amount of water exceeds a given level, large areas can be flooded. On the other hand, if the water levels are below given values, the environment can be seriously damaged. ∎

2.2.3 Binomial Distribution

Suppose now that n Bernoulli experiments are run such that the following conditions hold:

1. The experiments are identical, that is, the probability of success p is the same for all trials.

2. The experiments are independent, that is, the outcome of an experiment has no influence on the outcomes of the others.

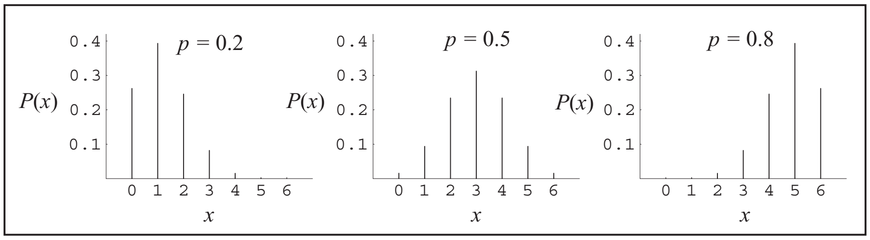

Figure 2.2: Examples of probability mass functions of binomial random variables with $n = 6$ and three values of p.

Let X be the number of successes in these n experiments. Then X is a random variable. To obtain the pmf of X, we first consider the event of obtaining x successes. If we obtained x successes, it also means that we obtained $n - x$ failures. Because the experiments are identical and independent, the probability of obtaining x successes and $n - x$ failures is

$$p^x(1-p)^{n-x}.$$

Note also that the number of possible ways of obtaining x successes (and $n - x$ failures) is obtained using the combinations formula:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Therefore, the pmf of X is

$$P(x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x = 0, 1, \ldots, n. \tag{2.15}$$

This random variable is known as the *binomial random variable* and is denoted by $X \sim B(n, p)$ and the distribution in (2.15) is called the *binomial distribution*.

The mean and variance of a $B(n, p)$ random variable can be shown to be

$$\mu = np \qquad \text{and} \qquad \sigma^2 = np(1-p). \tag{2.16}$$

Figure 2.2 shows the graphs of the pmf of three binomial random variables with $n = 6$ and three values of p. From these graphs it can be seen that when $p = 0.5$, the pmf is symmetric; otherwise it is skewed.

Since X is the number of successes in these n identical and independent Bernoulli experiments, one may think of X as the sum of n identical and independent Bernoulli random variables, that is, $X = X_1 + X_2 + \ldots + X_n$, where X_i is a Bernoulli random variable with probability of success equal to p. Note that when $n = 1$, then a $B(1, p)$ random variable is a Bernoulli random variable.

Another important property of binomial random variables is *reproductivity* with respect to the parameter n. This means that the sum of two independent binomial random variables with the same p is also a binomial random variable. More precisely, if $X_1 \sim B(n_1, p)$ and $X_2 \sim B(n_2, p)$, then

$$X_1 + X_2 \sim B(n_1 + n_2, p).$$

Example 2.10 (Exceedances). An interesting practical problem consists of determining the probability of r exceedances over a value u in n identical and independent repetitions of the experiment. Since there are only two possible outcomes (exceedance or not exceedance), these are Bernoulli experiments. Consequently, the number of exceedances M_u over the value u of the associated random variable X is a $B(n, p_u)$ random variable with parameters n and p_u, where p_u is the probability of an exceedance over the level u of X. Therefore, the pmf of M_u is

$$\Pr(M_u = r) = \binom{n}{r} (p_u)^r (1 - p_u)^{n-r}, \quad r = 0, 1, \ldots, n. \tag{2.17}$$

Moreover, since p_u can be written as

$$p_u = \Pr(X > u) = 1 - F(u), \tag{2.18}$$

where $F(\cdot)$ is the cdf of X, (2.17) becomes

$$\Pr(M_u = r) = \binom{n}{r} [1 - F(u)]^r [F(u)]^{n-r}. \tag{2.19}$$

■

Example 2.11 (Concrete structures). Suppose that an engineer examined $n = 6$ concrete structures to determine which ones are in poor condition. As in Example 2.4, the probability that a given structure is in poor condition is $p = 0.2$. If X is the number of structures that are in poor condition, then X is a binomial random variable $B(6, 0.2)$. From (2.15) the pmf is

$$P(x) = \binom{6}{x} p^x (1 - p)^{6-x}, \quad x = 0, 1, \ldots, 6. \tag{2.20}$$

The graph of this pmf is given in Figure 2.2. For example, the probability that none of the six structures are found to be in poor condition is

$$P(0) = \binom{6}{0} 0.2^0 \, 0.8^6 = 0.2621,$$

and the probability that only one of the six structures are found to be in poor condition is

$$P(1) = \binom{6}{1} 0.2^1 \, 0.8^5 = 0.3932.$$

■

Example 2.12 (Yearly maximum wave height). Consider a breakwater that is to be designed for a lifetime of 50 years. Assume also that the probability of yearly exceedance of a wave height of 9 m is 0.05. Then, the probability of having 5 years with exceedances during its lifetime is given by

$$P(5) = \binom{50}{5} 0.05^5 (1 - 0.05)^{50-5} = 0.06584.$$

Note that we have admitted the two basic assumptions of the binomial model, that is, identical and independent Bernoulli experiments. In this case both assumptions are reasonable. Note, however, that if the considered period were one day or one month instead of one year this would not be the case, because the wave heights in consecutive days are not independent events. (Assume both days belong to the same storm, then the maximum wave heights would be both high. On the other hand, if the periods were calm, both would be low.)

It is well known that there are some periodical phenomena ruling the waves that can last for more than one year. For that reason it would be even better to consider periods of longer duration. ■

Example 2.13 (Earthquake epicenter). From past experience, the epicenters of 10% of the earthquakes are within 50 km from a nuclear power plant. Now, consider a sequence of 10 such earthquakes and let X be the number of earthquakes whose epicenters are within 50 km from the nuclear power plant. Assume for the moment that the distances associated with different earthquakes are independent random variables and that all the earthquakes have the same probabilities of having their epicenters at distances below 50 km. Then, X is a $B(10, 0.1)$ random variable. Accordingly, the probability that none of the 10 earthquakes will occur within 50 km is

$$P(0) = \binom{10}{0} 0.1^0 (1 - 0.1)^{10-0} = 0.348678.$$

Note that this probability is based on two assumptions:

1. The distances associated with any two earthquakes are independent random variables.
2. The occurrence of an earthquake does not change the possible locations for others.

Both assumptions are not very realistic, because once an earthquake has occurred some others usually occur in the same or nearby location until the accumulated energy is released. After then, the probability of occurrence at the same location becomes much smaller, because no energy has been built up yet. ■

2.2.4 Geometric or Pascal Distribution

Consider again a series of identical and independent Bernoulli experiments, which are repeated until the first success is obtained. Let X be the number

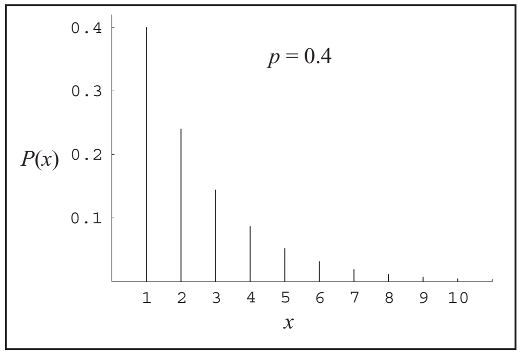

Figure 2.3: Graph of the probability mass function of the geometric random variable with $p = 0.4$.

of trial on which the first success occurs. What is the pmf of the random variable X? Note that if the first success has occurred at the trial number x, then the first $(x-1)$ trials must have been failures. Since the probability of a success is p and the probability of the $(x-1)$ failures is $(1-p)^{x-1}$ (because the trials are identical and independent), the pmf of X is

$$P(x) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots. \tag{2.21}$$

This random variable is called a *geometric* or *Pascal* random variable and is denoted by $G(p)$. The pmf of $G(p)$ random variable is decreasing in x, which means that the largest value of $P(x)$ is at $x = 1$. A graph of the pmf of $G(0.4)$ is shown in Figure 2.3. The mean and variance of $G(p)$ are

$$\mu = \frac{1}{p} \quad \text{and} \quad \sigma^2 = \frac{1-p}{p^2}. \tag{2.22}$$

Example 2.14 (Job interviews). A company has one vacant position to fill. It is known that 80% of job applicants for this position are actually qualified for the job. The company interviews the applicants one at a time as they come in. The interview process stops as soon as one qualified applicant is found. How many interviews will have to be conducted until the first qualified applicant is found? This can be thought of having a series of identical and independent Bernoulli trials each with success probability $p = 0.8$. If X is the number of interviews required to find the first qualified applicant, then X is $G(0.8)$ random variable. For example, the probability that the first qualified applicant is found on the third interview is

$$P(3) = 0.8(1-0.8)^{3-1} = 0.032.$$

Also, the probability that the company will conduct at least three interviews to find the first qualified applicant is

$$\Pr(X \geq 3) = 1 - \Pr(X \leq 2) = 1 - F(2) = 1 - [P(1) + P(2)] = 0.04.$$

If each interview costs the company \$500, then the expected costs of filling the vacant position is

$$500\ E(X) = 500\ (1/0.8) = \$625.$$

■

Assume now that a given event (flood, dam failure, exceedance over a given temperature, etc.) is such that its probability of occurrence during a period of unit duration (normally one year) is a small value p. Assume also that the occurrences of such event in nonoverlapping periods are independent. Then, as time passes, we have a sequence of identical Bernoulli experiments (occurrence or not of the given event). Thus, the time measured in the above units until the first occurrence of this event is the number of experiments until the first occurrence and then it can be considered as a geometric $G(p)$ random variable, whose mean is $1/p$. This suggests the following definition.

Definition 2.3 (Return period) *Let A be an event, and T be the random time between successive occurrences of A. The mean value, μ, of the random variable T is called the return period of A (note that it is the mean time for the return of such an event).*

For the return period to be approximately $1/p$, the following conditions must hold:

1. The probability of one event occurring during a short period of time is small.

2. The probability of more than one event occurring during a short period of time is negligible.

2.2.5 Negative Binomial Distribution

The geometric distribution arises when we are interested in the number of Bernoulli trials that are required until we get the first success. Now suppose that we define the random variable X as the number of identical and independent Bernoulli trials that are required until we get the rth success. For the rth success to occur at the trial number x, we must have $r-1$ successes in the $x-1$ previous trials and one success in the trial number x. The number of possible ways of obtaining $r-1$ successes in $x-1$ trials is obtained using the combinations formula:

$$\binom{x-1}{r-1} = \frac{(x-1)!}{(r-1)!\,(x-r)!}.$$

Therefore, the pmf of X is

$$P(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \ldots. \tag{2.23}$$

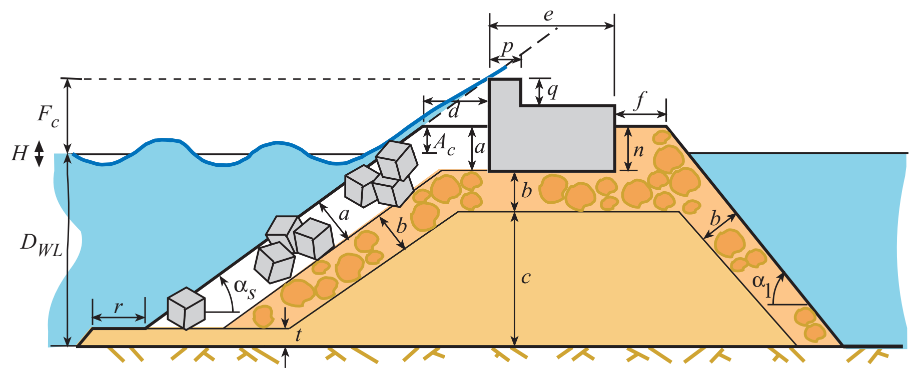

Figure 2.4: Illustration of the parts of a breakwater.

This random variable is called a *negative binomial* random variable and is denoted by $X \sim NB(r, p)$. Note that the geometric distribution is a special case of the negative binomial distribution obtained by setting $(r = 1)$, that is, $G(p) \sim NB(1, p)$. The mean and variance of an $NB(r, p)$ variable are

$$\mu = \frac{r\ (1-p)}{p} \quad \text{and} \quad \sigma^2 = \frac{r\ (1-p)}{p^2}. \tag{2.24}$$

Example 2.15 (Job interviews). Suppose that the company in Example 2.14 wishes to fill two vacant positions. Thus, the interview process stops as soon as two qualified applicants are found. If X is the number of interviews needed to fill the two vacant positions, then X is an $NB(2, 0.8)$ random variable. For example, the probability that the second qualified applicant is found on the third interview is

$$P(3) = \binom{3-1}{2-1}\, 0.8^2\; 0.2^{3-2} = 0.128.$$

■

Example 2.16 (Rubble-mound breakwater). A rubble-mound breakwater is made of a supported crownwall on an earthfill that is protected by a mound armor (large pieces of stone to protect the earthfill from the waves) (see Fig. 2.4). The geometrical connectivity and structural stress transmission in the armor occurs by friction and interlocking between units. While failure of rigid breakwaters occurs when a single wave exceeds a given threshold value, a rubble-mound breakwater fails after the occurrence of several waves above a given threshold value. This is because the failure is progressive, that is, the first wave produces some movement of the stone pieces, the second increases the damage, etc.

Then, a failure occurs when the rth Bernoulli event (wave height exceeding the threshold) occurs. Thus, the negative binomial random variable plays a key role. Assume that a wave produces some damage on the armor if its height exceeds 7 m, and that the probability of this event is 0.001. Then, if the rubble-mound breakwater fails after, say, eight such waves, then the number of waves occurring until failure is a negative binomial random variable. Consequently, the pmf of the number of waves until failure is

$$P(x) = \binom{x-1}{8-1} 0.001^8 (1-0.001)^{x-8}, \quad x = 8, 9, \ldots.$$

■

Like the binomial random variable, the negative binomial random variable is also reproductive with respect to parameter r. This means that the sum of independent negative binomial random variables with the same probability of success p is a negative binomial random variable. More precisely, if $X_1 \sim NB(r_1, p)$ and $X_2 \sim NB(r_2, p)$, then

$$X_1 + X_2 \sim NB(r_1 + r_2, p).$$

2.2.6 Hypergeometric Distribution

Suppose we have a finite population consisting of N elements, where each element can be classified into one of two distinct groups. Say, for example, that we have N products of which D products are defective and the remaining $N - D$ are acceptable (nondefective). Suppose further that we wish to draw a random sample of size $n < N$ from this population without replacement. The random variable X, which is the number of defective items in the sample, is called a *hypergeometric* random variable and is denoted $HG(N, p, n)$, where $p = D/N$ is the proportion of defective items in the population.

It is clear that the number of defective elements, X, cannot exceed either the total number of defective elements D, or the sample size n. Also, it cannot be less than 0 or less than $n - (N - D)$. Therefore, the support of X is

$$\max(0, n - qN) \le X \le \min(n, D),$$

where $q = 1 - p$ is the proportion of acceptable items in the population. The probability mass function of X is

$$P(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \quad \max(0, n - qN) \le x \le \min(n, Np). \tag{2.25}$$

The numerator is the number of samples that can be obtained with x defective elements and $n-x$ nondefective elements. The denominator is the total number of possible samples of size n that can be drawn. The mean and variance of $HG(N, p, n)$ are

$$\mu = np \quad \text{and} \quad \sigma^2 = \frac{N-n}{N-1}\, n\, p\, (1-p). \tag{2.26}$$

When N tends to infinity this distribution tends to the binomial distribution.

Example 2.17 (Urn problem). An urn contains 20 balls, 5 white and 15 black. We draw a sample of size 10 without replacement. What is the probability that the drawn sample contains exactly 2 white balls? Here $N = 20$, $p = 5/20$, and $n = 10$. Letting X be the number of white balls, then X is $HG(20, 1/4, 10)$. From (2.25), we have

$$P(2) = \frac{\binom{5}{2}\binom{15}{8}}{\binom{20}{10}} = 0.348.$$

∎

2.2.7 Poisson Distribution

Suppose we are interested in the number of occurrences of an event over a given interval of time or space. For example, let X be the number of traffic accidents occurring during a time interval t, or the number of vehicles arriving at a given intersection during a time interval of duration t. Then, X is a random variable and we are interested in finding its pmf. The experiment here consists of counting the number of times an event occurs during a given interval of duration t. Note that t does not have to be time; it could be location, area, volume, etc.

To derive the pmf of X we make the following *Poissonian* assumptions:

1. The probability p of the occurrence of a single event in a short interval d is proportional to its duration, that is, $p = \alpha d$, where α is a positive constant, known as the *arrival* or *intensity rate*.

2. The probability of the occurrence of more than one event in the same interval is negligible.

3. The number of occurrences in one interval is independent of the number of occurrences in other nonoverlapping intervals.

4. The number of events occurring in two intervals of the same duration have the same probability distribution.

Now, divide the interval t into n small and equal subintervals of duration $d = t/n$. Then, with the above assumptions, we may think of the n subintervals as n identical and independent Bernoulli trials $X_1, X_2, \ldots, X_n$, with $\Pr(X_i = 1) = p$ for $i = 1, 2, \ldots, n$. Letting $X = \sum_{i=1}^{n} X_i$, then X is a binomial random variable with parameters n and $p = \alpha t/n$. To guarantee that no more than a single event occurs in a subinterval d, the interval t may have to be divided into a very large number of intervals. So, we are interested in the pmf of X as $n \to \infty$. Under the above assumptions, one can show that

$$P(x) = \lim_{n\to\infty} \binom{n}{x} p^x \,(1-p)^{n-x} = \frac{e^{-\alpha t}(\alpha t)^x}{x!}, \quad x = 0, 1, \ldots, \lambda > 0. \quad (2.27)$$

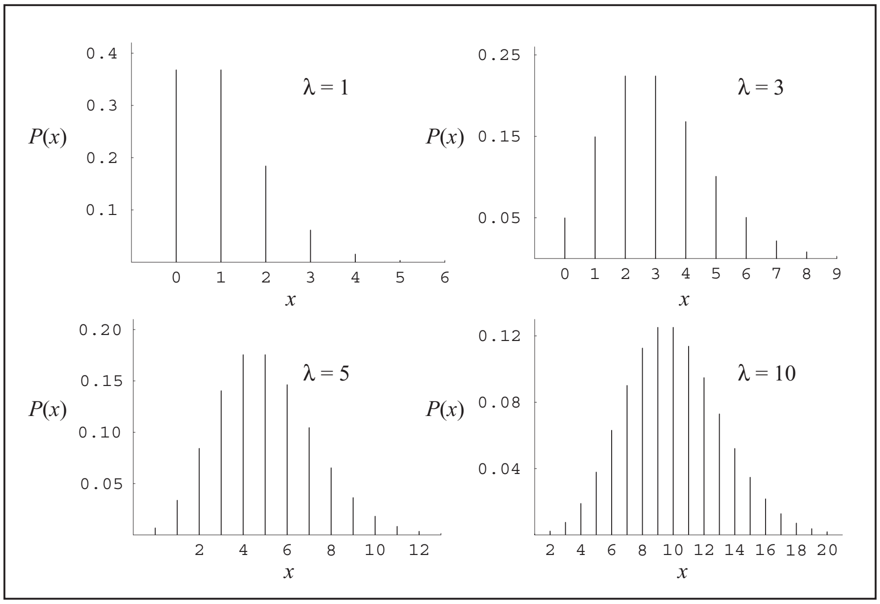

Figure 2.5: Some examples of probability mass functions of the Poisson random variable with four different values of λ.

Letting $\lambda = \alpha t$, we obtain

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots, \lambda > 0. \tag{2.28}$$

This random variable, which is the number of events occurring in period of a given duration t, is known as a Poisson random variable with parameter $\lambda = \alpha t$ and is denoted by $P(\lambda)$. Note that the parameter λ is equal to the intensity α times the duration t.

Figure 2.5 shows the graphs of the pmf of some Poisson random variables. It can be seen that as the parameter λ gets larger, the pmf becomes more symmetric.

The mean and variance of a $P(\lambda)$ variable are

$$\mu = \lambda \quad \text{and} \quad \sigma^2 = \lambda. \tag{2.29}$$

Like the binomial random variable, the Poisson random variables are also reproductive, that is, if $X_1 \sim P(\lambda_1)$ and $X_2 \sim P(\lambda_2)$ are independent, then

$$X_1 + X_2 \sim P(\lambda_1 + \lambda_2).$$

The Poisson random variable is particularly appropriate for modeling the number of occurrences of rare events such as storms, earthquakes, and floods.

Example 2.18 (Storms). Suppose that storms of a certain level occur once every 50 years on the average. We wish to compute the probability of no such storm will occur during a single year. Assuming that X has a Poisson random variable with parameter λ, then $\lambda = 1/50 = 0.02$ and, using (2.28), we have

$$P(0) = \frac{e^{-1/50}(1/50)^0}{0!} = e^{-0.02} = 0.98.$$

That is, it is highly likely that no storms will occur in a single year.

For this model to be correct we need to check the assumptions of the above Poisson model. The first two assumptions are reasonable, because if several storms occur during a short interval, they could be considered as a single storm. The third and fourth assumptions are not true for close intervals, but they are true for far enough intervals. ∎

Example 2.19 (Parking garage). A parking garage has three car entrances. Assume that the number of cars coming into the garage using different entrances are independent Poisson random variables with parameters λ_1, λ_2, and λ_3. Then, using the reproductivity property of Poisson random variables, the total number of cars entering the garage is a Poisson random variable $P(\lambda_1+\lambda_2+\lambda_3)$.

The definition of reproductivity assumes that the random variables being considered in the sum are independent. Then, the number of cars entering at each entrance must be independent. This assumption must be checked before the above Poisson model is used. ∎

Poisson Approximation of the Binomial Distribution

If X is a $B(n,p)$ random variable, but p is small, say $p \le 0.01$, and $np \le 5$, then the pmf of X can be approximated by the pmf of the Poisson random variable with $\lambda = np$, the mean of the binomial random variable, that is,

$$P(x) \approx \frac{e^{-\lambda}\,\lambda^x}{x!}, \quad x = 0, 1, \ldots, n. \tag{2.30}$$

This is why the Poisson process is known as the *rare* events process.

Example 2.20 (Storms). Consider the storms in Example 2.18, and suppose that we are interested in the number of years with storms over a 40-year period. Although X is $B(40, 1/50)$, it can be approximated by a $P(40/50)$ random variable. For example, $P(3)$ can be computed either exactly, using the binomial pmf

$$P(x) = \binom{40}{3}\left(\frac{1}{50}\right)^3\left(1-\frac{1}{50}\right)^{40-3} = 0.0374293,$$

or approximately, using the Poisson pmf,

$$P(x) \approx \frac{(40/50)^3 e^{-40/50}}{3!} = 0.0383427.$$

The error of approximation in this case is $0.0374293 - 0.0383427 = -0.0009134$. ∎

Table 2.4: Some Discrete Random Variables that Arise in Engineering Applications, Together with Their Probability Mass Functions, Parameters, and Supports.

Distribution	$P(x)$	Parameters and Support
Bernoulli	$P(x) = p^x(1-p)^{1-x}$	$0 < p < 1$ $x = 0, 1$
Binomial	$\binom{n}{x} p^x(1-p)^{n-x}$	$n = 1, 2, \ldots$ $0 < p < 1$ $x = 0, 1, \ldots, n$
Geometric	$p(1-p)^{x-1}$	$0 < p < 1$ $x = 1, 2, \ldots$
Negative Binomial	$\binom{x-1}{r-1} p^r(1-p)^{x-r}$	$0 < p < 1$ $x = r, r+1, \ldots$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	$\lambda > 0$ $x = 0, 1, \ldots$
Nonzero Poisson	$\frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})}$	$\lambda > 0$ $x = 1, 2, \ldots$

2.2.8 Nonzero Poisson Distribution

In certain practical applications we are interested in the number of occurrences of an event over a period of duration t, but we also know that at least one event has to occur during the period. If the Poissonian assumptions hold, then it can be shown that the random variable X has the following pmf

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})}, \quad x = 1, 2, \ldots, \lambda > 0. \tag{2.31}$$

This distribution is known as the *nonzero* or the *zero-truncated* Poisson distribution and is denoted by $P_0(\lambda)$. The mean and variance of $P_0(\lambda)$ are

$$\mu = \frac{\lambda}{1-e^{-\lambda}} \quad \text{and} \quad \sigma^2 = \frac{e^{\lambda}\lambda(-1+e^{\lambda}-\lambda)}{(1-e^{\lambda})^2}.$$

A summary of the random variables we discussed in this section is given in Table 2.4.

2.3 Multivariate Discrete Random Variables

In Section 2.2 we have dealt with random variables individually, that is, one random quantity at a time. In some practical situations, we may need to deal with several random quantities simultaneously. In this section we describe models that deal with multidimensional random variables. For a detailed discussion on various multivariate discrete models, see the book by Johnson et al. (1997).

Table 2.5: The Joint Probability Mass Function and the Marginal Probability Mass Functions of (X_1, X_2) in Example 2.21.

	X_2			
X_1	1	2	3	$P_1(x_1)$
0	0.1	0.3	0.2	0.6
1	0.2	0.1	0.1	0.4
$P_2(x_2)$	0.3	0.4	0.3	1.0

2.3.1 Joint Probability Mass Function

Let $X = \{X_1, X_2, \ldots, X_n\}$ be an n-dimensional discrete random variable, taking values $x_i \in S(X_i), i = 1, 2, \ldots, n$. The pmf of this multivariate random variable is denoted by $P(x_1, x_2, \ldots, x_n)$, which means $\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$. This is called the *joint* probability mass function. The joint pmf has n arguments, $x_1, x_2, \ldots, x_n$, one for each variable. When $n = 2$, we have a *bivariate* random variable.

Example 2.21 (Bivariate pmf). Suppose that $n = 2$ and the supports of X_1 and X_2 are $S(X_1) = \{0, 1\}$ and $S(X_2) = \{1, 2, 3\}$, respectively. The joint pmf can be displayed in a table such as the one given in Table 2.5. It has two arguments, $x_1 = 0, 1$ and $x_2 = 1, 2, 3$. From Table 2.5 we see, for example, that $P(0, 1) = 0.1$ and $P(0, 3) = 0.2$. ■

2.3.2 Marginal Probability Mass Function

From the joint pmf we can obtain *marginal* probability mass functions, one marginal for each variable. The marginal pmf of X_1, $P_1(x_1)$, is shown in the last column in Table 2.5. It is obtained by adding across the rows. Similarly, the marginal pmf of X_2, $P_2(x_2)$, is shown in the last row in Table 2.5. It is obtained by adding across the columns. More generally, the marginal of the jth variable, X_j, is obtained by summing the joint pmf over all possible values of all other variables. For example, the marginal pmf of X_1 is

$$P_1(x_1) = \sum_{x_2 \in S(X_2)} \cdots \sum_{x_n \in S(X_n)} P(x_1, x_2, \ldots, x_n), \quad x_1 \in S(X_1) \tag{2.32}$$

and the marginal of (X_1, X_2) is

$$P_{12}(x_1, x_2) = \sum_{x_3 \in S(X_3)} \cdots \sum_{x_n \in S(X_n)} P(x_1, x_2, \ldots, x_n), \;\; x_1 \in S(X_1);\; x_2 \in S(X_2). \tag{2.33}$$

2.3.3 Conditional Probability Mass Function

In some situations we wish to compute the pmf of some random variables given that some other variables are known to have certain values. For example, in Example 2.21, we may wish to find the pmf of X_2 given that $X_1 = 0$. This is known as the *conditional* pmf and is denoted by $P(x_2|x_1)$, which means $\Pr(X_2 = x_2|X_1 = x_1)$. The conditional pmf is the ratio of the joint pmf to the marginal pmf, that is,

$$P(x_2|x_1) = \frac{P(x_1, x_2)}{P(x_1)}, \quad x_2 \in S(X_2), \tag{2.34}$$

where $P(x_1, x_2)$ is the joint density of X_1 and X_2 and x_1 is assumed to be given.

Thus, for example, $P(1|1) = 0.2/0.4 = 0.5$, $P(2|1) = 0.1/0.4 = 0.25$, and $P(3|1) = 0.1/0.4 = 0.25$. Note that

$$\sum_{x_2=1}^{3} P(x_2|x_1) = 1 \quad \forall x_1 \in S(X_1)$$

because every conditional pmf is a pmf, that is, $P(x_2|x_1)$ must satisfy (2.1).

2.3.4 Covariance and Correlation

We have seen that from the joint pmf one can obtain the marginal pmf for each of the variables, $P_1(x_1), P_2(x_2), \ldots, P_n(x_n)$. From these marginals, one can compute the means, $\mu_1, \mu_2, \ldots, \mu_n$, and variances, $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, using (2.4) and (2.7), respectively. In addition to the means and variances, one can also compute the *covariance* between every pair of variables. The covariance between X_i and X_j, denoted by σ_{ij}, is defined as

$$\sigma_{ij} = E(X_i - \mu_i)(X_j - \mu_j) \tag{2.35}$$

$$= \sum_{x_i \in S(X_i)} \sum_{x_j \in S(X_j)} (x_i - \mu_i)(x_j - \mu_j) P(x_i, x_j), \tag{2.36}$$

where $P(x_i, x_j)$ is the joint pmf of X_i and X_j, which is obtained by summing the joint pmf over all possible values of all variables other than X_i and X_j. Note that

$$\sigma_{ii} = E(X_i - \mu_i)(X_i - \mu_i) = E(X_i - \mu_i)^2 = \sigma_i^2,$$

which shows that the covariance of a variable and itself is the variance of the variable.

Example 2.22 (Means, variances, and covariances). Consider the joint pmf in Table 2.5. The computations for the means, variances, and covariance are shown in Tables 2.6 and 2.7, from which we can see that

$$\mu_1 = 0.4, \quad \mu_2 = 2, \quad \sigma_1^2 = 0.240, \quad \sigma_2^2 = 0.6, \quad \text{and} \quad \sigma_{12} = -0.1. \tag{2.37}$$

∎

Table 2.6: Computations of the Means and Variances.

Variable X_1					
x_1	$P(x_1)$	$x_1P_1(x_1)$	$x_1-\mu_1$	$(x_1-\mu_1)^2$	$(x_1-\mu_1)^2P_1(x_1)$
0	0.6	0.0	-0.4	0.16	0.096
1	0.4	0.4	0.6	0.36	0.144
Total	1.0	0.4			0.240
Variable X_2					
x_2	$P(x_2)$	$x_2P_2(x_2)$	$x_2-\mu_2$	$(x_2-\mu_2)^2$	$(x_2-\mu_2)^2P_2(x_2)$
1	0.3	0.3	-1	1	0.3
2	0.4	0.8	0	0	0.0
3	0.3	0.9	1	1	0.3
Total	1.0	2.0			0.6

Table 2.7: Computations of the Covariance Between X_1 and X_2.

x_1	x_2	$P(x_1,x_2)$	$x_1-\mu_1$	$x_2-\mu_2$	$(x_1-\mu_1)(x_2-\mu_2)P(x_1,x_2)$
0	1	0.1	-0.4	-1	0.04
0	2	0.3	-0.4	0	0.00
0	3	0.2	-0.4	1	-0.08
1	1	0.2	0.6	-1	-0.12
1	2	0.1	0.6	0	0.00
1	3	0.1	0.6	1	0.06
		1			-0.10

The covariance between two variables gives information about the *direction* of the relationship between the two variables. If it is positive, the two variables are said to be *positively correlated* and, if it is negative, they are said to be *negatively correlated*. Because σ_{12} in the above example is negative, X_1 and X_2 are negatively correlated.

A graphical interpretation of the covariance between two variables X and Y is as follows. Let us draw all points with positive probabilities in a Cartesian plane. A typical point (x, y) is shown in Figure 2.6. A vertical line at $x = \mu_X$ and a horizontal line at $y = \mu_Y$ divide the plane into four quadrants. Note that the absolute value of the product $(x - \mu_X)(y - \mu_Y)$ is equal to the area of the shaded rectangle shown in Figure 2.6. Note that this area is zero when $x = \mu_X$ or $y = \mu_Y$. The area gets larger as the point (x, y) gets farther away from the point (μ_X, μ_Y). Note also that the product $(x - \mu_X)(y - \mu_Y)$ is positive in the first and third quadrants and negative in the second and fourth quadrants. This is indicated by the + and – signs in Figure 2.6. The covariance is the weighted sum of these products with weights equal to $\Pr(X = x, Y = y)$. If the sum of the

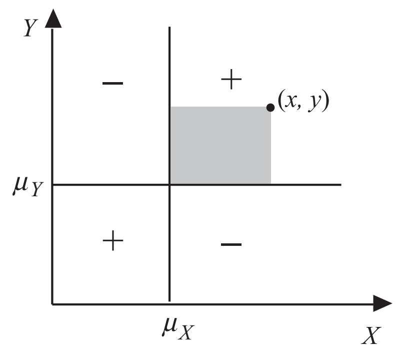

Figure 2.6: A graphical illustration of the covariance between X and Y.

weighted positive terms (those in the first and third quadrants) is equal to the sum of the weighted negative terms (those in the second and fourth quadrants), then the covariance is zero (the negative terms annihilate the positive ones). On the other hand if sum of the weighted positive terms exceeds that of the sum of the weighted negative terms, then the covariance is positive; otherwise it is negative.

Although the covariance between two variables gives information about the direction of the relationship between the two variables, it does not tell us much about the strength of the relationship between the two variables because it is affected by the unit of measurements. That is, if we change the unit of measurement (e.g., from dollars to thousands of dollars), the covariance will change accordingly.

A measure of association that is not affected by changes in unit of measurement is the *correlation coefficient.* The correlation coefficient between two variables X_i and X_j, denoted by ρ_{ij}, is defined as

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}, \tag{2.38}$$

that is, it is the covariance divided by the product of the two standard deviations.

It can be shown that $-1 \leq \rho_{ij} \leq 1$. The correlation ρ_{ij} measures linear association between the two variables. That is, if $\rho_{ij} = \pm 1$, then one variable is a linear function of the other. If $\rho_{ij} = 0$, it means only that the two variables are not linearly related (they may be nonlinearly related, however). In the above example, $\rho_{12} = -0.1/(\sqrt{0.144240 \times 0.6}) = -0.264$, hence X_1 and X_2 are negatively, but mildly correlated.

All considerations made for the graphical interpretation of the covariance are also valid for the correlation coefficient because of its definition. Figure 2.7 is an illustration showing the correspondence between the scatter diagram and the values of σ_X, σ_Y, and ρ_{XY}.

When we deal with a multivariate random variable $X = \{X_1, X_2, \ldots, X_k\}$, it is convenient to summarize their means, variances, covariances, and correlations

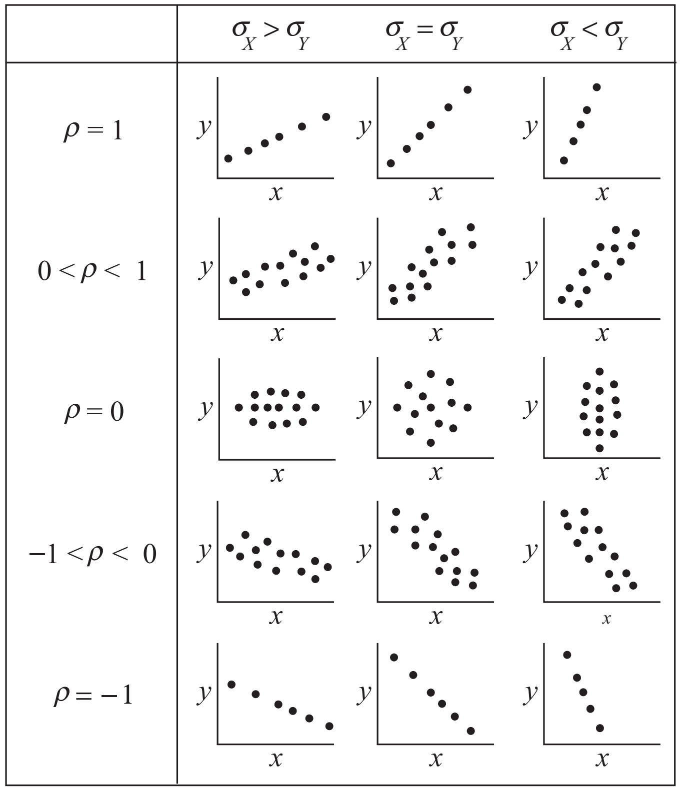

Figure 2.7: Graphical illustration of the correlation coefficient.

as follows. The means are displayed in a $k \times 1$ vector and the variances and covariances are displayed in a $k \times k$ matrix as follows:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_k^2 \end{bmatrix}, \quad \boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1k} \\ \rho_{21} & 1 & \dots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \dots & 1 \end{bmatrix}. \tag{2.39}$$

The vector $\boldsymbol{\mu}$ is called the *mean vector*, the matrix $\boldsymbol{\Sigma}$ is called the *variance-covariance* matrix or just the *covariance matrix*, and $\boldsymbol{\rho}$ is called the *correlation* matrix. The covariance matrix contains the variances on the diagonal and the covariances on the off-diagonal. Note that the covariance and correlation matrices are symmetric, that is, $\sigma_{ij} = \sigma_{ji}$ and $\rho_{ij} = \rho_{ji}$. Note also that the diagonal elements of $\boldsymbol{\rho}$ are all ones because X_i has a perfect correlation with itself.

For example, the means, variances, and covariance in (2.37) can be summarized as

$$\boldsymbol{\mu} = \begin{bmatrix} 0.4 \\ 2.0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.240 & -0.1 \\ -0.1 & 0.6 \end{bmatrix}, \text{and} \quad \boldsymbol{\rho} = \begin{bmatrix} 1 & -0.264 \\ -0.264 & 1 \end{bmatrix}.$$

2.4 Common Discrete Multivariate Models

In this section we deal with some multivariate models of interest.

2.4.1 Multinomial Distribution

We have seen in Section 2.2.3 that the binomial random variable results from random experiments that each has two possible outcomes. If a random experiment has more than two outcomes, the resultant random variable is called a *multinomial* random variable. Suppose that we perform an experiment with k possible outcomes $r_1, \ldots, r_k$, with probabilities $p_1, \ldots, p_k$, respectively. Since the outcomes are mutually exclusive and collectively exhaustive, these probabilities must satisfy $\sum_{i=1}^{k} p_i = 1$. If we repeat this experiment n times and let X_i be the numbers of times we obtain outcomes r_i, for $i = 1, \ldots, k$, then $X = \{X_1, \ldots, X_k\}$ is a multinomial random variable, which is denoted by $M(n; p_1, \ldots, p_k)$. The pmf of $M(n; p_1, \ldots, p_k)$ is

$$P(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1! x_2! \ \ldots \ x_k!} p_1^{x_1} \ p_2^{x_2} \ \ldots \ p_k^{x_k}, \tag{2.40}$$

where $n!/(x_1!x_2! \ \ldots \ x_k!)$ is the number of possible combinations that lead to the desired outcome. Note that $P(x_1, x_2, \ldots, x_k)$ means $\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$. Note also that the support of multinomial random variable X, whose pmf is given by (2.40), is

$$S(X) = \{x_1, x_2, \ldots, x_k : x_i = 0, 1, \ldots, n, \sum_{i=1}^{k} x_i = n\}. \tag{2.41}$$

Example 2.23 (Different failure types). Suppose that we wish to determine the strengths of six plates of fiberglass. Suppose also that there are three possible types of failures. The probabilities of these types of failures are $0.2, 0.3$, and 0.5, respectively. Let X_i be the number of plates with failure of type i. Then, $X = \{X_1, X_2, X_3\}$ is $M(6; 0.2, 0.3, 0.5)$. Thus, for example, the probability of having $2, 1$, and 3 failures of the three types is given by

$$P(2, 1, 3) = \frac{6!}{2!1!3!} \ 0.2^2 \ 0.3^1 \ 0.5^3 = 0.09.$$

■

Example 2.24 (Traffic in a square). A car when arriving at a square can choose among four different streets S_1, S_2, S_3, and S_4 with probabilities

p_1, p_2, p_3, and p_4, respectively. Find the probability that of 10 cars arriving at the square, 3 will take street S_1, 4 will take street S_2, 2 will take street S_3, and 1 will take street S_4.

Since we have four possible outcomes per car, the random variable of the number of cars taking each of the four streets is $M(10; p_1, p_2, p_3, p_4)$. Then, the required probability is

$$P(3, 4, 2, 1) = \frac{10!}{3!4!2!1!} \, p_1^3 \, p_2^4 \, p_3^2 \, p_4^1.$$

■

In a multinomial random variable, the mean and variance of X_i, and the covariance of X_i and X_j are

$$\mu_i = np_i, \quad \sigma_{ii}^2 = np_i(1 - p_i), \quad \text{and} \quad \sigma_{ij} = -np_i p_j,$$

respectively. Thus, all pairs of variables are negatively correlated. This is expected because they are nonnegative integers that sum to n; hence, if one has a large value, the others must necessarily have small values.

The multinomial family of random variables is reproductive with respect to parameter n, that is, if $X_1 \sim M(n_1; p_1, \ldots, p_k)$ and $X_2 \sim M(n_2; p_1, ..., p_k)$, then

$$X_1 + X_2 \sim M(n_1 + n_2; p_1, \ldots, p_k).$$

2.4.2 Multivariate Hypergeometric Distribution

We have seen in Section 2.2.6 that the hypergeometric distribution arises from sampling without replacement from a finite population with two groups (defective and nondefective). If the finite population consists of k groups, the resulting distribution is a *multivariate hypergeometric* distribution. Suppose the population consists of N products of which $D_1, D_2, \ldots, D_k$ are of the k types, with $\sum_{i=1}^{k} D_i = N$. Suppose we wish to draw a random sample of size $n < N$ from this population without replacement. The random variable $X = \{X_1, \ldots, X_k\}$, where X_i is the number of products of type i in the sample, is the multivariate hypergeometric variable and is denoted $MHG(N, p_1, \ldots, p_k, n)$, where $p_i = D_i/N$ is the proportion of products of type i in the population. The pmf of $MHG(N, p_1, \ldots, p_k, n)$ is

$$P(x_1, \ldots, x_k) = \frac{\binom{D_1}{x_1} \cdots \binom{D_k}{x_k}}{\binom{N}{n}}, \quad \sum_{i=1}^{k} x_i = n, \quad \sum_{i=1}^{k} D_i = N. \tag{2.42}$$

Example 2.25 (Urn problem). Suppose an urn contains 20 balls of which 5 are white, 10 are black, and 5 are red. We draw a sample of size 10 without replacement. What is the probability that the drawn sample contains exactly 2 white, 5 black and 3 red balls? Here, $N = 20$, $p_1 = 5/20 = 1/4$, $p_2 = 10/20 = 1/2$, $p_3 = 5/20 = 1/4$, and $n = 10$. Letting $X = \{X_1, X_2, X_3\}$ be the number of white, black and red balls, then X is $MHG\left(20, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 10\right)$. From (2.42), we then have

$$P(2,5,3) = \frac{\binom{5}{2}\binom{10}{5}\binom{5}{3}}{\binom{20}{10}} = \frac{6300}{46189} = 0.136.$$

Exercises

2.1 Show that the mean and variance of a Bernoulli random variable, with success probability p, are $\mu = p$ and $\sigma^2 = p(1-p)$.

2.2 Show that the mean and variance of a $B(n,p)$ random variable are $\mu = np$ and $\sigma^2 = np(1-p)$.

2.3 For the $B(6, 0.2)$ random variable in Exercise 2.9, whose pmf is given in (2.20), compute the probability of each of the following events:

(a) At least one structure is in poor condition.

(b) At most four structures are in poor conditions.

(c) Between two and four structures are in poor conditions.

2.4 Let $X = X_1 + \ldots + X_n$, where $X_i, i = 1, \ldots, n$, are identical and independent Bernoulli random variables with probability of success equal to p. Use the reproductivity property of the binomial random variable to show that $X \sim B(n,p)$.

2.5 Show that the pmf of any $G(p)$ random variable is decreasing in x.

2.6 Suppose that the company in Example 2.14 wishes to fill two vacant positions.

(a) What is the probability that at least four interviews will be required to fill the two vacant positions?

(b) If an interview costs the company \$500 on the average, what is the expected cost of filling the two vacant positions?

2.7 Prove the result in Equation (2.27).

2.8 Use (2.34) to show that the pmf of the nonzero Poisson random variable is given by (2.31).

2.9 The probability of a river to be flooded at a certain location during a period of 1 year is 0.02.

(a) Find the pmf of the number of floods to be registered during a period of 20 years.

(b) Find the pmf of the number of years required until the first with a flood.

(c) Find the pmf of the number of years required until the fifth with a flood.

2.10 Assume that the occurrence of earthquakes with intensity above a given threshold is Poissonian with rate of two earthquakes per year. Compute:

(a) The probability of having no earthquakes in a period of 6 months (the time a dam is being repaired).

(b) The pmf of the number of earthquakes occurring in a period of 5 years.

(c) Discuss the validity of the assumptions for using the Poisson model.

2.11 It is said that a system (e.g., a dam or a dike) has been designed for the N-year if it withstands floods that occur once in N years, that is, its probability of occurrence is $1/N$ in any year. Assuming that floods in different years are independent, calculate:

(a) The probability of having a flood larger or equal to the N-year flood during a period of 50 years.

(b) The probability of having one or more such a floods in 50 years.

(c) If a company designs 20 independent systems (located far enough) for the 500-year flood, what is the cdf of the number of systems that will fail in 50 years?

2.12 Compute the mean vector, covariance matrix, and correlation matrix for each of the following multinomial random variables:

(a) The three variables in Example 2.23.

(b) The four variables in Example 2.24.

2.13 For the $M(n; p_i, \ldots, p_k)$ random variable, show that the correlation coefficient between X_i and X_j is

$$\rho_{ij} = -\sqrt{p_i\, p_j/[(1-p_i)(1-p_j)]}.$$

2.14 The presence of cracks in a long cable is ruled by a Poisson process of intensity of two cracks per meter. Twenty percent of them are of critical size (large). Determine:

(a) The pmf of the number of cracks in a piece of 10 m.

(b) The length up to the first crack.

(c) The distance from the origin of the piece to the third crack.

(d) The number of cracks of critical size in the piece of 10 m.

2.15 The number of waves of height above 8 m at a given location and during a storm follows a Poisson process of intensity of three waves per hour. Determine:

(a) The pmf of the number of waves of height above 8 m during a storm of 5 hours.

(b) The time up to the first wave of height above 8 m.

(c) The time between the third and the sixth waves of height above 8 m.

(d) If a rouble mound breakwater fails after 12 such waves, obtain the probability of failure of the breakwater during a storm of 2 hours duration.

2.16 The number of important floods during the last 50 years at a given location was 20. Determine:

(a) The pmf of the number floods during the next 10 years.

(b) The mean number of yearly floods required for having a probability of 0.9 of no floods in the next 10 years.

2.17 Thirty percent of car accidents involve deaths (or fatal). Determine:

(a) The number of fatal accidents of a set of 20 accidents.

(b) The number of accidents up to a fatal accident.

(c) The number of accidents up to the fifth fatal accident.

2.18 Derive the mean vector, covariance matrix, and correlation matrix for the multivariate hypergeometric distribution in (2.42).

Chapter 3

Continuous Probabilistic Models

In Chapter 2 we discussed several commonly used discrete random variables. This chapter deals with continuous random variables. We start in Section 3.1 with a discussion of some methods for defining the probability of univariate continuous random variables that include probability density function, cumulative distribution functions, and moments of random variables. Then, commonly used continuous univariate random variables are presented in Section 3.2. They are viewed with an eye on their application to extremes. Section 3.3 is devoted to truncated distributions, which have important applications. Section 3.4 present four important functions associated with random variables. These are the survival, hazard, moment generating, and characteristic functions. General multivariate continuous random variables are dealt with in Section 3.5. Finally, some commonly used multivariate models are presented in Section 3.6.

3.1 Univariate Continuous Random Variables

As in the case of discrete random variables, there are several alternative ways for assigning and calculating probabilities associated with continuous random variables; the most important of them are given below.

3.1.1 Probability Density Function

As mentioned in Chapter 2, *continuous* random variables take an uncountable set of real values. Every continuous random variable has a *probability density function* (pdf). The pdf of a continuous random variable X is denoted by $f_X(x)$. For notational simplicity we sometimes use $f(x)$ instead of $f_X(x)$. Note that $f(x)$ is not $\Pr(X = x)$, as in the discrete case. But it is the height of the density curve at the point x. Also, if we integrate $f(x)$ on a given set A, we obtain $\Pr(X \in A)$.

Every pdf $f(x)$ must satisfy two conditions:

$$f(x) \geq 0 \quad \forall x \tag{3.1}$$

and

$$\int_{x \in S(X)} f(x)dx = 1, \tag{3.2}$$

where $S(X)$ is the support of the random variable X, the set of all values x for which $f(x) > 0$.

3.1.2 Cumulative Distribution Function

Every random variable also has a *cumulative distribution function* (cdf). The cdf of a random variable X, denoted by $F(x)$, is a function that assigns to each real value x the probability of X having values less than or equal to x, that is,

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt, \tag{3.3}$$

which implies that

$$f(x) = \frac{dF(x)}{dx}. \tag{3.4}$$

The probability that the random variable X takes values in the interval $(a, b]$, with $a \leq b$, is given by

$$\Pr(a < X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a). \tag{3.5}$$

Thus, $\Pr(a < X \leq b)$ is the area under the pdf on top of the interval $(a, b]$, as can be seen in Figure 3.1, which shows the graphs of the pdf and cdf of a continuous random variable X. Note that, while $f(x)$ is the height of the density curve at x, $F(x)$ is the area under the curve to the left of x. From (3.2), the area under the pdf of any continuous random variable is 1.

Note also that

$$\Pr(X = x) = \Pr(x < X \leq x) = F(x) - F(x) = 0, \tag{3.6}$$

that is, while it is possible for a continuous random variable X to take a given value in its support, it is improbable that it will take this exact value. This is due to the fact that there are uncountably many possible values.

The cdf has the following properties as a direct consequence of the definitions of cdf and probability:

1. $F(-\infty) = 0$ and $F(\infty) = 1$.

2. $F(x)$ is nondecreasing and right continuous.

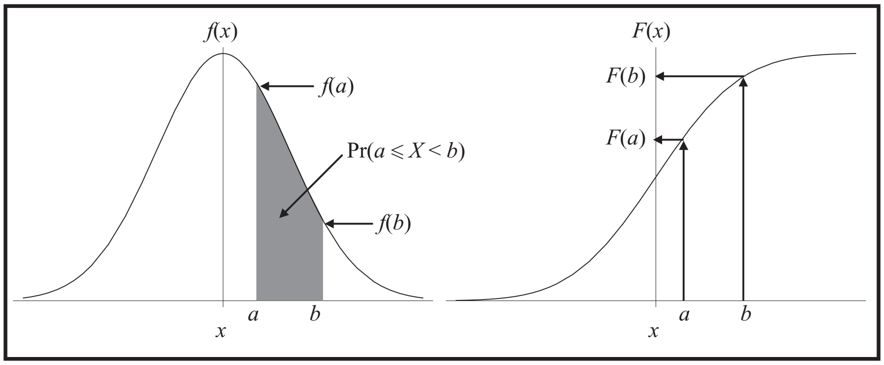

Figure 3.1: Graphs of the pdf and cdf of a continuous random variable X. The pdf, $f(x)$, is the height of the curve at x, and the cdf, $F(x)$, is the area under $f(x)$ to the left of x. Then $\Pr(a < X \leq b) = F(b) - F(a)$ is the area under the pdf on top of the interval $(a, b]$.

3.1.3 Moments

Let $g(X)$ be a function of a continuous random variable X. The *expected value* of $g(X)$ is defined by

$$E[g(X)] = \int_{x \in S(X)} g(x) f(x) dx. \tag{3.7}$$

For example, letting $g(X) = X^r$, we obtain the rth *moment* of the continuous random variable X,

$$E(X^r) = \int_{x \in S(X)} x^r f(x) dx. \tag{3.8}$$

When $r = 1$, we obtain the *mean*, μ, of the continuous random variable X,

$$\mu = E(X) = \int_{x \in S(X)} x\, f(x) dx. \tag{3.9}$$

Letting $g(X) = (X - \mu)^r$, we obtain the rth *central moment*,

$$E[(X - \mu)^r] = \int_{x \in S(X)} (x - \mu)^r f(x) dx. \tag{3.10}$$

When $r = 2$, we obtain the second central moment of the continuous random variable X, that is,

$$\sigma^2 = E[(X - \mu)^2] = \int_{x \in S(X)} (x - \mu)^2 f(x) dx, \tag{3.11}$$

which is known as the *variance*. The *standard deviation*, σ, of the random variable X is the positive square root of its variance. The variance can also be expressed as

$$\sigma^2 = E(X^2) - \mu^2, \tag{3.12}$$

where

$$E(X^2) = \int_{x \in S(X)} x^2 f(x) dx.$$

The expected value operator in the continuous case has the same properties that it has in the discrete case (see page 25).

3.2 Common Continuous Univariate Models

In this section we present several important continuous random variables that often arise in extreme value applications. For more detailed descriptions as well as additional models, see, for example, the books by Christensen (1984), Balakrishnan and Nevzorov (2003) Johnson et al. (1994, 1995), Ross (1992), and Wackerly et al. (2002).

3.2.1 Continuous Uniform Distribution

The continuous uniform random variable on the interval $[\alpha, \beta]$, denoted by $U(\alpha, \beta)$, has the following pdf

$$f(x) = \frac{1}{\beta - \alpha}, \quad \alpha \le x \le \beta, \tag{3.13}$$

from which it follows that the cdf can be written as

$$F(x) = \begin{cases} 0, & \text{if } x < \alpha \\ \dfrac{x - \alpha}{\beta - \alpha}, & \text{if } \alpha \le x < \beta, \\ 1, & \text{if } x \ge \beta. \end{cases}$$

The mean and variance of X are

$$\mu = \frac{\alpha + \beta}{2} \quad \text{and} \quad \sigma^2 = \frac{(\beta - \alpha)^2}{12}. \tag{3.14}$$

A special case of $U(\alpha, \beta)$ is the *standard uniform* random variable, $U(0, 1)$ obtained by setting $\alpha = 0$ and $\beta = 1$. The pdf and cdf of $U(0, 1)$ are

$$f(x) = 1, \quad 0 \le x \le 1, \tag{3.15}$$

and

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

Figure 3.2 shows the pdf and cdf of the standard uniform random variable.

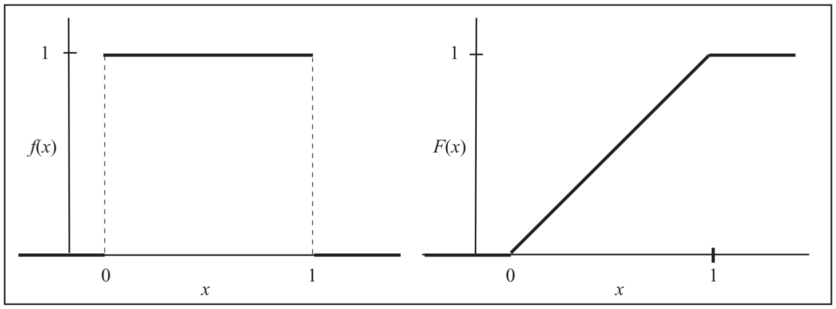

Figure 3.2: The pdf and cdf of the standard uniform random variable.

Example 3.1 (Birth time). If the times of birth are random variables assumed to be uniform on the interval $[0, 24]$, that is, all times in a given 24-hour period are equally possible, then the time of birth X is a uniform random variable, $U(0, 24)$, with pdf

$$f(x) = 1/24, \quad 0 \leq x \leq 24.$$

Note that the uniform model is valid so long as births occur naturally, that is, no induced births, for example. ■

Example 3.2 (Accidents). Let X be the distance in km from a hospital to the location where an accident occurs on a highway of 20 km length. Then, we may assume that X is $U(0, 20)$ random variable. The validity of this assumption requires certain conditions such as the road be straight and homogeneous and the drivers' abilities are constant over the 20-km highway, see Example 3.21. ■

The family of uniform random variables is stable with respect to changes of location and scale, that is, if X is $U(\alpha, \beta)$, then the variable $Y = cX + d$ is uniform $U(c\alpha + d, c\beta + d)$.

Example 3.3 (Temperatures). Suppose that the temperature, in degrees Celsius, at a given time and location is $U(30, 40)$ random variable. Since $F = 1.8C + 32$, where F and C are the temperatures measured in degrees Fahrenheit and Celsius, respectively, then the temperature in degrees Fahrenheit is an $U(86, 104)$ random variable. ■

3.2.2 Exponential Distribution

Let X be the time between two consecutive Poisson events with intensity λ events per unit of time (see Section 2.2.7) such as the time between failures of machines or the time between arrivals at checkout counter. That is, we start at the time when the first event occurs and measure the time to the next event. In

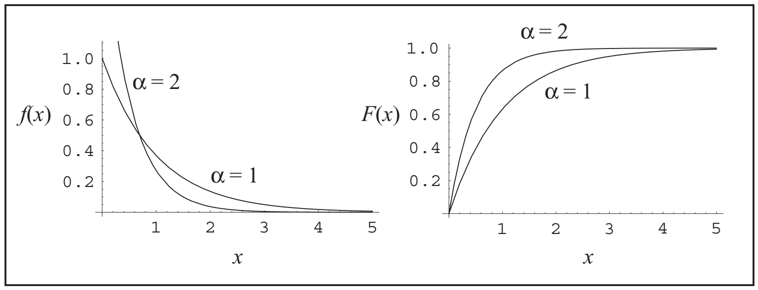

Figure 3.3: An example of the pdf and cdf of two exponential random variables.

other words, X is the *interarrival* time. Then X is a continuous random variable. What is the pdf and cdf of X? Consider the event that X exceeds x, that is, the second event occurs after time x since the occurrence of the first event. The probability of this event is $\Pr(X > x) = 1 - \Pr(X \leq x) = 1 - F(x)$, where $F(x)$ is the cdf of the random variable X. This event, however, is equivalent to saying that no Poisson events have occurred before time x. Replacing λ by λx in the Poisson pmf in (2.28), the probability of obtaining zero Poisson events is $P(0) = e^{-\lambda x}$. Therefore, we have

$$1 - F(x) = e^{-\lambda x},$$

from which it follows that the cdf of X is

$$F(x) = \begin{cases} 0, & \text{if} \quad x < 0, \\ 1 - e^{-\lambda x}, & \text{if} \quad x \geq 0, \ \lambda > 0. \end{cases}$$

Taking the derivative of $F(x)$ with respect to x, we obtain the pdf

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0, & \text{if} \quad x < 0, \\ \lambda e^{-\lambda x}, & \text{if} \quad x \geq 0. \end{cases} \tag{3.16}$$

The random variable X whose pdf is given in (3.16) is called an *exponential* random variable with parameter λ. When X is replaced by $-Y$ in (3.16), we obtain the pdf of the *reversed* exponential random variable,

$$f(y) = \begin{cases} \lambda e^{\lambda y}, & \text{if} \quad y < 0, \ \lambda > 0, \\ 0, & \text{if} \quad y \geq 0. \end{cases} \tag{3.17}$$

The graphs of the pdf and cdf of two exponential random variables are shown in Figure 3.3. It can be shown that the mean and variance of the exponential random variable are

$$\mu = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = \frac{1}{\lambda^2}. \tag{3.18}$$

The pdf of the exponential distribution in (3.16) can also be expressed as

$$f(x) = \begin{cases} 0, & \text{if} \quad x < 0, \\ \dfrac{1}{\delta} e^{-x/\delta}, & \text{if} \quad x \geq 0, \ \delta > 0. \end{cases} \tag{3.19}$$

This is simply a reparameterization of (3.16), where λ is replaced by $1/\delta$. In this form the cdf is

$$F(x) = \begin{cases} 0, & \text{if} \quad x < 0, \\ 1 - e^{-x/\delta}, & \text{if} \quad x \geq 0, \ \delta > 0, \end{cases}$$

and the mean and variance are simply $\mu = \delta$ and $\sigma^2 = \delta^2$, respectively.

Exponential random variables have the so-called *memoryless* or *no-aging* property, that is,

$$\Pr(X > a + b | X > a) = \Pr(X > b).$$

In words, if X is associated with lifetime the probability of X exceeding a given time b is the same no matter which time origin a is considered, from which the terminology *no-aging* was derived.

Example 3.4 (Waiting time at an intersection). When a car arrives at the intersection of two roads, it stops and then it needs a minimum time of t_0 seconds without passing cars to initiate the movement. If the arrival time, X, is assumed to be exponential with intensity λ cars/second, the probability of the waiting time to be zero is given by

$$\Pr(X > t_0) = 1 - \Pr(X \leq t_0) = 1 - F(t_0) = e^{-\lambda t_0}.$$

■

Example 3.5 (Time between consecutive storms). Assume that the occurrence of storms is Poissonian with rate λ storms/year. Then, the time until the occurrence of the first storm and the time between consecutive storms are exponential random variables with parameter λ. For example, assume that $\lambda = 5$ storms/year. Then, the probability of the time until the occurrence of the first storm or the time between consecutive storms to be smaller than 1 month is

$$p = \Pr(X < 1/12) = F(1/12) = 1 - e^{-\lambda/12} = 1 - e^{-5/12} = 0.3408.$$

■

For more properties and applications of the exponential distribution, the interested reader may refer to the book by Balakrishnan and Basu (1995).

3.2.3 Gamma Distribution

The Gamma distribution is a generalization of the exponential distribution. Consider a Poisson time process with intensity λ events per unit time. The time it takes for the first event to occur is an exponential random variable with parameter λ. Now, let X be the time up to the occurrence of θ Poisson events. If $\theta = 1$, then X is an exponential random variable, but if $\theta > 1$, then X is a Gamma random variable. What is then the pdf of a Gamma random variable?

To derive the pdf of a Gamma random variable, we first introduce a useful function called the *Gamma function*, which is defined as

$$\Gamma(\theta) = \int_0^\infty y^{\theta-1}e^{-y}dy. \tag{3.20}$$

Some important properties of the Gamma function are

$$\begin{aligned} \Gamma(0.5) &= \sqrt{\pi}, & (3.21)\\ \Gamma(\theta) &= (\theta-1)\Gamma(\theta-1), \quad \text{if } \theta \ge 1, & (3.22)\\ \Gamma(\theta) &= (\theta-1)!, \quad \text{if } \theta \text{ is a positive integer.} & (3.23) \end{aligned}$$

Now, if X is the time it takes for the θ Poisson events to occur, then the probability that X is in the interval $(x, x+dx)$ is $\Pr(x \le X \le x+dx) = f(x)dx$. But this probability is equal to the probability of having $\theta - 1$ Poisson events occurred in a period of duration x times the probability of the occurrence of one event in a period of duration dx. Thus, we have

$$f(x)dx = \frac{e^{-\lambda x}(\lambda x)^{\theta-1}}{(\theta-1)!}\,\lambda\,dx,$$

from which we obtain

$$f(x) = \frac{\lambda^\theta x^{\theta-1}e^{-\lambda x}}{(\theta-1)!}, \quad 0 \le x < \infty. \tag{3.24}$$

Using the property of the Gamma function in (3.23), Equation (3.24) can be written as

$$f(x) = \frac{\lambda^\theta x^{\theta-1}e^{-\lambda x}}{\Gamma(\theta)}, \quad 0 \le x < \infty, \tag{3.25}$$

which is valid for any real positive θ. The pdf in (3.25) is known as the Gamma distribution with parameters θ and λ, and is denoted by $G(\theta, \lambda)$. Note that when $\theta = 1$, the pdf in (3.25) becomes (3.16), which is the pdf of the exponential distribution. The pdf of some Gamma distributions are graphed in Figure 3.4.

In general, the cdf of the Gamma distribution

$$F(x) = \int_0^x f(t)dt = \int_0^x \frac{\lambda^\theta t^{\theta-1}e^{-\lambda t}}{\Gamma(\theta)}dt, \tag{3.26}$$

which is called the *incomplete Gamma function*, does not have a closed form, but can be obtained by numerical integration. For integer θ, (3.26) has a closed-form formula (see Exercise 3.7).

The mean and variance of $G(\theta, \lambda)$ are

$$\mu = \frac{\theta}{\lambda} \quad \text{and} \quad \sigma^2 = \frac{\theta}{\lambda^2}. \tag{3.27}$$

The family of Gamma distributions has the following important properties:

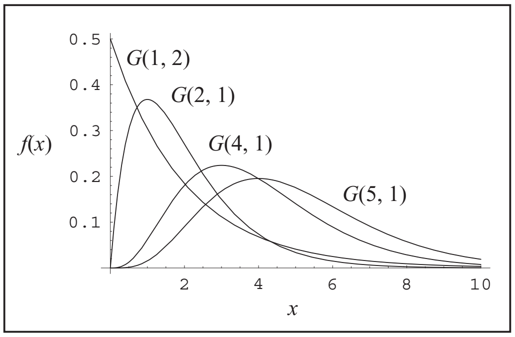

Figure 3.4: Examples of the pdf of some Gamma random variables.

1. It is reproductive with respect to parameter θ, that is, if $X_1 \sim G(\theta_1, \lambda)$ and $X_2 \sim G(\theta_2, \lambda)$ are independent, then
$$X_1 + X_2 \sim G(\theta_1 + \theta_2, \lambda).$$
2. It is stable with respect to scale changes, that is, if X is $G(\theta, \lambda)$, then cX is $G(\theta, \lambda/c)$, see Example 3.20.
3. It is not stable with respect to changes of location. In other words, if X is a Gamma random variable, then $X + a$ is not a Gamma random variable.

Example 3.6 (Parallel system of lamps). A lamp, L, consists of a parallel system of n individual lamps $L_1, \ldots, L_n$ (see Fig. 3.5). A lamp is lit once the previous one fails. Assuming no replacement after failure, the useful life of the lamp is the sum of the lives of the individual lamps. If the life of each of the individual lamps is assumed to be $G(\theta, \lambda)$, then the life of the lamp L is $G(\theta, n\lambda)$ by Property 1. ■

Example 3.7 (Structure). Consider the structure in Figure 3.6, which consists of three bars a, b, and c forming a right-angled triangle with sides $3, 4$, and 5 m, respectively.

The structure is subjected to a vertical pressure or load L. This creates three axial forces (two compression forces F_a and F_b, and one traction force F_c). If the force L is assumed to be $G(2, 1)$, the axial loads F_a, F_b, and F_c are all Gamma random variables. It can be shown that the equilibrium conditions lead to the following system of linear equations:

$$\begin{bmatrix} \cos x & -\cos y & 0 \\ \sin x & \sin y & 0 \\ 0 & \cos y & -1 \end{bmatrix} \begin{bmatrix} F_a \\ F_b \\ F_c \end{bmatrix} = \begin{bmatrix} 0 \\ L \\ 0 \end{bmatrix}. \tag{3.28}$$

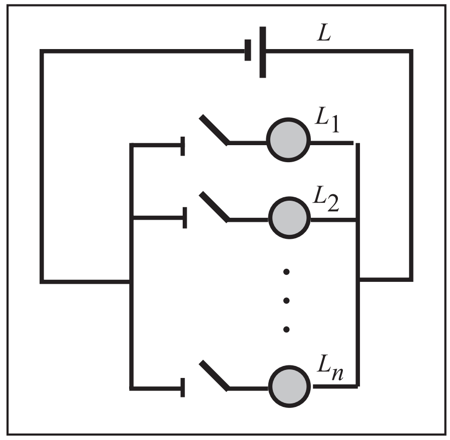

Figure 3.5: A lamp, L consisting of a system of n parallel lamps, $L_1, \ldots, L_n$.

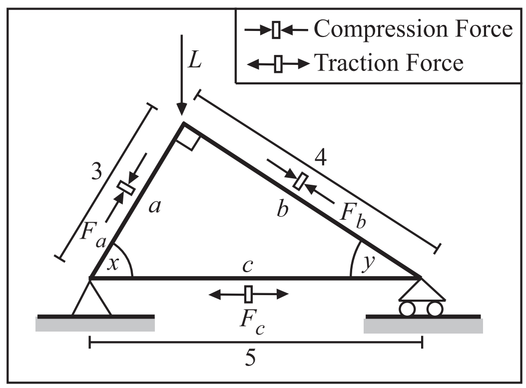

Figure 3.6: A three-bar structure subjected to a vertical force or load L.

Since $\cos x = 3/5$, $\cos y = 4/5$, $\sin x = 4/5$, and $\sin y = 3/5$, the solution of (3.28) gives

$$F_a = \frac{4}{5}\,L, \quad F_b = \frac{3}{5}\,L, \quad \text{and} \quad F_c = \frac{12}{25}\,L.$$

Now, since $G(\theta, \lambda)$ is stable with respect to scale changes, then cL is $G(2, 1/c)$. Therefore, $F_a \sim G(2, 5/4)$, $F_b \sim G(2, 5/3)$ and $F_c \sim G(2, 25/12)$. ■

Example 3.8 (Farmers subsidy). Farmers receive a monetary subsidy after having three floods. Assume that the floods are Poisson events with rate 0.5 floods/year. Then, the time to the third flood is a $G(3, 0.5)$. Thus, the mean time until farmers receive subsidy is $\mu = \theta/\lambda = 3/0.5 = 6$ years. Also, the probability that a subsidy will be received after 10 years is

$$\Pr(X > 10) = 1 - F(10) = 1 - \int_0^{10} \frac{0.5(0.5x)^2 e^{-0.5x}}{2!} \, dx = 1 - 0.8753 = 0.1247.$$

■

3.2.4 Log-Gamma Distribution

The standard form of the log-gamma family of distributions has pdf

$$f(x) = \frac{1}{\Gamma(k)} \, e^{kx} \, \exp\{-e^x\}, \quad -\infty < x < \infty, \; k > 0. \tag{3.29}$$

The pdf in (3.29) can be derived from the logarithmic transformation of gamma distribution, and hence the name *log-gamma distribution*. A three-parameter log-gamma family can be obtained from (3.29), by introducing a location parameter μ and a scale parameter σ, with pdf

$$f(x) = \frac{1}{\sigma \, \Gamma(k)} \, e^{k(x-\mu)/\sigma} \, \exp\{-e^{(x-\mu)/\sigma}\}, \; -\infty < x, \mu < \infty, \; \sigma > 0. \tag{3.30}$$

The corresponding cdf is

$$F(x) = I_{\exp\{(x-\mu)/\sigma\}}(k),$$

where $I_y(k)$ is the incomplete gamma ratio defined as

$$I_y(k) = \frac{1}{\Gamma(k)} \int_0^y e^{-t} \, t^{k-1} \, dt, \quad 0 < y < \infty, \; k > 0.$$

The mean and variance of X are

$$E(X) = \mu + \sigma \, \psi(k) \quad \text{and} \quad Var(X) = \sigma^2 \, \psi'(k), \tag{3.31}$$

where $\psi(z) = \dfrac{d}{dz} \ln \Gamma(z) = \dfrac{\Gamma'(z)}{\Gamma(z)}$ is the digamma function and $\psi'(z)$ is its derivative (trigamma function). Since, for large k, $\psi(k) \sim \ln k$ and $\psi'(k) \sim 1/k$, Prentice (1974) suggested a reparameterized form of the log-gamma density in (3.30) as

$$\begin{aligned} f(x) \quad = \quad & \frac{1}{\sigma \, \Gamma(k)} \, k^{k-\frac{1}{2}} \, e^{\sqrt{k}(x-\mu)/\sigma} \, \exp\{-k e^{(x-\mu)/(\sigma\sqrt{k})}\}, \\ & -\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma > 0. \end{aligned} \tag{3.32}$$

It can be shown that as $k \to \infty$, the density in (3.32) tends to the $Normal(\mu, \sigma^2)$ density function.

Lawless (2003) and Nelson (2004) illustrate the usefulness of the log-gamma density in (3.32) as a lifetime model. Inferential procedures for this model have been discussed by Balakrishnan and Chan (1994, 1995b,a, 1998), DiCiccio (1987), Lawless (1980), Prentice (1974), and Young and Bakir (1987).

3.2.5 Beta Distribution

The Beta random variable is useful for modeling experimental data with range limited to the interval $[0, 1]$. For example, when X is the proportion of impurities in a chemical product or the proportion of time that a machine is under repair, then, X such that $0 \leq X \leq 1$ and a Beta distribution is used to model experimental data collected on such variables. Its name is due to the presence of the *Beta function* in its pdf. The Beta function is defined as

$$\beta(\lambda, \theta) = \int_0^1 x^{\lambda-1}(1-x)^{\theta-1}\, dx, \quad \lambda > 0,\ \theta > 0. \tag{3.33}$$

Note that the Beta function is related to the Gamma function by

$$\beta(\lambda, \theta) = \frac{\Gamma(\lambda)\Gamma(\theta)}{\Gamma(\lambda+\theta)}. \tag{3.34}$$

The pdf of a Beta random variable is given by

$$f(x) = \frac{x^{\lambda-1}(1-x)^{\theta-1}}{\beta(\lambda, \theta)} = \frac{\Gamma(\lambda+\theta)}{\Gamma(\lambda)\Gamma(\theta)} x^{\lambda-1}(1-x)^{\theta-1}, \quad 0 \leq x \leq 1, \tag{3.35}$$

where $\lambda > 0$ and $\theta > 0$. The Beta random variable is denoted by Beta(λ, θ). The cdf of the Beta(λ, θ) is

$$F(x) = \int_0^x f(t)dt = \int_0^x \frac{t^{\lambda-1}(1-t)^{\theta-1}}{\beta(\lambda, \theta)} dt = I_x(\lambda, \theta), \tag{3.36}$$

where $I_x(\lambda, \theta)$ is called the *incomplete Beta ratio*, which can not be given in closed form, but can be obtained by numerical integration.

The mean and variance of the Beta random variable are

$$\mu = \frac{\lambda}{\lambda+\theta} \quad \text{and} \quad \sigma^2 = \frac{\lambda\theta}{(\lambda+\theta+1)(\lambda+\theta)^2},$$

respectively.

The fact that $0 \leq X \leq 1$ does not restrict the use of the Beta random variable because if Y is a random variable defined on the interval $[a, b]$, then

$$X = \frac{Y-a}{b-a}$$

defines a new variable such that $0 \leq X \leq 1$. Therefore, the Beta density function can be applied to a random variable defined on the interval $[a, b]$ by translation and a change of scale.

The interest in this variable is also based on its flexibility, because it can take many different shapes, which can fit different sets of experimental data very well. For example, Figure 3.7 shows different examples of the pdf of the Beta distribution. Two particular cases of the Beta distribution are interesting. Setting $(\lambda = 1, \theta = 1)$, gives the *standard uniform* random variable, $U(0, 1)$, while setting $(\lambda = 2, \theta = 1$ or $\lambda = 1, \theta = 2)$ gives the *triangular* random variable whose cdf is given by $f(x) = 2x$ or $f(x) = 2(1-x)$, $0 \leq x \leq 1$, respectively.

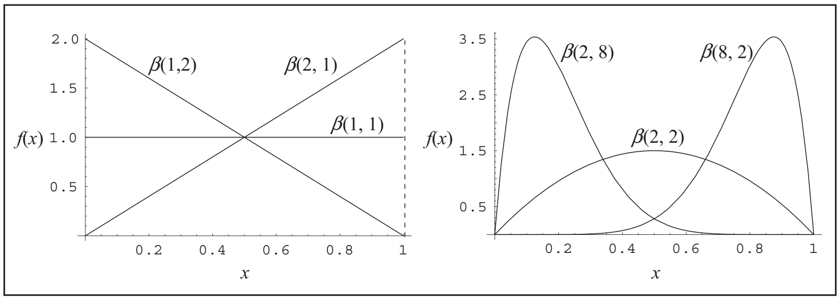

Figure 3.7: Examples showing that the probability density functions of Beta random variables take wide range of different shapes.

3.2.6 Normal or Gaussian Distribution

One of the most important distributions in probability and statistics is the *normal* distribution (also known as the *Gaussian* distribution), which arises in various applications. For example, consider the random variable, X, which is the sum of n *independently and identically distributed* (iid) random variables $X_1, \ldots, X_n$. Then, by the *central limit theorem*, X is asymptotically (as $n \to \infty$) normal, regardless of the form of the distribution of the random variables $X_1, \ldots, X_n$. In fact, the normal distribution also arises in many cases where the random variables to be summed are dependent.

The normal random variable with mean μ and variance σ^2 is denoted by $X \sim N(\mu, \sigma^2)$ and its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty, \tag{3.37}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$. The mean and variance of a normal random variable are μ and σ^2, respectively. Figure 3.8 is a graph of the pdf of a $N(50, 25)$. Note that the pdf is symmetric about the mean $\mu = 50$. Also, the pdf has two inflection points; one on each side of the mean μ and equi-distant from μ. The standard deviation σ is equal to the distance between the mean and the inflection point. Like any other continuous random variable, the area under the curve is 1. The cdf of the normal random variable does not exist in closed form, but can be obtained by numerical integration.

The effect of the parameters μ and σ on the pdf and cdf can be seen in Figure 3.9, which shows the pdf and cdf of two normal random variables with the same mean (zero) but different standard deviations. The higher the standard deviation, the flatter the pdf.

If X is $N(\mu, \sigma^2)$, then the random variable

$$Z = \frac{X - \mu}{\sigma} \tag{3.38}$$

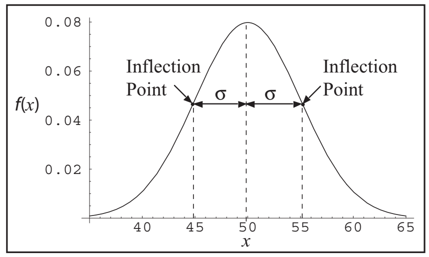

Figure 3.8: The pdf of $N(\mu, \sigma^2)$, where $\mu = 50$ and $\sigma = 5$.

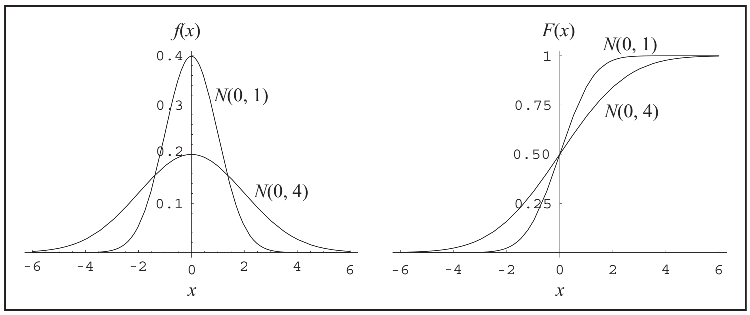

Figure 3.9: Some examples of normal pdfs and cdfs.

is $N(0,1)$. The normal random variable with mean 0 and standard deviation 1 is called the *standard normal* distribution. From (3.37), the pdf of Z is

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}, \quad -\infty < z < \infty, \tag{3.39}$$

and the corresponding cdf is

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} dx, \quad -\infty < z < \infty. \tag{3.40}$$

The pdf and cdf of the standard normal random variable are shown in Figure 3.10. This cdf also does not exist in closed form. However, it has been computed numerically and is given in the Appendix as Table A.1. Note that because of the symmetry of the normal density, we have $\Phi(-z) = 1 - \Phi(z)$.

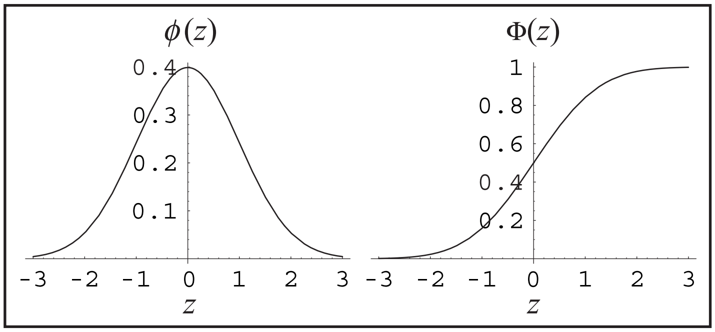

Figure 3.10: The pdf, $\phi(z)$, and the cdf, $\Phi(z)$, of the standard normal random variable, $N(0,1)$.

The main interest of the change of variable in (3.38) is that we can use Table A.1 to calculate probabilities for any other normal distribution. For example, if $X \sim N(\mu, \sigma^2)$, then

$$\Pr(X \le x) = \Pr\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Pr\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where $\Phi(z)$ is the cdf of the standard normal distribution in (3.40), which can be obtained from Table A.1 in the Appendix.

Example 3.9 (Normal variable). Assume that a simple compression strength is a normal random variable with mean $\mu = 200$ kg/cm^2 and a standard deviation 40 kg/cm^2. Then, the probability that the compression strength is at most 140 kg/cm^2 is

$$\begin{aligned} \Pr(X \le 140) &= F(140) = \Phi\left(\frac{140-200}{40}\right) \\ &= \Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668, \end{aligned}$$

where $\Phi(1.5)$ is obtained from Table A.1. Figure 3.11 shows that $\Pr(X \le 140) = \Pr(Z \le -1.5)$. This probability is equal to the shaded areas under the two curves. ∎

The family of normal distributions is reproductive with respect to the parameters μ and σ, that is, if $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, and X_1 and X_2 are independent, then

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

If the random variables X_j, $j = 1, \ldots, n$, are independent and normal $N(\mu_j, \sigma_j^2)$, then the random variable

$$Y = \sum_{j=1}^{n} c_j X_j, \quad c_j \in \mathbb{R}, \quad j = 1, 2, \ldots, n \tag{3.41}$$

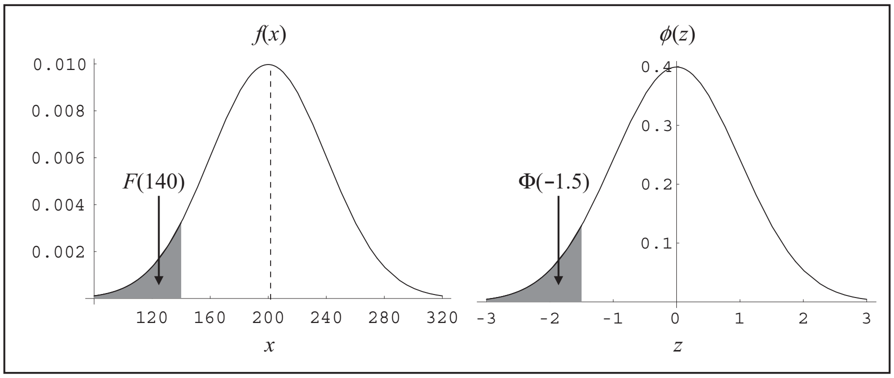

Figure 3.11: The pdf, $f(x)$, where $X \sim N(200, 40^2)$, and the pdf, $\phi(z)$, where $Z \sim N(0, 1)$. The shaded area under $f(x)$ to the left of $x = 140$ is equal to the shaded area under $\phi(z)$ to the left of $z = -1.5$.

is normal with

$$\mu = \sum_{j=1}^{n} c_j \mu_j \quad \text{and} \quad \sigma^2 = \sum_{j=1}^{n} c_j^2 \sigma_j^2 \ .$$

This shows that the normal family is stable with respect to linear combinations.

Normal Approximation to the Binomial Distribution

We know from Section 2.2.3 that the mean and variance of a binomial random variable are $\mu = np$ and $\sigma^2 = np(1-p)$. If the parameter n is large and neither p nor $(1-p)$ are very close to zero, the variable

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{3.42}$$

is approximately $N(0, 1)$. This allows approximating the binomial probabilities using the normal probabilities. In practice, good approximations are obtained if $np, n(1-p) > 5$.

Example 3.10 (Normal approximation). Suppose that 30% of patients entering a hospital with myocardial infarction dies in the hospital. If 2000 patients enter in one year and X is the number of these patients who will die in the hospital, then X is $B(2000, 0.3)$. Since n is large, $np = 600 > 5$, and $n(1-p) = 1400 > 5$, we can use the normal approximation to the binomial. Since $\mu = 2000 \times 0.3 = 600$ patients and $\sigma^2 = 2000 \times 0.3 \times 0.7 = 420$, then X can be approximated by $N(600, 420)$. Thus, for example, the probability that a maximum of 550 patients die in the hospital is

$$\begin{aligned} \Pr(X \leq 550) &\approx \Pr\left(Z \leq (550 - 600)/\sqrt{420}\right) \\ &= \Phi(-2.44) = 1 - \Phi(2.44) \\ &= 1 - 0.9927 = 0.0073, \end{aligned}$$

where $\Phi(2.44)$ is obtained from Table A.1 in the Appendix. ∎

3.2.7 Log-Normal Distribution

A random variable X is log-normal when its logarithm, $\log(X)$, is normal. The pdf of the log-normal random variable can be expressed as

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right], \quad x \geq 0, \tag{3.43}$$

where the parameters μ and σ are the mean and the standard deviation of the initial normal random variable. The mean and variance of the log-normal random variable are

$$e^{\mu+\sigma^2/2} \quad \text{and} \quad e^{2\mu}(e^{2\sigma^2} - e^{\sigma^2}). \tag{3.44}$$

In some applications, the random variables of interest are defined to be the products (instead of sums) of iid positive random variables. In these cases, taking the logarithm of the product yields the sum of the logarithms of its components. Thus, by the central limit theorem, the logarithm of the product of n iid random variables is asymptotically normal.

The log-normal random variable is not reproductive with respect to its parameters μ and σ^2, but stable with respect to products of independent variables, that is, if $X_1 \sim LN(\mu_1, \sigma_1^2)$ and $X_2 \sim LN(\mu_2, \sigma_2^2)$, then

$$X_1 \times X_2 \sim LN(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

3.2.8 Logistic Distribution

A random variable X is said to have a logistic distribution if its cdf is given by:

$$F(x) = \frac{1}{1+\exp\{-(x-\alpha)/\beta\}}, \quad \beta > 0, \quad -\infty < x, \alpha < \infty, \tag{3.45}$$

where α and β are location and scale parameters, respectively. Note that the logistic distribution (3.45) is symmetric about $x = \alpha$ and has a shape similar to that of the normal distribution.

The use of logistic function as a growth curve can be justified as follows. Consider the differential equation:

$$\frac{dF(x)}{dx} = k\left[F(x) - a\right]\left[b - F(x)\right], \tag{3.46}$$

where k, a, and b are constants with $k > 0$ and $b > a$. In other words, the rate of growth is equal to the excess over the initial asymptotic value a times the deficiency compared with final asymptotic value b. The solution of the differential equation (3.46) with $a = 0$ and $b = 1$ (the asymptotic limits of the cdf) is

$$F(x) = \left(1 + ce^{-x/k}\right)^{-1},$$

where c is a constant. This is the same as the logistic distribution in (3.45) with $k = \beta$ and $c = e^{\alpha/\beta}$. Equation (3.46) is used as a model of autocatalysis (see Johnson et al. (1995)).

From (3.45), the pdf of the logistic random variable is

$$f(x) = \frac{\exp\{-(x-\alpha)/\beta\}}{\beta\left[1+\exp\{-(x-\alpha)/\beta\}\right]^2}, \quad \beta > 0, \quad -\infty < x, \alpha < \infty. \tag{3.47}$$

The mean and variance of the logistic random variable are

$$\mu = \alpha \quad \text{and} \quad \sigma^2 = \frac{\pi^2\beta^2}{3}. \tag{3.48}$$

A simple relationship between the cdf (3.45) and the pdf (3.47) is

$$f(x) = \frac{1}{\beta}F(x)[1-F(x)].$$

This relation is useful to establish several properties of the logistic distribution; see, for example, Balakrishnan (1992).

3.2.9 Chi-Square and Chi Distributions

Let $Y_1, \ldots, Y_n$ be independent random variables, where Y_i is distributed as $N(\mu_i, 1)$. Then, the variable

$$X = \sum_{i=1}^{n} Y_i^2$$

is called a *noncentral* χ^2 random variable with n degrees of freedom and *noncentrality parameter* $\lambda = \sum_{i=1}^{n} \mu_i^2$. It is denoted as $\chi_n^2(\lambda)$. When $\mu_i = 0$ for all i, then $\lambda = 0$ and we obtain the *central* χ^2 random variable, which is denoted by χ_n^2. The pdf of the central χ^2 random variable with n degrees of freedom is

$$f(x) = \frac{x^{(n/2)-1}\, e^{-x/2}}{2^{n/2}\, \Gamma(n/2)}, \quad x \geq 0, \tag{3.49}$$

where $\Gamma(.)$ is the Gamma function defined in (3.20). The cdf $F(x)$ can not be given in closed form in general. However, it is available numerically and is given in the Appendix as Table A.3. The mean and variance of a χ_n^2 random variable are

$$\mu = n + \lambda \quad \text{and} \quad \sigma^2 = 2(n + 2\lambda). \tag{3.50}$$

The $\chi_n^2(\lambda)$ variable is reproductive with respect to n and λ, that is, if $X_1 \sim \chi_{n_1}^2(\lambda_1)$ and $X_2 \sim \chi_{n_2}^2(\lambda_2)$, then

$$X_1 + X_2 \sim \chi_{n_1+n_2}^2(\lambda_1 + \lambda_2).$$

The positive square root of a $\chi_n^2(\lambda)$ random variable is called a χ random variable and is denoted by $\chi_n(\lambda)$.

3.2.10 Rayleigh Distribution

An interesting particular case of the χ_n random variable is the *Rayleigh* random variable, which is obtained when $n = 2$. The pdf of the Rayleigh random variable is given by

$$f(x) = \frac{1}{\delta^2} x \exp\left[-\frac{1}{2}\left(\frac{x}{\delta}\right)^2\right], \quad \delta > 0, x \geq 0. \tag{3.51}$$

The corresponding cdf is

$$F(x) = 1 - \exp\left[-\frac{1}{2}\left(\frac{x}{\delta}\right)^2\right]. \tag{3.52}$$

The mean and variance are

$$\mu = \delta\sqrt{\pi/2} \quad \text{and} \quad \sigma^2 = \delta^2(4-\pi)/2. \tag{3.53}$$

The Rayleigh distribution is used, for example, to model wave heights; see, for example, Longuet-Higgins (1975).

3.2.11 Student's t Distribution

Let Z be the normal $N(\lambda, 1)$ and Y be a χ_n^2. If Z and Y are independent, then the random variable

$$X = \frac{Z}{\sqrt{Y/n}}$$

is called the *noncentral Student's t* random variable with n degrees of freedom and noncentrality parameter λ and is denoted by $t_n(\lambda)$. When $\lambda = 0$, we obtain the *central Student's t* random variable, which is denoted by t_n, and its pdf is

$$f(x) = \frac{\Gamma(n+1)/2}{\Gamma(n/2)\sqrt{n\pi}}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad -\infty < x < \infty, \tag{3.54}$$

where $\Gamma(.)$ is the Gamma function defined in (3.20). The cdf $F(x)$ is not simple. However, it is available numerically and is given in the Appendix as Table A.2. The mean and variance of the central t random variable are $\mu = 0$, for $n > 1$, and $\sigma^2 = n/(n-2)$, for $n > 2$, respectively.

3.2.12 F Distribution

Let Y_1 be a χ_m^2 and Y_2 be a χ_n^2, where m and n are positive integers. If Y_1 and Y_2 are independent, then the random variable

$$X = \frac{Y_1/m}{Y_2/n}$$

has an F distribution with m and n degrees of freedom, and is denoted by $F_{(m,n)}$. The corresponding pdf is

$$f(x) = \frac{\Gamma(m+n)}{\Gamma(m/2)\Gamma(n/2} \left(\frac{m}{n}\right)^{m/2} \frac{x^{(m/)-1}}{[1+(m/n)x]^{(m+n)/2}}, \quad x > 0, \tag{3.55}$$

where $\Gamma(.)$ is the Gamma function defined in (3.20). The cdf $F(x)$ is available numerically and three quantiles of which are given in Tables A.4–A.6 in the Appendix. The mean and variance of F random variable are

$$\mu = \frac{n}{n-2}, \quad \text{for } n > 2,$$

and

$$\sigma^2 = \frac{2n^2(n+m-2)}{m(n-2)^2(n-4)}, \quad \text{for } n > 4,$$

respectively.

3.2.13 Weibull Distribution

The Weibull distribution appears very frequently in practical problems when we observe data that represent minima values. The reason why this occurs is given in Chapter 8, where it is shown that, for many parent populations with limited left tail, the limit of the minima of independent samples converges to a Weibull distribution. The pdf of the Weibull random variable is given by

$$f(x) = \frac{\beta}{\delta} \exp\left[-\left(\frac{x-\lambda}{\delta}\right)^{\beta}\right] \left(\frac{x-\lambda}{\delta}\right)^{\beta-1}, \quad x > \lambda, \tag{3.56}$$

and the cdf by

$$F(x) = 1 - \exp\left[-\left(\frac{x-\lambda}{\delta}\right)^{\beta}\right], \quad x \geq \lambda, \tag{3.57}$$

with mean and variance

$$\mu = \lambda + \delta\Gamma\left(1+\frac{1}{\beta}\right) \quad \text{and} \quad \sigma^2 = \delta^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right]. \tag{3.58}$$

Also of interest is the *reversed* Weibull distribution with pdf

$$f(x) = \frac{\beta}{\delta} \exp\left[-\left(\frac{\lambda-x}{\delta}\right)^{\beta}\right] \left(\frac{\lambda-x}{\delta}\right)^{\beta-1}, \quad x < \lambda, \tag{3.59}$$

and cdf

$$F(x) = \exp\left[-\left(\frac{\lambda-x}{\delta}\right)^{\beta}\right], \quad x \leq \lambda, \tag{3.60}$$

with mean and variance

$$\mu = \lambda - \delta\Gamma\left(1+\frac{1}{\beta}\right) \quad \text{and} \quad \sigma^2 = \delta^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right]. \tag{3.61}$$

3.2.14 Gumbel Distribution

The Gumbel distribution appears very frequently in practical problems when we observe data that represent maxima values. The reason why this occurs is presented in Chapter 8, where it is shown that for many parent populations with limited or unlimited left tail, the limit of the maxima of independent samples converges to a Gumbel distribution. The pdf of the Gumbel random variable is given by

$$f(x) = \frac{1}{\delta} \exp\left[\frac{\lambda - x}{\delta} - \exp\left(\frac{\lambda - x}{\delta}\right)\right], \quad -\infty < x < \infty, \tag{3.62}$$

and the cdf by

$$F(x) = \exp\left[-\exp\left(\frac{\lambda - x}{\delta}\right)\right], \quad -\infty < x < \infty, \tag{3.63}$$

with mean and variance

$$\mu = \lambda + 0.57772\delta \quad \text{and} \quad \sigma^2 = \frac{\pi^2\delta^2}{6}. \tag{3.64}$$

Also of interest is the *reversed* Gumbel distribution with pdf

$$f(x) = \frac{1}{\delta} \exp\left[\frac{x - \lambda}{\delta} - \exp\left(\frac{x - \lambda}{\delta}\right)\right], \quad -\infty < x < \infty, \tag{3.65}$$

and cdf

$$F(x) = 1 - \exp\left[-\exp\left(\frac{x - \lambda}{\delta}\right)\right], \quad -\infty < x < \infty, \tag{3.66}$$

with mean and variance

$$\mu = \lambda - 0.57772\delta \quad \text{and} \quad \sigma^2 = \pi^2\delta^2/6. \tag{3.67}$$

3.2.15 Fréchet Distribution

The Fréchet distribution appears very frequently in practical problems when we observe data that represent maxima values. The reason why this occurs is provided in Chapter 8, where it is shown that for many parent populations with unlimited left tail, the limit of the maxima of independent samples converges to a Fréchet distribution.

The pdf of the Fréchet random variable is given by

$$f(x) = \frac{\beta\delta}{(x - \lambda)^2} \exp\left[-\left(\frac{\delta}{x - \lambda}\right)^{\beta}\right] \left(\frac{\delta}{x - \lambda}\right)^{\beta - 1}, \quad x > \lambda, \tag{3.68}$$

and the cdf by

$$F(x) = \exp\left[-\left(\frac{\delta}{x - \lambda}\right)^{\beta}\right], \quad x > \lambda, \tag{3.69}$$

with mean and variance

$$\mu = \lambda + \delta\Gamma\left(1 - \frac{1}{\beta}\right), \quad \text{for } \beta > 1, \tag{3.70}$$

and

$$\sigma^2 = \delta^2\left[\Gamma\left(1 - \frac{2}{\beta}\right) - \Gamma^2\left(1 - \frac{1}{\beta}\right)\right], \quad \text{for } \beta > 2. \tag{3.71}$$

Also of interest is the *reversed* Fréchet distribution with pdf

$$f(x) = \frac{\beta\delta}{(\lambda - x)^2}\exp\left[-\left(\frac{\delta}{\lambda - x}\right)^{\beta}\right]\left(\frac{\delta}{\lambda - x}\right)^{\beta-1}, \quad x < \lambda, \tag{3.72}$$

and cdf

$$F(x) = 1 - \exp\left[-\left(\frac{\delta}{\lambda - x}\right)^{\beta}\right], \quad x < \lambda, \tag{3.73}$$

with mean and variance

$$\mu = \lambda - \delta\Gamma\left(1 - \frac{1}{\beta}\right), \quad \text{for } \beta > 1, \tag{3.74}$$

and

$$\sigma^2 = \delta^2\left[\Gamma\left(1 - \frac{2}{\beta}\right) - \Gamma^2\left(1 - \frac{1}{\beta}\right)\right], \quad \text{for } \beta > 2. \tag{3.75}$$

3.2.16 Generalized Extreme Value Distributions

The generalized extreme value distributions include all distributions that can be obtained as the limit of sequences of maxima and minima values (see Chapter 8). The cdf of the maximal generalized extreme value distribution (GEVD) is given by

$$H(x;\lambda,\delta,\kappa) = \begin{cases} \exp\left\{-\left[1 - \kappa\left(\frac{x - \lambda}{\delta}\right)\right]^{1/\kappa}\right\}, & 1 - \kappa\left(\frac{x - \lambda}{\delta}\right) \ge 0, \kappa \ne 0, \\ \exp\left\{-\exp\left(\frac{\lambda - x}{\delta}\right)\right\}, & -\infty < x < \infty, \kappa = 0, \end{cases} \tag{3.76}$$

where the support is $x \le \lambda + \delta/\kappa$ if $\kappa > 0$, or $x \ge \lambda + \delta/\kappa$ if $\kappa < 0$. The corresponding p-quantile is

$$x_p = \begin{cases} \lambda + \delta\left[1 - (-\log p)^{\kappa}\right]/\kappa, & \text{if } \kappa \ne 0, \\ \lambda - \delta\log(-\log p), & \text{if } \kappa = 0. \end{cases} \tag{3.77}$$

The Gumbel, reversed Weibull and Fréchet distributions are particular cases of the maximal GEVD.

Also of interest is the minimal GEVD with cdf, $H(x;\lambda,\delta,\kappa)$, which is given by

$$\begin{cases} 1-\exp\left\{-\left[1+\kappa\left(\dfrac{x-\lambda}{\delta}\right)\right]^{1/\kappa}\right\}, & 1+\kappa\left(\dfrac{x-\lambda}{\delta}\right)\geq 0, \kappa\neq 0, \\ 1-\exp\left[-\exp\left(\dfrac{x-\lambda}{\delta}\right)\right], & -\infty < x < \infty, \end{cases} \tag{3.78}$$

where the support is $x \geq \lambda - \delta/\kappa$ if $\kappa > 0$, or $x \leq \lambda - \delta/\kappa$ if $\kappa < 0$.

The corresponding p-quantile is

$$x_p = \begin{cases} \lambda - \delta\left[1-(-\log(1-p))^{\kappa}\right]/\kappa, & \text{if } \kappa\neq 0, \\ \lambda + \delta\log(-\log(1-p)), & \text{if } \kappa = 0. \end{cases} \tag{3.79}$$

The reversed Gumbel, Weibull, and reversed Fréchet distributions are particular cases of the minimal GEVD.

3.2.17 Generalized Pareto Distributions

The generalized Pareto distribution arises when you consider excesses of a random variable above or below given thresholds (see Chapter 9). The cdf of the generalized Pareto distribution is given by

$$F(x;\lambda,\kappa) = \begin{cases} 1-\left(1-\dfrac{\kappa x}{\lambda}\right)^{1/\kappa}, & \left(1-\dfrac{\kappa x}{\lambda}\right)\geq 0, \\ & \kappa\neq 0, \lambda > 0, \\ 1-e^{-x/\lambda}, & x\geq 0, \kappa = 0, \lambda > 0, \end{cases} \tag{3.80}$$

where λ and κ are scale and shape parameters, respectively. For $\kappa \neq 0$, the range of x is $0 \leq x \leq \lambda/\kappa$ if $\kappa > 0$, and $x \geq 0$ if $\kappa \leq 0$.

Also of interest if the reversed generalized Pareto distribution with cdf

$$F(x;\lambda,\kappa) = \begin{cases} \left(1+\dfrac{\kappa x}{\lambda}\right)^{1/\kappa}, & \left(1+\dfrac{\kappa x}{\lambda}\right)\geq 0, x\leq 0, \\ & \kappa\neq 0, \lambda > 0, \\ e^{x/\lambda}, & x\leq 0, \kappa = 0, \lambda > 0, \end{cases} \tag{3.81}$$

where λ and κ are scale and shape parameters, respectively. For $\kappa \neq 0$, the range of x is $-\lambda/\kappa \leq x \leq 0$ if $\kappa > 0$, and $x < 0$ if $\kappa \leq 0$.

Finally, we conclude this section with a summary of all the univariate continuous distributions so far discussed in Tables 3.1, 3.2, and 3.3.

Table 3.1: The Probability Density Functions of Some Continuous Random Variables that Frequently Arise in Engineering Applications.

Distribution	$f(x)$
Uniform	$(\beta - \lambda)^{-1}$
Exponential	$\lambda e^{-\lambda x}$
Gamma	$\dfrac{\lambda^\theta x^{\theta-1} e^{-\lambda x}}{\Gamma(\theta)}$
Beta	$\dfrac{x^{\lambda-1}(1-x)^{\theta-1}}{\beta(\lambda,\theta)}$
Normal	$\dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]$
Log-normal	$\dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(\log(x)-\mu)^2}{2\sigma^2}\right]$
Logistic	$\dfrac{\exp\{-(x-\alpha)/\beta\}}{\beta\left[1+\exp\{-(x-\alpha)/\beta\}\right]^2}$
Central χ^2	$\dfrac{x^{(n/2)-1}\, e^{-x/2}}{2^{n/2}\, \Gamma(n/2)}$
Rayleigh	$\delta^{-2} x \exp\left[-\dfrac{1}{2}\left(\dfrac{x}{\delta}\right)^2\right]$
Student's t	$\dfrac{\Gamma(n+1)/2}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\dfrac{x^2}{n}\right)^{-(n+1)/2}$

3.3 Truncated Distributions

In this section, we introduce truncated distributions that are very useful when dealing with extremes wherein only values above or below certain threshold values are often of interest.

Definition 3.1 (Truncated Distributions) *Let X be a random variable. We call the random variables $X|X \leq x_0$, $X|X > x_0$, and $X|x_0 < X \leq x_1$ truncated at x_0 from the right, from the left, or truncated at x_0 from the left and at x_1 from the right, respectively.*

The following theorem gives the corresponding cdfs as a function of the cdf $F_X(x)$ of X.

Table 3.2: The Probability Density Functions of Some Continuous Random Variables that Frequently Arise in Engineering Applications.

Distribution	$f(x)$
Weibull	$\frac{\beta}{\delta}\exp\left[-\left(\frac{x-\lambda}{\delta}\right)^{\beta}\right]\left(\frac{x-\lambda}{\delta}\right)^{\beta-1}$
Reversed Weibull	$\frac{\beta}{\delta}\exp\left[-\left(\frac{\lambda-x}{\delta}\right)^{\beta}\right]\left(\frac{\lambda-x}{\delta}\right)^{\beta-1}$
Gumbel	$\frac{1}{\delta}\exp\left[\frac{\lambda-x}{\delta}-\exp\left(\frac{\lambda-x}{\delta}\right)\right]$
Reversed Gumbel	$\frac{1}{\delta}\exp\left[\frac{x-\lambda}{\delta}-\exp\left(\frac{x-\lambda}{\delta}\right)\right]$
Fréchet	$\frac{\beta\delta}{(x-\lambda)^2}\exp\left[-\left(\frac{\delta}{x-\lambda}\right)^{\beta}\right]\left(\frac{\delta}{x-\lambda}\right)^{\beta-1}$
Reversed Fréchet	$\frac{\beta\delta}{(\lambda-x)^2}\exp\left[-\left(\frac{\delta}{\lambda-x}\right)^{\beta}\right]\left(\frac{\delta}{\lambda-x}\right)^{\beta-1}$
Maximal GPD	$\frac{1}{\lambda}\left(1-\frac{\kappa x}{\lambda}\right)^{1/\kappa-1}$
Minimal GPD	$\frac{1}{\lambda}\left(1+\frac{\kappa x}{\lambda}\right)^{1/\kappa-1}$

Theorem 3.1 (Cdf of a truncated distribution) *The cdf of the truncated random variable $X|X \le x_0$ is*

$$F_{X|X\le x_0}(x) = \begin{cases} F_X(x)/F_X(x_0), & \text{if } x < x_0, \\ 1, & \text{if } x \ge x_0. \end{cases} \tag{3.82}$$

The cdf of the truncated random variable $X|X > x_0$ is

$$F_{X|X>x_0}(x) = \begin{cases} 0, & \text{if } x \le x_0, \\ \dfrac{F_X(x)-F_X(x_0)}{1-F_X(x_0)}, & \text{if } x > x_0, \end{cases} \tag{3.83}$$

Finally, the cdf of the truncated random variable $X|x_0 < X \le x_1$ is

$$F_{X|x_0<X\le x_1}(x) = \begin{cases} 0, & \text{if } x \le x_0, \\ \dfrac{F_X(x)-F_X(x_0)}{F_X(x_1)-F_X(x_0)}, & \text{if } x_0 < x \le x_1, \\ 1, & \text{if } x > 1. \end{cases}$$

Table 3.3: The Means and Variances of Some Continuous Random Variables that Frequently Arise in Engineering Applications.

Distribution	Mean	Variance
Uniform	$(\lambda+\beta)/2$	$(\beta-\lambda)^2/12$
Exponential	$1/\lambda$	$1/\lambda^2$
Gamma	θ/λ	θ/λ^2
Beta	$\lambda/(\lambda+\theta)$	$\lambda\theta/\left[(\lambda+\theta+1)(\lambda+\theta)^2\right]$
Normal	μ	σ^2
Log-normal	$e^{\mu+\sigma^2/2}$	$e^{2\mu}(e^{2\sigma^2}-e^{\sigma^2})$
Central χ^2	n	$2n$
Rayleigh	$\delta\sqrt{\pi/2}$	$\delta^2(4-\pi)/2$
Student's t	0	$n/(n-2)$
Weibull	$\lambda+\delta\Gamma\left(1+\frac{1}{\beta}\right)$	$\delta^2\left[\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right]$
Reversed Weibull	$\lambda-\delta\Gamma\left(1+\frac{1}{\beta}\right)$	$\delta^2\left[\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right]$
Gumbel	$\lambda+0.57772\delta$	$\pi^2\delta^2/6$
Reversed Gumbel	$\lambda-0.57772\delta$	$\pi^2\delta^2/6$
Fréchet	$\lambda+\delta\Gamma\left(1-\frac{1}{\beta}\right)$	$\delta^2\left[\Gamma\left(1-\frac{2}{\beta}\right)-\Gamma^2\left(1-\frac{1}{\beta}\right)\right]$
Reversed Fréchet	$\lambda-\delta\Gamma\left(1-\frac{1}{\beta}\right)$	$\delta^2\left[\Gamma\left(1-\frac{2}{\beta}\right)-\Gamma^2\left(1-\frac{1}{\beta}\right)\right]$
Logistic	α	$\pi^2\beta^2/3$

Proof. Because of the definition of conditional probability we have

$$
\begin{aligned}
F_{X|X\le x_0}(x) &= \Pr(X \le x | X \le x_0) \\
&= \frac{\Pr\left((X \le x) \cap (X \le x_0)\right)}{\Pr(X \le x_0)} \\
&= \frac{\Pr\left((X \le x) \cap (X \le x_0)\right)}{F_X(x_0)} \\
&= \begin{cases} \dfrac{\Pr(X \le x)}{F_X(x_0)} = \dfrac{F_X(x)}{F_X(x_0)}, & \text{if } x < x_0, \\ \dfrac{\Pr(X \le x_0)}{F_X(x_0)} = 1, & \text{if } x \ge x_0. \end{cases}
\end{aligned}
$$

For the left truncation we have

$$\begin{aligned}
F_{X|X>x_0}(x) &= \Pr(X \le x | X > x_0) \\
&= \frac{\Pr\left((X \le x) \cap (X > x_0)\right)}{\Pr(X > x_0)} \\
&= \frac{\Pr(x_0 < X \le x)}{1 - F_X(x_0)} \\
&= \begin{cases} 0, & \text{if } x \le x_0, \\ \dfrac{F_X(x) - F_X(x_0)}{1 - F_X(x_0)}, & \text{if } x > x_0. \end{cases}
\end{aligned}$$

Finally, for truncation on both sides, we have

$$\begin{aligned}
F_{X|x_0<X\le x_1}(x) &= \Pr(X \le x | x_0 < X \le x_1) \\
&= \frac{\Pr\left((X \le x) \cap (x_0 < X \le x_1)\right)}{\Pr(x_0 < X \le x_1)} \\
&= \frac{\Pr(x_0 < X \le \min(x, x_1))}{F_X(x_1) - F_X(x_0)} \\
&= \begin{cases} 0, & \text{if } x \le x_0, \\ \dfrac{F_X(x) - F_X(x_0)}{F_X(x_1) - F_X(x_0)}, & \text{if } x_0 < x \le x_1, \\ 1, & \text{if } x > x_1. \end{cases}
\end{aligned}$$

■

Example 3.11 (Lifetime). The remaining lifetime, X, in years of a patient after suffering a heart attack has a cdf

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \dfrac{1 - e^{-\lambda x}}{1 - e^{-50\lambda}}, & \text{if } 0 \le x < 50, \\ 1, & \text{if } x \ge 50. \end{cases}$$

If a given patient suffered a heart attack 30 years ago, determine the cdf of the remaining lifetime.

Before solving this problem, it is worthwhile mentioning that the given variable is a right-truncated exponential distribution, which implies that no patient can survive above 50 years after suffering a heart attack.

Since the patient suffered a heart attack 30 years ago, we must determine the cdf of the random variable $(X - 30)$ conditioned by $X \ge 30$. Then, we ask for the truncated distribution on the left at 30 and translated 30 units, that is,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \dfrac{e^{-30\lambda} - e^{-\lambda(x+30)}}{e^{-30\lambda} - e^{-50\lambda}}, & \text{if } 0 \le x < 20, \\ 1, & \text{if } x \ge 20. \end{cases}$$

∎

Example 3.12 (Hospital). Suppose that the age, X (in years) of patients entering a hospital has the following pdf:

$$f_X(x) = \begin{cases} \dfrac{\pi}{200} \sin\left(\dfrac{\pi x}{100}\right), & \text{if } 0 \le x < 100, \\ 0, & \text{otherwise.} \end{cases}$$

Then, the pdf for the children younger than 5 years old that enter the hospital is the same density but truncated on the right at $X = 5$. Thus, we have

$$f_X(x) = \begin{cases} \dfrac{\pi \sin(\pi x/100)}{100\,[1 - \cos(5\pi/100)]}, & \text{if } x < 5, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, the pdf for the patients above 60 years of age is the same density but truncated on the left at $X = 60$. Thus, we have

$$f_Z(z) = \begin{cases} \dfrac{\pi \sin(\pi z/100)}{100\,[1 + \cos(60\pi/100)]}, & \text{if } 60 \le z < 100, \\ 0, & \text{otherwise.} \end{cases}$$

∎

Example 3.13 (Screw strength). A factory producing screws states that the strength of screws, R^*, in kg/cm^2 has an exponential distribution $E(\lambda)$. If all the screws are subject to a quality test consisting of applying a test stress of 10 kg/cm^2 and those failing are discarded, determine the pdf of the strength, R, of the accepted screws.

Since after the test the screws with a strength less than 10 kg/cm^2, are discarded, the resulting strength is truncated on the left at 10 kg/cm^2, so that we have

$$f_R(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}}{e^{-10\lambda}} = \lambda e^{-\lambda(x-10)}, & \text{if } x > 10 \text{ kg/cm}^2, \\ 0, & \text{otherwise.} \end{cases}$$

Note that this is just an exponential distribution $E(\lambda)$ with a location shift of $10kg/cm^2$. ∎

3.4 Some Other Important Functions

In this section we present four important functions associated with random variables: the survival, hazard, moment generating and characteristics functions.

3.4.1 Survival and Hazard Functions

Let X be a nonnegative random variable with pdf $f(x)$ and cdf $F(x)$. This happens, for example, when the random variable X is the lifetime of an object (e.g., a person or a machine). We have the following definitions:

Definition 3.2 (Survival function) *The function*

$$S(x) = \Pr(X > x) = 1 - F(x) \tag{3.84}$$

is called the survival function.

The function $S(x)$ is called a survival function because it gives the probability that the object will survive beyond time x.

Definition 3.3 (Hazard function) *The hazard function (hazard rate or mortality function) is defined as*

$$H(x) = \frac{f(x)}{\Pr(X > x)} = \frac{f(x)}{1 - F(x)}. \tag{3.85}$$

Assume that X is the lifetime of an element. Then, the hazard function can be interpreted as the probability, per unit time, that the item will fail just after time x given that the item has survived up to time x:

$$\begin{aligned} H(x) &= \lim_{\varepsilon \to 0} \frac{\Pr(x < X \leq x + \varepsilon | X > x)}{\varepsilon} \\ &= \frac{\lim_{\varepsilon \to 0} \dfrac{F(x + \varepsilon) - F(x)}{\varepsilon}}{1 - F(x)} \\ &= \frac{\dfrac{d}{dx} F(x)}{1 - F(x)} = \frac{f(x)}{1 - F(x)}. \end{aligned}$$

In other words, the hazard function can be interpreted as the probability of instantaneous failure given that the item has survived up to time x.

From (3.84) and (3.85), it can be seen that

$$f(x) = H(x)S(x), \tag{3.86}$$

that is, the pdf is the product of the hazard and survival functions. There is also a one-to-one correspondence between the cdf, $F(x)$, and the hazard function, $H(x)$. To see this, note that

$$\frac{d}{dx} \log(1 - F(x)) = -H(x),$$

and integrating from 0 to x we have,

$$\log(1 - F(x)) = -\int_0^x H(t)dt,$$

and

$$F(x) = 1 - \exp\left\{-\int_0^x H(t)dt\right\}. \tag{3.87}$$

Consequently, there is a one-to-one correspondence between $F(x)$ and $H(x)$, and so one can be obtained from the other.

Note also that a comparison between (3.84) and (3.87) suggests the following relationship between the survival and hazard functions:

$$S(x) = \exp\left\{-\int_0^x H(t)dt\right\}. \tag{3.88}$$

Example 3.14 (Weibull). Consider a Weibull random variable with cdf

$$F(x) = 1 - \exp(-x^\beta), \quad \beta > 0, x > 0.$$

The corresponding pdf is given by

$$f(x) = \beta x^{\beta-1}\exp(-x^\beta).$$

Then, the hazard function is

$$H(x) = \frac{f(x)}{1-F(x)} = \frac{\beta x^{\beta-1}\exp(-x^\beta)}{\exp(-x^\beta)} = \beta x^{\beta-1}.$$

Note that $H(x)$ is increasing, constant, or decreasing according to $\beta > 1$, $\beta = 1$, or $\beta < 1$, respectively. ■

3.4.2 Moment Generating Function

As we have already seen in this chapter and in Chapter 2, every random variable (discrete or continuous) has a cumulative distribution function $F(x)$ and an associated pmf ($P(x)$ in the discrete case) or pdf ($f(x)$ in the continuous case). In addition to these functions, random variable have two other important functions; the *moment generating function* (MGF) and the *characteristic function* (CF). These are discussed below.

Definition 3.4 (Moment generating function) *Let X be a unidimensional random variable with distribution function $F(x)$. The MGF of X, denoted by $M(t)$, is defined for the discrete case as*

$$M(t) = E\left(e^{tX}\right) = \sum_{x \in S(X)} e^{tx}P(x), \tag{3.89}$$

where $P(x)$ is the pmf of X and $S(X)$ is the support of $P(x)$. For the continuous case, the MGF is defined as

$$M(t) = E\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx}f(x)dx, \tag{3.90}$$

where $f(x)$ is the pdf of X and $M(t)$ is a function of t.

For simplicity of notation we shall use $M(t)$ instead of $M(t)$, unless otherwise needed. The function $M(t)$ is called the moment generating function because it generates all the moments of the random variable X. Namely, the kth order moment of the random variable X is given by

$$E(X^k) = \frac{d^{(k)}}{dt} M(t)\bigg|_{t=0}. \tag{3.91}$$

In other words, the kth order moment of the random variable X is the kth derivative of the MGF with respect to t, evaluated at $t = 0$.

Example 3.15 (The MGF of a binomial random variable). The MGF of the binomial random variable $B(n, p)$ is

$$M(t) \quad = \quad \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x \, q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(p \, e^t\right)^x \, q^{n-x} = \left(p \, e^t + q\right)^n.$$

The first two derivatives of $M(t)$ are

$$\frac{d}{dt} M(t) = n \left(pe^t + (1-p)\right)^{n-1} pe^t$$

and

$$\frac{d^{(2)}}{dt} M(t) = n \left(pe^t + (1-p)\right)^{n-1} pe^t + n(n-1) \left(pe^t + (1-p)\right)^{n-2} p^2 e^{2t},$$

respectively. Then, substituting zero for t in the above two equations, we obtain the first two moments, that is,

$$E(X) = np \tag{3.92}$$

and

$$E(X^2) = np + n(n-1)p^2, \tag{3.93}$$

from which the variance is obtained as

$$\sigma^2 = E(X^2) - [E(X)]^2 = np(1-p). \tag{3.94}$$

The MGF for some other random variables are given in Table 3.4. ∎

3.4.3 Characteristic Function

The MGF does not always exist, which means that not every random variable has a MGF. A function that always exists is the characteristic function, which is defined next.

Table 3.4: Moment Generating Functions of Some Common Random Variables.

Discrete Distribution		Continuous Distribution	
Name	$M(t)$	Name	$M(t)$
Dirac	e^t	Uniform	$\dfrac{e^{tb}-e^{ta}}{t(b-a)}$
Bernoulli	$pe^t+(1-p)$	Exponential	$\left(1-\dfrac{t}{\lambda}\right)^{-1}$
Binomial	$\left(pe^t+(1-p)\right)^n$	Gamma	$\left(1-\dfrac{t}{\lambda}\right)^{-k}$
Geometric	$\dfrac{pe^t}{1-(1-p)e^t}$	Log-gamma	$e^{t\alpha}\Gamma(1+t\beta)\Gamma(1-t\beta)$
Negative binomial	$\left[\dfrac{pe^t}{1-(1-p)e^t}\right]^r$	Normal	$\exp\left\{t\mu-\dfrac{\sigma^2t^2}{2}\right\}$
Poisson	$\exp\left\{\lambda(e^t-1)\right\}$	χ^2	$\dfrac{\exp\{t\lambda/(1-2t)\}}{(1-2t)^{n/2}}$
Multinomial	$\left[\sum_{j=1}^{k} p_j e^{t_j}\right]^n$	Logistic	$e^{t\alpha}\Gamma(1+\beta t)\Gamma(1-\beta t)$

Definition 3.5 (Characteristic function) *Let X be a unidimensional random variable with distribution function $F(x)$; its characteristic function, denoted by $\psi_X(t)$, is defined for the discrete case as*

$$\psi_X(t) = \sum_{x\in S(X)} e^{itx}P(x), \tag{3.95}$$

where $P(x)$ is the pmf of X and $S(X)$ is the support of $P(x)$. For the continuous case, it is defined as

$$\psi_X(t) = \int_{-\infty}^{\infty} e^{itx}f(x)dx, \tag{3.96}$$

where $f(x)$ is the pdf of X. Note that, in both cases, $\psi_X(t)$ is a complex function.

For simplicity of notation we shall use $\psi(t)$ instead of $\psi_X(t)$, unless otherwise needed. Before we discuss the importance of the characteristic function, let us derive it for some random variables.

Example 3.16 (Characteristic function of a discrete uniform random variable). The characteristic function of the discrete uniform random variable (see Section 2.2.1, which has the pmf

$$P(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n,$$

is

$$\psi(t) = \sum_x e^{i\,tx} P(x) = \sum_x e^{itx} \frac{1}{n} = (e^{it} + e^{2it} + \ldots + e^{nit})/n.$$

■

Example 3.17 (Characteristic function of a continuous uniform random variable). The characteristic function of the continuous uniform random variable, $U(0, a)$, with pdf function

$$f(x) = \frac{1}{a}, \quad 0 \le x \le a,$$

is

$$\psi(t) = \int_0^a e^{itx}\, \frac{1}{a}\, dx = \frac{1}{a}\, \frac{e^{itx}}{it}\bigg|_0^a = \frac{e^{ita} - 1}{ita}.$$

■

Example 3.18 (Characteristic function of a binomial random variable). The characteristic function of the binomial random variable $B(n, p)$, with pmf

$$P(x) = \binom{n}{x} p^x\, q^{n-x}, \quad x = 0, 1, \ldots, n$$

is

$$\psi(t) = \sum_{x=0}^{n} e^{itx} \binom{n}{x} p^x\, q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(p\, e^{it}\right)^x\, q^{n-x} = \left(p\, e^{it} + q\right)^n.$$

■

Table 3.5 gives the characteristic functions of the most common distributions. The characteristic function makes the calculations of moments easy. It also helps sometimes in the identification of distributions of sums of independent random variables.

The most important properties (applications) of the characteristic function are:

1. The characteristic function always exists.

2. $\psi(0) = 1$.

3. $-1 \le \psi(t) \le 1$.

Table 3.5: Characteristic Functions of Some Common Random Variables.

Discrete Distribution		Continuous Distribution	
Name	$\psi(t)$	Name	$\psi(t)$
Dirac	e^{it}	Uniform	$\dfrac{e^{itb}-e^{ita}}{it(b-a)}$
Bernoulli	$pe^{it}+(1-p)$	Exponential	$\left(1-\dfrac{it}{\lambda}\right)^{-1}$
Binomial	$\left(pe^{it}+(1-p)\right)^n$	Gamma	$\left(1-\dfrac{it}{\lambda}\right)^{-k}$
Geometric	$\dfrac{pe^{it}}{1-(1-p)e^{it}}$	Log-gamma	$e^{it\alpha}\Gamma(1+it\beta)\Gamma(1-it\beta)$
Negative binomial	$\left[\dfrac{pe^{it}}{1-(1-p)e^{it}}\right]^r$	Normal	$\exp\left\{it\mu-\dfrac{\sigma^2t^2}{2}\right\}$
Poisson	$\exp\left\{\lambda(e^{it}-1)\right\}$	χ^2	$\dfrac{\exp\{it\lambda/(1-2it)\}}{(1-2it)^{n/2}}$
Multinomial	$\left[\sum_{j=1}^{k} p_j e^{it_j}\right]^n$	Logistic	$e^{it\alpha}\Gamma(1+i\beta t)\Gamma(1-i\beta t)$

4. If $Z = aX+b$, where X is a random variable and a and b are real constants, we have

$$\psi_Z(t) = e^{itb}\psi_X(at), \tag{3.97}$$

where $\psi_Z(t)$ and $\psi_X(t)$ are the characteristic functions of Z and X, respectively.

5. The characteristic function of the sum of two independent random variables is the product of their characteristic functions:

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t). \tag{3.98}$$

6. Suppose that $X_1, \ldots, X_n$ is a set of n independent random variables with characteristic functions $\psi_{X_1}(t), \ldots, \psi_{X_n}(t)$, respectively. Let $C = \sum_{i=1}^{n} a_i X_i$ be a linear combination of the random variables. Then, the characteristic function of C is given by

$$\psi_C(t) = \prod_{i=1}^{n} \psi_{X_i}(a_i t). \tag{3.99}$$

7. The characteristic function of the random variable R, which is the sum of a random number N of identically and independently distributed random variables $X_1, \ldots, X_N$ is given by

$$\psi_R(t) = \psi_N\left(\frac{\log \psi_X(t)}{i}\right), \tag{3.100}$$

where $\psi_X(t)$, $\psi_R(t)$ and $\psi_N(t)$ are the characteristic functions of X_i, R, and N, respectively.

Example 3.19 (Sum of normal random variables). Let $Z_1, \ldots, Z_n$ be independent normal random variables $N(0,1)$ with characteristic function

$$\psi_Z(t) = e^{-t^2/2}.$$

Also, let $S = Z_1 + \ldots + Z_n$ and $M = S/\sqrt{n}$. Then, according to Property 5, the characteristic function of S is

$$\psi_S(t) = [\psi_Z(t)]^n,$$

and, according to Property 4, the characteristic function of M is

$$\begin{aligned}\psi_M(t) &= \psi_S\left(t/\sqrt{n}\right) = \left[\psi_Z\left(t/\sqrt{n}\right)\right]^n \\ &= \exp\left(-\frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 n\right) = e^{-t^2/2},\end{aligned} \tag{3.101}$$

which shows that $M = (Z_1 + \ldots + Z_n)/\sqrt{n}$ has the same characteristic function as Z_i. ■

Example 3.20 (Stability of the Gamma family with respect to scale changes). Let X be a Gamma $G(\theta, \lambda)$ random variable with characteristic function

$$\psi_X(t) = \left(1 - \frac{it}{\theta}\right)^{-\lambda}$$

and $Y = cX$. Then, by Property 4 we have

$$\begin{aligned}\psi_Y(t) &= \psi_X(ct) = \psi_{G(\theta,\lambda)}(ct) = \left(1 - \frac{ict}{\theta}\right)^{-\lambda} \\ &= \left(1 - \frac{it}{\theta/c}\right)^{-\lambda} = \psi_{G(\theta/c,\lambda)}(t),\end{aligned}$$

which shows that the random variable Y is $G(\theta/c, \lambda)$, that is, the Gamma family is stable with respect to scale changes. ■

Example 3.21 (Stability of the uniform family). Let X be uniform $U(a,b)$, with characteristic function

$$\psi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

Then, by Property 4 the characteristic function of the random variable $Y = cX + d$ is

$$\begin{aligned}\psi_Y(t) &= e^{itd}\,\psi_X(ct) = e^{itd}\,\frac{e^{itbc}-e^{itac}}{ict(b-a)} \\ &= \frac{e^{it(cb+d)}-e^{it(ca+d)}}{it(cb-ca)} = \frac{e^{it(cb+d)}-e^{it(ca+d)}}{it[(cb+d)-(ca+d)]},\end{aligned}$$

which shows that $Y \sim U(ca+d, cb+d)$. ∎

As indicated, the kth order moment $E(X^k)$ can be easily calculated from the characteristic function as

$$E(X^k) = \frac{\psi^{(k)}(0)}{i^k}. \tag{3.102}$$

Example 3.22 (Moments of the Bernoulli random variable). Since the characteristic function of the Bernoulli random variable is

$$\psi(t) = pe^{it} + q,$$

using (3.102) we have

$$E(X^k) = \frac{\psi^{(k)}(0)}{i^k} = \frac{pi^k}{i^k} = p.$$

This shows that all moments are equal to p. ∎

Example 3.23 (Moments of the Gamma random variable). Since the characteristic function of the Gamma $G(\theta, \lambda)$ random variable is

$$\psi(t) = \left(1 - \frac{it}{\theta}\right)^{-\lambda},$$

its moments with respect to the origin are

$$E(X^k) = \frac{\psi^{(k)}(0)}{i^k} = \frac{\lambda\,(\lambda+1)\ldots(\lambda+k-1)}{\theta^k}.$$

∎

3.5 Multivariate Continuous Random Variables

In Section 3.2 we dealt with random variables, one at a time. In some practical situations, we may need to deal with several random variables simultaneously. In this section, we describe models that deal with multidimensional random variables.

3.5.1 Joint Probability Density Function

Let $X = \{X_1, \ldots, X_n\}$ be n-dimensional continuous random variable, with support $S(X_i), i = \{1, 2, \ldots, n\}$. The pdf of this multivariate random variable is given by $f(x_1, \ldots, x_n)$. This is called the *joint* pdf. The joint pdf has n arguments, $x_1, \ldots, x_n$, one for each of the variables.

The pdf satisfies the following properties:

$$f(x_1, \ldots, x_n) \geq 0,$$

and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1.$$

Note that for any continuous n-dimensional random variable, we have $\Pr(X_1 = x_1, \ldots, X_n = x_n) = 0$.

3.5.2 Joint Cumulative Distribution Function

The joint cdf is defined in an analogous manner to that used for the univariate case:

$$F(\mathbf{x}) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n) dt_1 \ldots dt_n.$$

When $n = 2$, we have a *bivariate* random variable.

Example 3.24 (Bivariate cumulative distribution function). The cdf of a bivariate random variable (X_1, X_2) is

$$F(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2.$$

The relationship between the pdf and cdf is

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \, \partial x_2}.$$

Among other properties of a two-dimensional cdf, we mention the following:

1. $F(\infty, \infty) = 1$.
2. $F(-\infty, x_2) = F(x_1, -\infty) = 0$.
3. $F(x_1 + a_1, x_2 + a_2) \geq F(x_1, x_2)$, where $a_1, a_2 \geq 0$.
4. $P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$. This formula is illustrated in Figure 3.12.

■

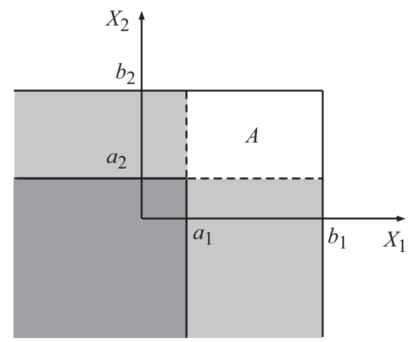

Figure 3.12: An illustration of computing the probability of a rectangle in terms of the joint cdf associated with its vertices.

3.5.3 Marginal Probability Density Functions

From the joint pdf we can obtain *marginal* probability density functions, one marginal for each variable. The marginal pdf of one variable is obtained by integrating the joint pdf over all other variables. Thus, for example, the marginal pdf of X_1, is

$$f(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_2 \ldots dx_n.$$

A multidimensional random variable is said to be continuous if its marginals are continuous.

The probabilities of any set B of possible values of a multidimensional continuous variable can be calculated if the pdf, written as $f(\mathbf{x})$ or $f(x_1, \ldots, x_n)$, is known just by integrating it in the set B. For example, the probability that X_i belongs to a given region, say, $a_i < X_i \leq b_i$ for all i, is the integral

$$\Pr(a_1 < X_1 \leq b_1, \ldots, a_n < X_n \leq b_n) = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

3.5.4 Conditional Probability Density Functions

We define the conditional pdf for the case of two-dimensional random variables. The extension to the n-dimensional case is straightforward. For simplicity of notation we use (X, Y) instead of (X_1, X_2). Let then (X, Y) be a two-dimensional random variable. The random variable Y given $X = x$ is denoted by $(Y|X = x)$. The corresponding probability density and distribution functions are called the *conditional* pdf and cdf, respectively. The following expressions give the conditional pdf for the random variables $(Y|X = x)$ and $(X|Y = y)$:

$$f_{(Y|X=x)}(y|x) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}, \quad f_X(x) > 0,$$

and

$$f_{(X|Y=y)}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}, \quad f_Y(y) > 0.$$

It may also be of interest to compute the pdf conditioned on events different from $Y = y$. For example, for the event $Y \leq y$, we get the conditional cdf:

$$F_{(X|Y \leq y)}(x,y) = \Pr(\, X \leq x | Y \leq y) = \frac{\Pr(X \leq x, Y \leq y)}{\Pr(Y \leq y)} = \frac{F_{(X,Y)}(x,y)}{F_Y(y)}.$$

The corresponding pdf is given by

$$f_{(X|Y \leq y)}(x,y) = \frac{\partial}{\partial x}\, F_{(X|Y \leq y)}(x,y) = \frac{\partial}{\partial x}\, \frac{F_{(X,Y)}(x,y)}{F_Y(y)}.$$

Two random variables X and X are said to be *independent* if

$$f_{(Y|X=x)}(y|x) = f_Y(y), \tag{3.103}$$

or

$$f_{(X|Y=y)}(x) = f_X(x), \tag{3.104}$$

otherwise, they are said to be *dependent*. This means that X and Y are independent if the conditional pdf is equal to the marginal pdf. Note that (3.103) and (3.104) are equivalent to

$$f_{(X,Y)}(x,y) = f_X(x) f_Y(y), \tag{3.105}$$

that is, if two variables are independent, then their joint pdf is equal to the product of their marginals. This is also true for $n > 2$ random variables. That is, if $X_1, \ldots, X_n$ are independent random variables, then

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n). \tag{3.106}$$

3.5.5 Covariance and Correlation

Using the marginal pdfs, $f_{X_1}(x_1), \ldots, f_{X_n}(x_n)$, we can compute the means, $\mu_1, \ldots, \mu_n$, and variances, $\sigma_1^2, \ldots, \sigma_n^2$ using (3.9) and (3.11), respectively. We can also compute the *covariance* between every pair of variables. The covariance between X_i and X_j, denoted by σ_{ij}, is defined as

$$\sigma_{ij} = \int_{x_i \in S(X_i)} \int_{x_j \in S(X_j)} (x_i - \mu_i)(x_j - \mu_j) f(x_i, x_j) dx_j dx_i, \tag{3.107}$$

where $f(x_i, x_j)$ is the joint pdf of X_i and X_j, which is obtained by integrating the joint pdf over all variables other than X_i and X_j. As in the discrete case, the *correlation coefficient* is

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}. \tag{3.108}$$

For convenience, we usually arrange the means, variances, covariances, and correlations in matrices. The means are arranged in a column vector and the variances and covariances are arranged in a matrix as follows:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}, \tag{3.109}$$

where we use σ_{ii} instead of σ_i^2, for convenience. The vector $\boldsymbol{\mu}$ is known as the mean vector and the matrix $\boldsymbol{\Sigma}$ is known as the variance-covariance matrix. Similarly, the correlation coefficients can be arranged in the following matrix:

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{pmatrix}, \tag{3.110}$$

which is known as the *correlation* matrix. Note that both $\boldsymbol{\Sigma}$ and $\boldsymbol{\rho}$ are symmetric matrices. The relationship between them is

$$\boldsymbol{\rho} = \mathbf{D}\boldsymbol{\Sigma}\,\mathbf{D}, \tag{3.111}$$

where $\mathbf{D}$ is a diagonal matrix whose ith diagonal element is $1/\sqrt{\sigma_{ii}}$.

3.5.6 The Autocorrelation Function

In this section we introduce the concept of autocorrelation function, that will be used in dependent models to be described later in Chapter 8.

Definition 3.6 (Autocorrelation function) *Let $X_1, \dots$ be a set of random variables with the same mean and variance, and given by $E(X_i) = \mu$ and $Var(X_i) = \sigma_X^2$. The covariance between the random variables X_i and X_{i+k} separated by k intervals (of time), which under the stationarity assumption must be the same for all i, is called the autocovariance at lag k and is defined by*

$$\gamma_k = cov(X_i, X_{i+k}) = E[(X_i - \mu)(X_{i+k} - \mu)]. \tag{3.112}$$

The autocorrelation function at lag k is

$$\rho_k = \frac{E[(X_i - \mu)(X_{i+k} - \mu)]}{\sqrt{E[(X_i - \mu)^2]E[(X_{i+k} - \mu)^2]}} = \frac{\gamma_k}{\sigma_X^2}. \tag{3.113}$$

3.5.7 Bivariate Survival and Hazard Functions

Let (X, Y) be a bivariate random variable, where X and Y are nonnegative lifetime random variables, and let $F(x, y)$ be an absolutely continuous bivariate distribution function with density function $f(x, y)$.

Definition 3.7 *The bivariate survival function is given by*

$$S(x,y) = \Pr(X > x, Y > y). \tag{3.114}$$

Thus, the bivariate survival function gives the probability that the object X will survive beyond time x and the object will survive beyond time y.

Definition 3.8 (Bivariate hazard function) *The bivariate hazard function or bivariate failure rate is given by*

$$H(x,y) = \frac{f(x,y)}{\Pr(X > x, Y > y)}. \tag{3.115}$$

The above definition is due to Basu (1971). From (3.114) and (3.115), we see that

$$H(x,y) = \frac{f(x,y)}{S(x,y)}. \tag{3.116}$$

Note that if X and Y are independent random variables, then we have

$$H(x,y) = H_X(x)H_Y(y),$$

where $H_X(x)$ and $H_Y(y)$ are the corresponding univariate hazard functions. Similar to the univariate case, $H(x,y)$ can be interpreted as the probability of failure of both items in the intervals of time $[x, x+\varepsilon_1)$ and $[y, y+\varepsilon_2)$, on condition that they did not fail before time x and Y, respectively:

$$\begin{aligned} H(x,y) &= \lim_{x \to \varepsilon_1, y \to \varepsilon_2} \frac{\Pr(x < X \le x+\varepsilon_1, y < Y \le y+\varepsilon_2 | X > x, Y > y)}{\varepsilon_1 \varepsilon_2} \\ &= \frac{f(x,y)}{\Pr(X > x, Y > y)}. \end{aligned}$$

Unlike in the univariate case, the bivariate hazard function does not define $F(x,y)$, and other types of hazard functions may be taken into consideration.

Example 3.25 (Bivariate survival function). Consider a bivariate random variable with bivariate survival function

$$S(x,y) = \Pr(X > x, Y > y) = (1+x+y)^{-\alpha}, \quad \alpha > 0,\ x,y > 0,$$

where $\alpha > 0$. The joint pdf is

$$f(x,y) = \frac{\partial^2 S(x,y)}{\partial x \partial y} = \alpha(\alpha+1)\,(1+x+y)^{-(\alpha+2)},$$

and the bivariate hazard function is

$$H(x,y) = \frac{f(x,y)}{\Pr(X > x, Y > y)} = \alpha(\alpha+1)\,(1+x+y)^{-2}.$$

∎

3.5.8 Relationship Between Bivariate CDF and Survival Functions

Let (X, Y) be a bivariate random variable with joint cdf

$$F(x, y) = \Pr(X \leq x, Y \leq y),$$

and joint survival function

$$S(x, y) = \Pr(X > x, Y > y).$$

The relationship between $S(x, y)$ and $F(x, y)$ is given by (see Fig. 3.12)

$$S(x, y) = 1 + F(x, y) - F_X(x) - F_Y(y), \tag{3.117}$$

where $F_X(x)$ and $F_Y(y)$ are the cdf of the marginals.

3.5.9 Joint Characteristic Function

The characteristic function can be generalized to n dimensions as follows.

Definition 3.9 (n-dimensional characteristic function) *Let $X = (X_1, \ldots, X_n)$ be a n-dimensional random variable. Its characteristic function is defined as*

$$\psi_{\mathbf{X}}(\mathbf{t}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i(t_1 x_1 + \ldots + t_n x_n)} dF_{\mathbf{X}}(x_1, \ldots, x_n),$$

where $F_{\mathbf{X}}(x_1, \ldots, x_n)$ is the cdf of $\mathbf{X}$ and $\mathbf{t} = (t_1, \ldots, t_n)$.

3.6 Common Continuous Multivariate Models

3.6.1 Bivariate Logistic Distribution

The bivariate logistic distribution with joint cdf,

$$F(x, y) = \frac{1}{1 + e^{-(x-\lambda)/\sigma} + e^{-(y-\delta)/\tau}}, \quad -\infty < x, y < \infty, \tag{3.118}$$

where $-\infty < \lambda, \delta < \infty$ and $\sigma, \tau > 0$, was introduced by Gumbel (1961). The corresponding joint density function is

$$f(x, y) = \frac{2e^{-(x-\lambda)/\sigma-(y-\delta)/\tau}}{\sigma\tau \left[1 + e^{-(x-\lambda)/\sigma} + e^{-(y-\delta)/\tau}\right]^3}, \quad -\infty < x, y < \infty. \tag{3.119}$$

From Equation (3.118), by letting x or y go to ∞, we obtain the marginal cumulative distribution functions of X and Y:

$$F_X(x) = \frac{1}{1 + e^{-(x-\lambda)/\sigma}} \quad \text{and} \quad F_Y(y) = \frac{1}{1 + e^{-(y-\delta)/\tau}}, \tag{3.120}$$

which are univariate logistic distributions. The conditional density function of $X|Y$ is

$$f_{X|Y}(x|y) = \frac{2e^{-(x-\lambda)/\sigma}(1+e^{-(y-\delta)/\tau})^2}{\sigma\left[1+e^{-(x-\lambda)/\sigma}+e^{-(y-\delta)/\tau}\right]^3}, \tag{3.121}$$

and the conditional mean of X given $Y = y$ is

$$E(X|Y=y) = \lambda + \sigma - \sigma \log(1+e^{-(y-\delta)/\tau}). \tag{3.122}$$

3.6.2 Multinormal Distribution

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an n-dimensional normal random variable, which is denoted by $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix, respectively. The pdf of $\mathbf{X}$ is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left[-(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})/2\right],$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\det(\boldsymbol{\Sigma})$ is the determinant of $\boldsymbol{\Sigma}$. The following theorem gives the conditional mean and variance-covariance matrix of any conditional variable, which is normal.

Theorem 3.2 (Conditional mean and covariance matrix) *Let Y and Z be two sets of random variables having a multivariate normal distribution with mean vector and covariance matrix given in partitioned forms by*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_Y \\ \boldsymbol{\mu}_Z \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YZ} \\ \boldsymbol{\Sigma}_{ZY} & \boldsymbol{\Sigma}_{ZZ} \end{pmatrix},$$

where $\boldsymbol{\mu}_Y$ and $\boldsymbol{\Sigma}_{YY}$ are the mean vector and covariance matrix of Y; $\boldsymbol{\mu}_Z$ and $\boldsymbol{\Sigma}_{ZZ}$ are the mean vector and covariance matrix of Z; and $\boldsymbol{\Sigma}_{YZ}$ is the covariance of Y and Z. Then the conditional pdf of Y given $Z = z$ is multivariate normal with mean vector $\boldsymbol{\mu}_{Y|Z=z}$ and covariance matrix $\boldsymbol{\Sigma}_{Y|Z=z}$, where

$$\begin{aligned} \boldsymbol{\mu}_{Y|Z=z} &= \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YZ}\, \boldsymbol{\Sigma}_{ZZ}^{-1}\, (z - \boldsymbol{\mu}_Z), \\ \boldsymbol{\Sigma}_{Y|Z=z} &= \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YZ}\, \boldsymbol{\Sigma}_{ZZ}^{-1}\, \boldsymbol{\Sigma}_{ZY}. \end{aligned} \tag{3.123}$$

For other properties of multivariate normal random variables, see any multivariate analysis book such as Rencher (2002) or the multivariate distribution theory book by Kotz et al. (2000).

3.6.3 Marshall-Olkin Distribution

Due to its importance, we include here the Marshall-Olkin distribution (see Marshall and Olkin (1967)), which has several interesting physical interpretations. One such interpretation is as follows. Suppose we have a system with two components in series. Both components are subject to a Poissonian processes of fatal shocks, such that if a component is affected by a shock it fails.

Component 1 is subject to a Poissonian process with intensity λ_1, Component 2 is subject to a Poissonian process with intensity λ_2, and both components are subject to a Poissonian process with intensity λ_{12}. Let $N_1(x;\lambda_1)$, $N_2(y;\lambda_2)$, and $N_{12}(\max(x,y);\lambda_{12})$, be the number of shocks associated with first, second, and third Poissonian processes during a period of duration x, y, and $\max(x,y)$, respectively. Then, $N_1(x;\lambda_1)$, $N_2(y;\lambda_2)$, and $N_{12}(\max(x,y);\lambda_{12})$ are Poisson random variables with means $x\lambda_1$, $y\lambda_2$, and $\max(x,y)\lambda_{12}$, respectively. Thus, it follows from the pdf of the Poisson random variable in (2.28) that the bivariate survival function in (3.114) in this case becomes

$$\begin{aligned} S(x,y) &= \Pr(X>x, Y>y) \\ &= \Pr\left[N_1(x;\lambda_1)=0, N_2(y;\lambda_2)=0, N_{12}(\max(x,y);\lambda_{12})=0\right] \\ &= \exp\left[-\lambda_1 x - \lambda_2 y - \lambda_{12}\max(x,y)\right]. \end{aligned}$$

This model has another interpretation using nonfatal shocks. Consider the same model as before, but now the shocks are not fatal. Once a shock coming from the Poisson process with intensity λ_1 has occurred, there is a probability p_1 of failure of Component 1, once a shock coming from the Poisson process with intensity λ_2 has occurred, there is a probability p_2 of failure of Component 2. Finally, once a shock coming from the Poisson process with intensity λ_{12} has occurred, there are probabilities p_{00}, p_{01}, p_{10}, and p_{11} of failure of both components, only Component 1, only Component 2, and no failure, respectively. In this case we have

$$\begin{aligned} S(x,y) &= \Pr\left(X>x, Y>y\right) \\ &= \exp\left[-\delta_1 x - \delta_2 y - \delta_{12}\max(x,y)\right], \end{aligned} \tag{3.124}$$

where

$$\delta_1 = \lambda_1 p_1 + \lambda_{12} p_{01}, \quad \delta_2 = \lambda_2 p_2 + \lambda_{12} p_{10}, \quad \text{and} \quad \delta_{12} = \lambda_{12} p_{00}. \tag{3.125}$$

The following is a straightforward generalization of this model to n dimensions:

$$\begin{aligned} S(x_1,\ldots,x_n) = \exp\Bigg[&-\sum_{i=1}^{n} \lambda_i x_i - \sum_{i<j} \lambda_{ij}\max(x_i,x_j) \\ &- \sum_{i<j<k} \lambda_{ijk}\max(x_i,x_j,x_k) - \ldots - \lambda_{12\ldots n}\max(x_1,\ldots,x_n)\Bigg]. \end{aligned}$$

3.6.4 Freund's Bivariate Exponential Distribution

Freund (1961) constructed an alternate bivariate exponential model in the following manner. Suppose a system has two components (C_1 and C_2) with their lifetimes X_1 and X_2 having exponential densities

$$f_{X_i}(x) = \theta_i \ \exp\{-\theta_i \ x\}, \quad x \geq 0, \ \theta_i > 0 \ (i=1,2). \tag{3.126}$$

The dependence between X_1 and X_2 is introduced by the assumption that when Component C_i (with lifetime X_i) fails, the parameter for X_{3-i} changes from θ_{3-i}

to θ'_{3-i} (for $i = 1, 2$). In this set-up, the joint density function of X_1 and X_2 is

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \theta_1\ \theta'_2\ \exp\{-\theta'_2\ x_2 - \gamma_2\ x_1\} \quad \text{for} \quad 0 \le x_1 < x_2 \\ &= \theta'_1\ \theta_2\ \exp\{-\theta'_1\ x_1 - \gamma_1\ x_2\} \quad \text{for} \quad 0 \le x_2 < x_1, \end{aligned} \tag{3.127}$$

where $\gamma_i = \theta_1 + \theta_2 - \theta'_i$ $(i = 1, 2)$. The corresponding joint survival function is

$$\begin{aligned} S_{X_1,X_2}(x_1,x_2) &= \frac{1}{\gamma_2}\left\{\theta_1\ e^{-\theta'_2 x_2 - \gamma_2 x_1} + (\theta_2 - \theta'_2)\ e^{-(\theta_1+\theta_2)x_2}\right\} \quad 0 \le x_1 < x_2 \\ &= \frac{1}{\gamma_1}\left\{\theta_2\ e^{-\theta'_1 x_1 - \gamma_1 x_2} + (\theta_1 - \theta'_1)\ e^{-(\theta_1+\theta_2)x_1}\right\} \quad 0 \le x_2 < x_1. \end{aligned} \tag{3.128}$$

It should be noted that, under this model, the probability that Component C_i is the first to fail is $\theta_i/(\theta_1 + \theta_2), i = 1, 2$, and that the time to first failure is distributed as $\text{Exp}(\theta_1 + \theta_2)$. Further, the distribution of the time from first failure to failure of the other component is a mixture of $\text{Exp}(\theta'_1)$ and $\text{Exp}(\theta'_2)$ with proportions $\theta_2/(\theta_1 + \theta_2)$ and $\theta_1/(\theta_1 + \theta_2)$, respectively.

Block and Basu (1974) constructed a system of absolutely continuous bivariate exponential distributions by modifying the above presented Marshall-Olkin's bivariate exponential distributions (which do have singular part). This system is a reparameterization of Freund's bivariate exponential distribution in (3.128) with

$$\lambda_i = \theta_1 + \theta_2 - \theta'_{3-i}\ (i = 1, 2) \quad \text{and} \quad \lambda_{12} = \theta'_1 + \theta'_2 - \theta_1 - \theta_2.$$

For an elaborate discussion on various forms of bivariate and multivariate exponential distributions and their properties, one may refer to Chapter 47 of Kotz et al. (2000).

Exercises

3.1 Show that:

(a) The mean and variance of the uniform $U(\lambda, \beta)$ random variable are

$$\mu = \frac{\lambda + \beta}{2} \quad \text{and} \quad \sigma^2 = \frac{(\beta - \lambda)^2}{12}. \tag{3.129}$$

(b) The mean and variance of an exponential random variable with parameter λ are

$$\mu = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = \frac{1}{\lambda^2}. \tag{3.130}$$

3.2 The simple compression strength (measured in kg/cm^2) of a given concrete is a normal random variable:

(a) If the mean is 300 kg/cm^2 and the standard deviation is 40 kg/cm^2, determine the 15th percentile.

(b) If the mean is 200 kg/cm^2 and the standard deviation is 30 kg/cm^2, give the percentile associated with a strength of 250 kg/cm^2.

(c) If the mean is 300 kg/cm^2, obtain the standard deviation if the 80th percentile is 400 kg/cm^2.

(d) If an engineer states that 400 kg/cm^2 is the 20th percentile in the previous case, is he right?

3.3 The occurrence of earthquakes of intensity above five in a given region is Poissonian with mean rate 0.5 earthquakes/year.

(a) Determine the pdf of the time between consecutive earthquakes.

(b) If an engineering work fails after five earthquakes of such an intensity, obtain the pdf of the lifetime of such a work in years.

(c) Obtain the pmf of the number of earthquakes (of intensity five or larger) that occur in that region during a period of 10 years.

3.4 The arrivals of cars to a gas station follow a Poisson law of mean rate five cars per hour. Determine:

(a) The probability of five arrivals between 17.00 and 17.30.

(b) The pdf of the time up to the first arrival.

(c) The pdf of the time until the arrival of the 5th car.

3.5 If the height T, of an asphaltic layer is normal with mean 6 cm and standard deviation 0.5 cm, determine:

(a) The pdf value $f_T(5)$.

(b) The probability $\Pr(T \leq 2.5)$.

(c) The probability $\Pr(|T-6| < 2.5)$.

(d) The probability $\Pr(|T-6| < 2.5 | T \leq 5)$.

3.6 Show that, as $k \to \infty$, the log-gamma density in (3.32) tends to the Normal(μ, σ^2) density function.

3.7 Show that the cdf of Gamma in (3.26) has the following closed form for integer θ:

$$F(x) = 1 - \sum_{y=0}^{\theta-1} \frac{e^{-x} x^y}{y!}, \tag{3.131}$$

which shows that $F(x)$ is related to the Poisson probabilities.

3.8 Starting with the gamma pdf

$$\frac{1}{\Gamma(k)}\; e^{-y}\; y^{k-1}, \quad y > 0,\; k > 0,$$

show that the pdf in (3.29) is obtained by a logarithmic transformation.

3.9 A random variable X has the density in (3.29). Show that $Y = -X$ has the pdf

$$f(y) = \frac{1}{\Gamma(k)}\; e^{-ky}\; \exp\{-e^{-y}\}, \quad -\infty < y < \infty,\; k > 0,$$

which includes the Gumbel extreme value distribution in (3.65) as a special case. [Hint: The shape parameter $k = 1$.]

3.10 A random variable X has the pdf in (3.30). Show that the MGF of X is

$$E\left(e^{tX}\right) = e^{t\mu}\; \Gamma(k + t\sigma)/\Gamma(k). \tag{3.132}$$

3.11 Show that a generalized Pareto distribution truncated from the left is also a generalized Pareto distribution.

3.12 The grades obtained by students in a statistics course is a random variable with cdf

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \dfrac{x}{10}, & \text{if } 0 \le x < 10, \\ 1, & \text{if } x \ge 10. \end{cases}$$

(a) Obtain the cdf of the students with grade below 5.

(b) If the students receive at least one point just for participating in the evaluation, obtain the new cdf for this case.

3.13 Obtain the hazard function of the exponential distribution. Discuss the result.

3.14 A cumulative distribution function $F(x)$ is said to be an increasing (decreasing) hazard function (IHF and DHF, respectively) distribution if its hazard function is nondecreasing (nonincreasing) in x. Show that the following properties hold:

(a) If X_i, $i = 1, 2$, are IHF random variables with hazard functions given by $H_i(x)$, $i = 1, 2$, then the random variable $X = X_1 + X_2$ is also IHF with hazard function $H_X(x) \le \min\{H_1(x), H_2(x)\}$.

3.15 Let X be a random variable with survival function defined by (see Glen and Leemis (1997))

$$S(x) = \Pr(X > x) = \frac{\arctan[\alpha(\phi - x)] + (\pi/2)}{\arctan(\alpha\phi) + (\pi/2)}, \quad x \ge 0,$$

where $\alpha > 0$ and $-\infty < \phi < \infty$.

i. Show that $S(x)$ is a genuine survival function, that is, it satisfies the conditions $S(0) = 1$, $\lim_{x\to\infty} S(x) = 0$ and $S(x)$ is nonincreasing.

ii. Show that the hazard function

$$H(x) = \frac{\alpha}{\{\arctan[\alpha(\phi - x)] + (\pi/2)\}\{1 + \alpha^2(x - \phi)^2\}}, \quad x \geq 0,$$

has an upside-down bathtub form.

(b) A mixture of DHF distributions is also DHF. This property is not necessarily true for IHF distributions.

(c) Parallel and series systems of identical IHF units are IHF. For the series systems, the units do not have to have identical distributions.

3.16 Use MGFs in Table 3.4 to derive the mean and variance of the corresponding random variables in Table 3.3. [Hint: Find the first two derivatives of the MGF.]

3.17 Use CFs in Table 3.5 to derive the mean and variance of the corresponding random variables in Table 3.3. [Hint: Find the first two derivatives of the CF.]

3.18 Let X and Y be independent random variables.

(a) Show that the characteristic function of the random variable $Z = aX + bY$ is $\psi_Z(t) = \psi_X(at)\psi_Y(bt)$, where $\psi_Z(t)$ is the characteristic function of the random variable Z.

(b) Use this property to establish that a linear combination of normal random variables is normal.

3.19 Use the properties of the characteristic function to show that a linear combination of independent normal random variables is another normal random variable.

3.20 Let X_1 and X_2 be iid random variables with Gumbel distribution given by

$$F_i(x_i) = \exp[-\exp(-x_i)], \quad -\infty < x_i < \infty; \quad i = 1, 2.$$

Show that the random variable $X = X_1 - X_2$ has a logistic distribution with cdf

$$F_X(x) = \frac{1}{1 + e^{-x}}, \quad -\infty < x < \infty.$$

Hint: Use the characteristic function.

3.21 Consider a Gumbel bivariate exponential distribution defined by

$$\Pr(X > x, Y > y) = \exp(-\varepsilon_1 x - \varepsilon_2 y - \gamma xy), \quad x, y > 0,$$

where $\varepsilon_1, \varepsilon_2 > 0$ and $0 \leq \gamma \leq \varepsilon_1\varepsilon_2$.

(a) Obtain the marginal distributions and show that X and Y are independent if and only if $\gamma = 0$.

(b) Show that the bivariate hazard function is given by

$$H(x, y) = (\varepsilon_1 + \gamma x)(\varepsilon_2 + \gamma y) - \gamma.$$

3.22 The bivariate survival function of the Morgenstern distribution is

$$S(x, y) = \exp(-x - y)\left\{1 + \alpha\left[1 - \exp(-x)\right]\left[1 - \exp(-y)\right]\right\}.$$

Obtain:

(a) The cdf $F_{X,Y}(x, y)$.

(b) The marginal cdfs $F_X(x)$ and $F_Y(y)$.

(c) The conditional cdfs $F_{X|Y=y}(x|y)$ and $F_{Y|X=x}(y|x)$.

3.23 The bivariate survival function of the Gumbel type I distribution is

$$S(x, y) = \exp(-x - y + \theta xy).$$

Obtain:

(a) The cdf $F_{X,Y}(x, y)$.

(b) The marginal cdfs $F_X(x)$ and $F_Y(y)$.

(c) The conditional cdfs $F_{X|Y=y}(x|y)$ and $F_{Y|X=x}(y|x)$.

3.24 Show that the joint survival function corresponding to (3.127) is as given in (3.128).

3.25 Show that, when $\theta_1 + \theta_2 \neq \theta'_i$ $(i = 1, 2)$, the marginal density function of X_i $(i = 1, 2)$ is

$$f_{X_i}(x) = \frac{1}{\theta_1 + \theta_2 - \theta'_i}\left\{(\theta_i - \theta'_i)(\theta_1 + \theta_2)\, e^{-(\theta_1+\theta_2)x} + \theta'_i\theta_{3-i}e^{-\theta'_i x}\right\},\ x \geq 0,$$

which is indeed a mixture of two exponential distributions if $\theta_i > \theta'_i$.

3.26 For the bivariate exponential distribution in (3.127), show that the joint MGF of (X_1, X_2) is

$$E\left(e^{s_1X_1+s_2X_2}\right) = \frac{1}{\theta_1 + \theta_2 - s_1 - s_2}\left\{\frac{\theta_2}{1 - s_1/\theta'_1} + \frac{\theta_1}{1 - s_2/\theta'_2}\right\}.$$

Find $E(X_i)$, $\mathrm{Var}(X_i)$, $\mathrm{Cov}(X_1,\, X_2)$, and $\mathrm{Corr}(X_1,\, X_2)$.