3

Probability Models and Distribution Functions

3.1 Basic probability

3.1.1 Events and sample spaces: Formal presentation of random measurements

Experiments or trials of interest, are those which may yield different results with outcomes that are not known ahead of time with certainty. We have seen in the previous chapters a large number of examples in which outcomes of measurements vary. It is of interest to find, before conducting a particular experiment, what are the chances of obtaining results in a certain range. In order to provide a quantitative answer to such a question, we have to formalize the framework of the discussion so that no ambiguity is left.

When we say a "trial" or "experiment," in the general sense, we mean a well defined process of measuring certain characteristic(s), or variable(s). For example, if the experiment is to measure the compressive strength of concrete cubes, we must specify exactly how the concrete mixture was prepared, that is, proportions of cement, sand, aggregates and water in the batch. Length of mixing time, dimensions of mold, number of days during which the concrete has hardened. The temperature and humidity during preparation and storage of the concrete cubes, etc. All these factors influence the resulting compressive strength. Well-documented protocol of an experiment enables us to replicate it as many times as needed. In a well-controlled experiment we can assume that the variability in the measured variables is due to **randomness**. We can think of the random experimental results as sample values from a hypothetical population. The set of all possible sample values is called the **sample space**. In other words, the sample space is the set of all possible outcomes of a specified experiment. The outcomes do not have to be numerical. They could be names, categorical values, functions, or collection of items. The individual outcome of an experiment will be called an **elementary event** or a **sample point** (element). We provide a few examples.

Example 3.1. The experiment consists of choosing ten names (without replacement) from a list of 400 undergraduate students at a given university. The outcome of such an experiment is a list of ten names. The sample space is the collection of **all** possible such sublists that can be drawn from the original list of 400 students. ■

Example 3.2. The experiment is to produce twenty concrete cubes, under identical manufacturing conditions, and count the number of cubes with compressive strength above 200 [kg/cm^2]. The sample space is the set $S = \{0, 1, 2, \cdots, 20\}$. The elementary events, or sample points, are the elements of S. ■

Modern Industrial Statistics: with applications in R, MINITAB and JMP, Second Edition. Ron S. Kenett and Shelemyahu Zacks.

Companion website: www.wiley.com/go/modern_industrial_statistics

Example 3.3. The experiment is to choose a steel bar from a specific production process, and measure its weight. The sample space S is the interval (ω_0, ω_1) of possible weights. The weight of a particular bar is a sample point. ■

Thus, sample spaces could be finite sets of sample points, or countable or non-countable infinite sets.

Any subset of the sample space, S, is called an **event**. S itself is called the **sure event**. The empty set, $\emptyset$, is called the **null event**. We will denote events by the letters $A, B, C, \cdots$ or $E_1, E_2, \cdots$. All events under consideration are subsets of the same sample space S. Thus, events are sets of sample points.

For any event $A \subseteq S$, we denote by A^c the **complementary event**, that is, the set of all points of S which are not in A.

An event A is said to **imply** an event B, if all elements of A are elements of B. We denote this **inclusion relationship** by $A \subset B$. If $A \subset B$ **and** $B \subset A$, then the two events are **equivalent**, $A \equiv B$.

Example 3.4. The experiment is to select a sequence of 5 letters for transmission of a code in a money transfer operation. Let $A_1, A_2, \ldots, A_5$ denote the first, second, $\ldots$, fifth letter chosen. The sample space is the set of all possible sequences of five letters. Formally,

$$S = \{(A_1A_2A_3A_4A_5) : A_i \in \{a, b, c, \cdots, z\}, \quad i = 1, \cdots, 5\}$$

This is a finite sample space containing 26^5 possible sequences of 5 letters. Any such sequence is a sample point.

Let E be the event that all the 5 letters in the sequence are the same. Thus

$$E = \{aaaaa, bbbbb, \cdots, zzzzz\}.$$

This event contains 26 sample points. The complement of E, E^c, is the event that at least one letter in the sequence is different from the other ones. ■

3.1.2 Basic rules of operations with events: Unions, intersections

Given events $A, B, \cdots$ of a sample space S, we can generate new events, by the operations of union, intersection and complementation.

The **union** of two events A and B, denoted $A \cup B$, is an event having elements which belong **either** to A **or** to B.

The intersection of two events, $A \cap B$, is an event whose elements belong both to A **and** to B. By pairwise union or intersection we immediately extend the definition to finite number of events $A_1, A_2, \cdots, A_n$, that is,

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$

and

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n.$$

The finite union $\bigcup_{i=1}^{n} A_i$ is an event whose elements belong to **at least one** of the n events. The finite intersection $\bigcap_{i=1}^{n} A_i$ is an event whose elements belong to **all** the n events.

Any two events, A and B, are said to be mutually **exclusive** or **disjoint** if $A \cap B = \emptyset$, that is, they do not contain common elements. Obviously, by definition, any event is disjoint of its complement, i.e., $A \cap A^c = \emptyset$. The operations of union and intersection are:

1. **Commutative**:

$$A \cup B = B \cup A,$$

$$A \cap B = B \cap A;$$

2. **Associative**:
$$\begin{aligned}(A \cup B) \cup C &= A \cup (B \cup C)\\ &= A \cup B \cup C\\ (A \cap B) \cap C &= A \cap (B \cap C)\\ &= A \cap B \cap C\end{aligned} \tag{3.1}$$

3. **Distributive**:
$$\begin{aligned}A \cap (B \cup C) &= (A \cap B) \cup (A \cap C)\\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C).\end{aligned} \tag{3.2}$$

The intersection of events is sometimes denoted as a product, that is,

$$A_1 \cap A_2 \cap \cdots \cap A_n \equiv A_1 A_2 A_3 \cdots A_n.$$

The following law, called the **De-Morgan Rule**, is fundamental to the algebra of events and yields the complement of the union, or intersection, of two events, namely:

1. $(A \cup B)^c = A^c \cap B^c$ (3.3)
2. $(A \cap B)^c = A^c \cup B^c$.
   Finally, we define the notion of **partition**. A collection of n events $E_1, \cdots, E_n$ is called a **partition** of the sample space S, if
   (i) $\bigcup_{i=1}^{n} E_i = S$,
   (ii) $E_i \cap E_j = \emptyset$ for all $i \neq j$ $(i, j = 1, \cdots, n)$.

That is, the events in any partition are mutually **disjoint**, and their union exhaust all the sample space.

Example 3.5. The experiment is to generate on the computer a random number, U, in the interval $(0, 1)$. A random number in $(0, 1)$ can be obtained as

$$U = \sum_{j=1}^{\infty} I_j 2^{-j},$$

where I_j is the random result of tossing a coin, that is,

$$I_j = \begin{cases} 1, & \text{if Head}\\ 0, & \text{if Tail.}\end{cases}$$

To generate random numbers from a set of integers, the summation index j is bounded by a finite number N. This method is, however, not practical for generating random numbers on a continuous interval. Computer programs generate "pseudo-random" numbers. Methods for generating random numbers are described in various books on simulation (see Bratley, Fox and Schrage (1983)). Most commonly applied is the **linear congruential generator**. This method is based on the recursive equation

$$U_i = (aU_{i-1} + c) \bmod m, \quad i = 1, 2, \cdots.$$

The parameters a, c and m depend on the computer's architecture. In many programs, $a = 65539$, $c = 0$ and $m = 2^{31} - 1$. The first integer X_0 is called the "seed." Different choices of the parameters a, c and m yield "pseudo-random" sequences with different statistical properties.

The sample space of this experiment is

$$S = \{u : 0 \leq u \leq 1\}$$

Let E_1 and E_2 be the events

$$E_1 = \{u : 0 \le u \le 0.5\},$$
$$E_2 = \{u : 0.35 \le u \le 1\}.$$

The union of these two events is

$$E_3 = E_1 \cup E_2 = \{u : 0 \le u \le 1\} = S.$$

The intersection of these events is

$$E_4 = E_1 \cap E_2 = \{u : 0.35 \le u < 0.5\}.$$

Thus, E_1 and E_2 are **not** disjoint.

The complementary events are

$$E_1^c = \{u : 0.5 \le u < 1\} \quad \text{and} \quad E_2^c = \{u : u < 0.35\}$$

$E_1^c \cap E_2^c = \emptyset$; i.e., the complementary events are disjoint. By DeMorgan's law

$$\begin{aligned}(E_1 \cap E_2)^c &= E_1^c \cup E_2^c \\ &= \{u : u < 0.35 \quad \textbf{or} \quad u \ge 0.5\}.\end{aligned}$$

However,

$$\emptyset = S^c = (E_1 \cup E_2)^c = E_1^c \cap E_2^c.$$

Finally, the following is a partition of S:

$$B_1 = \{u : u < 0.1\}, \quad B_2 = \{u : 0.1 \le u < 0.2\},$$
$$B_3 = \{u : 0.2 \le u < 0.5\}, \quad B_4 = \{u : 0.5 \le u < 1\}.$$

Notice that $B_4 = E_1^c$. ■

Different identities can be derived by the above rules of operations on events, a few will be given as exercises.

3.1.3 Probabilities of events

A probability function $\Pr\{\cdot\}$ assigns to events of S real numbers, following the following basic axioms.

1. $\Pr\{E\} \ge 0$
2. $\Pr\{S\} = 1.$
3. **If** $E_1, \cdots, E_n$ $(n \ge 1)$ **are mutually disjoint events, then**

$$\Pr\left\{\bigcup_{i=1}^{n} E_i\right\} = \sum_{i=1}^{n} \Pr\{E_i\}.$$

From these three basic axioms, we deduce the following results.

Result 1. If $A \subset B$ **then**

$$\Pr\{A\} \le \Pr\{B\}.$$

Indeed, since $A \subset B$, $B = A \cup (A^c \cap B)$. Moreover $A \cap A^c \cap B = \emptyset$. Hence, by Axioms 1 and 3, $\Pr\{B\} = \Pr\{A\} + \Pr\{A^c \cap B\} \ge \Pr\{A\}$.

Thus, if E, is any event, since $E \subset S$, $0 \le \Pr\{E\} \le 1$.

Result 2. For any event E, $\Pr\{E^c\} = 1 - \Pr\{E\}$.

Indeed $S = E \cup E^c$. Since $E \cap E^c = \emptyset$,

$$1 = \Pr\{S\} = \Pr\{E\} + \Pr\{E^c\}. \tag{3.4}$$

This implies the result.

Result 3. For any events A, B

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}. \tag{3.5}$$

Indeed, we can write

$$A \cup B = A \cup A^c \cap B,$$

where $A \cap (A^c \cap B) = \emptyset$. Thus, by the third axiom,

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{A^c \cap B\}.$$

Moreover, $B = A^c \cap B \cup A \cap B$, where again $A^c \cap B$ and $A \cap B$ are disjoint. Thus, $\Pr\{B\} = \Pr\{A^c \cap B\} + \Pr\{A \cap B\}$, or $\Pr\{A^c \cap B\} = \Pr\{B\} - \Pr\{A \cap B\}$. Substituting this above we obtain the result.

Result 4. If $B_1, \cdots, B_n$ $(n \geq 1)$ **is a partition of** S, **then for any event** E,

$$\Pr\{E\} = \sum_{i=1}^{n} \Pr\{E \cap B_i\}.$$

Indeed, by the distributive law,

$$\begin{aligned} E = E \cap S = E \cap \left(\bigcup_{i=1}^{n} B_i \right) \\ = \bigcup_{i=1}^{n} EB_i. \end{aligned}$$

Finally, since $B_1, \cdots, B_n$ are mutually disjoint, $(EB_i) \cap (EB_j) = E \cap B_i \cap B_j = \emptyset$ for all $i \neq j$. Therefore, by the third axiom

$$\Pr\{E\} = \Pr\left\{ \bigcup_{i=1}^{n} EB_i \right\} = \sum_{i=1}^{n} \Pr\{EB_i\}. \tag{3.6}$$

Example 3.6. Fuses are used to protect electronic devices from unexpected power surges. Modern fuses are produced on glass plates through processes of metal deposition and photographic lythography. On each plate several hundred fuses are simultaneously produced. At the end of the process the plates undergo precise cutting with special saws. A certain fuse is handled on one of three alternative cutting machines. Machine M_1 yields 200 fuses per hour. Machine M_2 yields 250 fuses per hour and machine M_3 yields 350 fuses per hour. The fuses are then mixed together. The proportions of defective parts that are typically produced on these machines are 0.01, 0.02, and 0.005, respectively. A fuse is chosen at random from the production of a given hour. What is the probability that it is compliant with the amperage requirements (non-defective)?

Let E_i be the event that the chosen fuse is from machine M_i $(i = 1, 2, 3)$. Since the choice of the fuse is random, each fuse has the same probability $\frac{1}{800}$ to be chosen. Hence, $\Pr\{E_1\} = \frac{1}{4}$, $\Pr\{E_2\} = \frac{5}{16}$ and $\Pr\{E_3\} = \frac{7}{16}$.

Let G denote the event that the selected fuse is non-defective. For example for machine M_1, $\Pr\{G\} = 1 - 0.01 = 0.99$. We can assign $\Pr\{G \cap M_1\} = 0.99 \times 0.25 = 0.2475$, $\Pr\{G \cap M_2\} = 0.98 \times \frac{5}{16} = 0.3062$ and $\Pr\{G \cap M_3\} = 0.995 \times \frac{7}{16} = 0.4353$. Hence, the probability of selecting a non-defective fuse is, according to Result 4,

$$\Pr\{G\} = \Pr\{G \cap M_1\} + \Pr\{G \cap M_2\} + \Pr\{G \cap M_3\} = 0.989.$$

■

Example 3.7. Consider the problem of generating random numbers, discussed in Example 3.5. Suppose that the probability function assigns any interval $I(a,b) = \{u : a < u < b\}$, $0 \leq a < b \leq 1$, the probability

$$\Pr\{I(a,b)\} = b - a.$$

Let $E_3 = I(0.1, 0.4)$ and $E_4 = I(0.2, 0.5)$. $C = E_3 \cup E_4 = I(0.1, 0.5)$. Hence,

$$\Pr\{C\} = 0.5 - 0.1 = 0.4.$$

On the other hand $\Pr\{E_3 \cap E_4\} = 0.4 - 0.2 = 0.2$.

$$\begin{aligned}\Pr\{E_3 \cup E_4\} &= \Pr\{E_3\} + \Pr\{E_4\} - \Pr\{E_3 \cap E_4\} \\ &= (0.4 - 0.1) + (0.5 - 0.2) - 0.2 = 0.4.\end{aligned}$$

This illustrates Result 3 ■

3.1.4 Probability functions for random sampling

Consider a finite population P, and suppose that the random experiment is to select a random sample from P, with or without replacement. More specifically let $L_N = \{w_1, w_2, \cdots, w_N\}$ be a **list** of the elements of P, where N is its size. w_j $(j = 1, \cdots, N)$ is an identification number of the j-th element.

Suppose that a sample of size n is drawn from L_N [respectively, P] **with** replacement. Let W_1 denote the first element selected from L_N. If j_1 is the index of this element, then $W_1 = w_{j_1}$. Similarly, let W_i $(i = 1, \ldots, n)$ denote the i-th element, of the sample. The corresponding sample space is the collection

$$S = \{(W_1, \cdots, W_n) : W_i \in L_N, \quad i = 1, 2, \cdots, n\}$$

of all samples, with replacement from L_N. The total number of possible samples is N^n. Indeed, w_{j_1} could be any one of the elements of L_N, and so are $w_{j_2}, \cdots, w_{j_n}$. With each one of the N possible choices of w_{j_1} we should combine the N possible choices of w_{j_2} and so on. Thus, there are N^n possible ways of selecting a sample of size n, with replacement. The sample points are the elements of S (possible samples). The sample is called **random with replacement**, RSWR, if each one of these N^n possible samples is assigned the same probability, $1/N^n$, of being selected.

Let $M(i)$ $(i = 1, \cdots, N)$ be the number of samples in S, which contain the i-th element of L_N (at least once). Since sampling is **with** replacement

$$M(i) = N^n - (N-1)^n.$$

Indeed, $(N-1)^n$ is the number of samples with replacement, which do not include w_i. Since all samples are equally probable, the probability that a RSWR $\mathbf{S}_n$ includes w_i $(i = 1, \cdots, N)$ is

$$\begin{aligned}\Pr\{w_i \in \mathbf{S}_n\} &= \frac{N^n - (N-1)^n}{N^n} \\ &= 1 - \left(1 - \frac{1}{N}\right)^n.\end{aligned}$$

If $n > 1$, then the above probability is larger than $1/N$ which is the probability of selecting the element W_i in any given trial, but smaller than n/N. Notice also that this probability does not depend on i, that is, all elements of L_N have the same probability of being included in an RSWR. It can be shown that the probability that w_i is included in the sample exactly once is $\frac{n}{N}\left(1 - \frac{1}{N}\right)^{n-1}$. If sampling is **without** replacement, the number of sample points in S is $N(N-1)\cdots(N-n+1)/n!$, since the order of selection is immaterial. The number of sample points which include w_i is $M(i) = (N-1)(N-2)\cdots(N-n+1)/(n-1)!$. A sample $\mathbf{S}_n$ is called **random without replacement**, RSWOR, if all possible samples are equally probable. Thus, under RSWOR,

$$\Pr\{w_i \in \mathbf{S}_n\} = \frac{n!M(i)}{N(N-1)\cdots(N-n+1)} = \frac{n}{N},$$

for all $i = 1, \cdots, N$.

We consider now events, which depend on the attributes of the elements of a population. Suppose that we sample to obtain information on the number of defective (non-standard) elements in a population. The attribute in this case is "the element complies to the requirements of the standard." Suppose that M out of N elements in L_N are non-defective (have the attribute). Let E_j be the event that j out of the n elements in the sample are non-defective. Notice that $E_0, \cdots, E_n$ is a partition of the sample space. What is the probability, under RSWR, of E_j? Let K_j^n denote the number of sample points in which j out of n are G elements (non-defective) and $(n - j)$ elements are D (defective). To determine K_j^n, we can proceed as follows:

Choose first j G's and $(n - j)$ D's from the population. This can be done in $M^j(N - M)^{n-j}$ different ways. We have now to assign the j G's into j out of n components of the vector $(w_1, \cdots, w_n)$. This can be done in $n(n - 1) \cdots (n - j + 1)/j!$ possible ways. This is known as the number of combinations of j out of n, i.e.,

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}, \quad j = 0, 1, \cdots, n \tag{3.7}$$

where $k! = 1 \cdot 2 \cdot \cdots \cdot k$ is the product of the first k positive integers, $0! = 1$. Hence, $K_j^n = \binom{n}{j} M^j(N - M)^{n-j}$. Since every sample is equally probable, under RSWR,

$$\Pr\{E_{j:n}\} = K_j^n/N^n = \binom{n}{j} P^j(1 - P)^{n-j}, \quad j = 0, \cdots, n \tag{3.8}$$

where $P = M/N$. If sampling is without replacement, then

$$K_j^n = \binom{M}{j}\binom{N-M}{n-j}$$

and

$$\Pr\{E_j\} = \frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}}. \tag{3.9}$$

These results are valid since the order of selection is immaterial for the event E_j.

These probabilities of E_j under RSWR and RSWOR are called, respectively, the **binomial** and **hypergeometric** probabilities.

Example 3.8. The experiment consists of randomly transmitting a sequence of binary signals, 0 or 1. What is the probability that 3 out of 6 signals are 1's? Let E_3 denote this event.

The sample space of 6 signals consists of 2^6 points. Each point is equally probable. The probability of E_3 is

$$\begin{aligned}\Pr\{E_3\} &= \binom{6}{3}\frac{1}{2^6} = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 64} \\ &= \frac{20}{64} = \frac{5}{16} = 0.3125.\end{aligned}$$

■

Example 3.9. Two out of ten television sets are defective. A RSWOR of $n = 2$ sets is chosen. What is the probability that the two sets in the sample are good (non-defective)? This is the hypergeometric probability of E_0 when $M = 2$, $N = 10$, $n = 2$, that is,

$$\Pr\{E_0\} = \frac{\binom{8}{2}}{\binom{10}{2}} = \frac{8 \cdot 7}{10 \cdot 9} = 0.622.$$

■

3.1.5 Conditional probabilities and independence of events

In this section we discuss the notion of conditional probabilities. When different events are related, the realization of one event may provide us relevant information to improve our probability assessment of the other event(s). In Section 3.1.3 we gave an example with three machines which manufacture the same part but with different production rates and different proportions of defective parts in the output of those machines. The random experiment was to choose at random a part from the mixed yield of the three machines.

We saw earlier that the probability that the chosen part is non-defective is 0.989. If we can identify, before the quality test, from which machine the part came, the probabilities of non-defective would be conditional on this information.

The probability of choosing at random a non-defective part from machine M_1, is 0.99. If we are given the information that the machine is M_2, the probability is 0.98, and given machine M_3, the probability is 0.995. These probabilities are called **conditional probabilities**. The information given changes our probabilities.

We define now formally the concept of conditional probability.

Let A **and** B **be two events such that** $\Pr\{B\} > 0$**. The conditional probability of** A**, given** B**, is**

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}. \tag{3.10}$$

Example 3.10. The random experiment is to measure the length of a steel bar.

The sample space is $S = (19.5, 20.5)$ [cm]. The probability function assigns any subinterval a probability equal to its length. Let $A = (19.5, 20.1)$ and $B = (19.8, 20.5)$. $\Pr\{B\} = 0.7$. Suppose that we are told that the length belongs to the interval B, and we have to guess whether it belongs to A. We compute the conditional probability

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} = \frac{0.3}{0.7} = 0.4286.$$

On the other hand, if the information that the length belongs to B is not given, then $\Pr\{A\} = 0.6$. Thus, there is a difference between the conditional and non-conditional probabilities. This indicates that the two events A and B are dependent. ■

Definition. Two events A, B are called **independent** if

$$\Pr\{A \mid B\} = \Pr\{A\}.$$

If A and B are independent events, then

$$\Pr\{A\} = \Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

or, equivalently,

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}.$$

If there are more than two events, $A_1, A_2, \cdots, A_n$, we say that the events are **pairwise independent** if

$$\Pr\{A_i \cap A_j\} = \Pr\{A_i\}\Pr\{A_j\} \quad \text{for all} \quad i \neq j, \qquad i,j = 1, \cdots, n.$$

The n events are said to be **mutually independent** if, for any subset of k events, $k = 2, \ldots, n$, indexed by $A_{i_1}, \ldots, A_{i_k}$,

$$\Pr\{A_{i_1} \cap A_{i_2} \cdots \cap A_{i_k}\} = \Pr\{A_{i_1}\} \cdots \Pr\{A_{i_n}\}.$$

In particular, if n events are mutually independent, then

$$\Pr\left\{\bigcap_{i=1}^{n} A_i\right\} = \prod_{i=1}^{n} \Pr\{A_i\}. \tag{3.11}$$

One can show examples of events which are pairwise independent but **not** mutually independent.

We can further show (see exercises) that if two events are independent, then the corresponding complementary events are independent. Furthermore, if n events are mutually independent, then any pair of events is pairwise independent, every three events are triplewise independent, etc.

Example 3.11. Five identical parts are manufactured in a given production process. Let $E_1, \cdots, E_5$ be the events that these five parts comply with the quality specifications (non-defective). Under the model of mutual independence the probability that **all** the five parts are indeed non-defective is

$$\Pr\{E_1 \cap E_2 \cap \cdots \cap E_5\} = \Pr\{E_1\}\Pr\{E_2\}\cdots\Pr\{E_5\}.$$

Since these parts come from the same production process, we can assume that $\Pr\{E_i\} = p$, all $i = 1, \cdots, 5$. Thus, the probability that **all** the 5 parts are non-defective is p^5.

What is the probability that one part is defective and all the other four are non-defective? Let A_1 be the event that one out of five parts is defective. In order to simplify the notation, we write the intersection of events as their product. Thus,

$$A_1 = E_1^c E_2 E_3 E_4 E_5 \cup E_1 E_2^c E_3 E_4 E_5 \cup E_1 E_2 E_3^c E_4 E_5 \cup E_1 E_2 E_3 E_4^c E_5 \cup E_1 E_2 E_3 E_4 E_5^c.$$

A_1 is the union of five **disjoint** events. Therefore

$$\begin{aligned}\Pr\{A_1\} &= \Pr\{E_1^c E_2 \cdots E_5\} + \cdots + \Pr\{E_1 E_2 \cdots E_5^c\}\\ &= 5p^4(1-p).\end{aligned}$$

Indeed, since $E_1, \cdots, E_5$ are **mutually** independent events

$$\Pr\{E_1^c E_2 \cdots E_5\} = \Pr\{E_1^c\}\Pr\{E_2\}\cdots\Pr\{E_5\} = (1-p)p^4.$$

Also,

$$\Pr\{E_1 E_2^c E_3 E_4 E_5\} = (1-p)p^4,$$

etc. Generally, if J_5 denotes the number of defective parts among the five ones,

$$\Pr\{J_5 = i\} = \binom{5}{i} p^{(5-i)}(1-p)^i, \quad i = 0, 1, 2, \cdots, 5.$$

■

In the context of a graph representing directed links between variables, a directed acyclic graph (DAG) represents a qualitative causality model. The model parameters are derived by applying the Markov property, where the conditional probability distribution at each node depends only on its parents. For discrete random variables, this conditional probability is often represented by a table, listing the local probability that a child node takes on each of the feasible values for each combination of values of its parents. The joint distribution of a collection of variables can be determined uniquely by these local conditional probability tables. A Bayesian Network (BN) is represented by a DAG. A BN reflects a simple **conditional independence** statement, namely that each variable is independent of its non-descendants in the graph given the state of its parents. This property is used to reduce, sometimes significantly, the number of parameters that are required to characterize the joint probability distribution of the variables. This reduction provides an efficient way to compute the posterior probabilities given the evidence present in the data. We do not cover here BN. For examples of applications of BN with an general introduction to this topic, see Kenett (2012).

3.1.6 Bayes formula and its application

Bayes formula, which is derived in the present section, provides us with a fundamental formula for weighing the evidence in the data concerning unknown parameters, or some unobservable events.

Suppose that the results of a random experiment depend on some event(s) which is (are) not directly observable. The observable event is related to the unobservable one(s) via the conditional probabilities. More specifically, suppose that $\{B_1, \cdots, B_m\}$ $(m \geq 2)$ is a partition of the sample space. The events $B_1, \cdots, B_m$ are not directly observable, or verifiable. The random experiment results in an event A (or its complement). The conditional probabilities $\Pr\{A \mid B_i\}$, $i = 1, \cdots, m$ are known. The question is whether, after observing the event A, can we assign probabilities to the events $B_1, \cdots, B_m$?

In order to weigh the evidence that A has on $B_1, \cdots, B_m$, we first assume some probabilities $\Pr\{B_i\}$, $i = 1, \cdots, m$, which are called **prior probabilities**. The prior probabilities express our degree of belief in the occurrence of the events B_i ($i = 1, \cdots, m$). After observing the event A we convert the prior probabilities of B_i ($i = 1, \cdots, m$) to **posterior probabilities** $\Pr\{B_i \mid A\}$, $i = 1, \cdots, m$ by using **Bayes formula**

$$\Pr\{B_i \mid A\} = \frac{\Pr\{B_i\} \Pr\{A \mid B_i\}}{\sum_{j=1}^{m} \Pr\{B_j\} \Pr\{A \mid B_j\}}, \quad i = 1, \cdots, m. \tag{3.12}$$

These posterior probabilities reflect the weight of evidence that the event A has concerning $B_1, \cdots, B_m$.

Bayes formula can be obtained from the basic rules of probability. Indeed, assuming that $\Pr\{A\} > 0$,

$$\begin{aligned}\Pr\{B_i \mid A\} &= \frac{\Pr\{A \cap B_i\}}{\Pr\{A\}} \\ &= \frac{\Pr\{B_i\} \Pr\{A \mid B_i\}}{\Pr\{A\}}.\end{aligned}$$

Furthermore, since $\{B_1, \cdots, B_m\}$ is a partition of the sample space,

$$\Pr = \{A\} = \sum_{j=1}^{m} \Pr\{B_j\} \Pr\{A \mid B_j\}.$$

Substituting this expression above, we obtain Bayes formula.

The following example illustrates the applicability of Bayes formula to a problem of decision-making.

Example 3.12. Two vendors B_1, B_2 produce ceramic plates for a given production process of hybrid micro circuits. The parts of vendor B_1 have probability $p_1 = 0.10$ of being defective. The parts of vendor B_2 have probability $p_2 = 0.05$ of being defective. A delivery of $n = 20$ parts arrives, but the label which identifies the vendor is missing. We wish to apply Bayes formula to assign a probability that the package came from vendor B_1.

Suppose that it is a-priori equally likely that the package was mailed by vendor B_1 or vendor B_2. Thus, the prior probabilities are $\Pr\{B_1\} = \Pr\{B_2\} = 0.5$. We inspect the twenty parts in the package, and find $J_{20} = 3$ defective items. A is the event $\{J_{20} = 3\}$. The conditional probabilities of A, given B_i ($i = 1, 2$) are

$$\begin{aligned}\Pr\{A \mid B_1\} &= \binom{20}{3} p_1^3 (1 - p_1)^{17} \\ &= 0.1901.\end{aligned}$$

Similarly

$$\begin{aligned}\Pr\{A \mid B_2\} &= \binom{20}{3} p_2^3 (1 - p_2)^{17} \\ &= 0.0596.\end{aligned}$$

According to Bayes formula

$$\begin{aligned}\Pr\{B_1 \mid A\} &= \frac{0.5 \times 0.1901}{0.5 \times 0.1901 + 0.5 \times 0.0596} = 0.7613 \\ \Pr\{B_2 \mid A\} &= 1 - \Pr\{B_1 \mid A\} = 0.2387.\end{aligned}$$

Thus, after observing three defective parts in a sample of $n = 20$ ones, we believe that the delivery came from vendor B_1. The posterior probability of B_1, given A, is more than 3 times higher than that of B_2 given A. The a-priori odds of B_1 against B_2 were 1:1. The a-posteriori odds are 19:6. ■

3.2 Random variables and their distributions

Random variables are formally defined as **real-valued functions**, $X(w)$, **over the sample space**, S, **such that, events** $\{w : X(w) \leq x\}$ **can be assigned probabilities, for all** $-\infty < x < \infty$, where w are the elements of S.

Example 3.13. Suppose that S is the sample space of all RSWOR of size n, from a finite population, P, of size N. $1 \leq n < N$. The elements w of S are subsets of distinct elements of the population P. A random variable $X(w)$ is some function which assigns w a finite real number, for example, the number of "defective" elements of w. In the present example $X(w) = 0, 1, \cdots, n$ and

$$\Pr\{X(w) = j\} = \frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}}, \quad j = 0, \cdots, n,$$

where M is the number of "defective" elements of P. ■

Example 3.14. Another example of random variable is the compressive strength of a concrete cube of a certain dimension. In this example, the random experiment is to manufacture a concrete cube according to a specified process. The sample space S is the space of all cubes, w, that can be manufactured by this process. $X(w)$ is the compressive strength of w. The probability function assigns each event $\{w : X(w) \leq \xi\}$ a probability, according to some mathematical model which satisfies the laws of probability. Any continuous non-decreasing function $F(x)$, such that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$ will do the job. For example, for the compressive strength of concrete cubes, the following model has been shown to fit experimental results

$$\Pr\{X(w) \leq x\} = \begin{cases} 0, & x \leq 0 \\ \dfrac{1}{\sqrt{2\pi}\sigma}\displaystyle\int_0^x \frac{1}{y}\exp\left\{-\frac{(\ln\ y - \mu)^2}{2\sigma^2}\right\} dy, & 0 < x < \infty. \end{cases}$$

The constants μ and σ, $-\infty < \mu < \infty$ and $0 < \sigma < \infty$, are called **parameters** of the model. Such parameters characterize the manufacturing process. ■

We distinguish between **two types** of random variables: **discrete** and **continuous**.

Discrete random variables, $X(w)$, are random variables having a finite or countable range. For example, the number of "defective" elements in a random sample is a discrete random variable. The number of blemishes on a ceramic plate is a discrete random variable. A **continuous random variable** is one whose range consists of whole intervals of possible values. The weight, length, compressive strength, tensile strength, cycle time, output voltage, etc. are continuous random variables.

3.2.1 Discrete and continuous distributions

3.2.1.1 Discrete random variables

Suppose that a discrete random variable can assume the distinct values $x_0, \cdots, x_k$ (k is finite or infinite). The function

$$p(x) = \Pr\{X(w) = x\}, \quad -\infty < x < \infty \tag{3.13}$$

is called the **probability distribution function** (p.d.f.) of X.

Notice that if x is not one of the values in the specified range $S_X = \{x_j; j = 0, 1, \cdots, k\}$, then $\{X(w) = x\} = \phi$ and $p(x) = 0$. Thus, $p(x)$ assumes positive values only on the specified sequence S_X (S_X is also called the sample space of X), such that

1. $p(x_j) \geq 0, j = 0, \cdots, k$ (3.14)

2. $\sum_{j=0}^{k} p(x_j) = 1.$

Example 3.15. Suppose that the random experiment is to cast a die once. The sample points are six possible faces of the die, $\{w_1, \cdots, w_6\}$. Let $X(w_j) = j, j = 1, \cdots, 6$, be the random variable, representing the face number. The probability model yields

$$p(x) = \begin{cases} \frac{1}{6}, & \text{if } x = 1, 2, \cdots, 6 \\ 0, & \text{otherwise.} \end{cases}$$

■

Example 3.16. Consider the example of Section 3.1.5, of drawing independently $n = 5$ parts from a production process, and counting the number of "defective" parts in this sample. The random variable is $X(w) = J_5$. $S_X = \{0, 1, \cdots, 5\}$ and the p.d.f. is

$$p(x) = \begin{cases} \binom{5}{x} p^{5-x}(1-p)^x, & x = 0, 1, \cdots, 5 \\ 0, & \text{otherwise.} \end{cases}$$

■

The probability of the event $\{X(w) \le x\}$, for any $-\infty < x < \infty$, can be computed by summing the probabilities of the values in S_X, which belong to the interval $(-\infty, x]$. This sum is called the **cumulative distribution function** (c.d.f.) of X, and denoted by

$$\begin{aligned} P(x) &= \Pr\{X(w) \le x\} \\ &= \sum_{\{x_j \le x\}} p(x_j), \end{aligned} \tag{3.15}$$

where $x_j \in S_X$.

The c.d.f. corresponding to Example 3.16 is

$$P(x) = \begin{cases} 0, & x < 0 \\ \sum_{j=0}^{[x]} \binom{5}{j} p^{5-j}(1-p)^j, & 0 \le x < 5 \\ 1, & 5 \le x \end{cases}$$

where $[x]$ denotes the **integer part of** x, that is, the **largest integer** smaller or equal to x.

Generally the graph of the p.d.f. of a discrete variable is a bar-chart (see Figure 3.1). The corresponding c.d.f. is a step function, as shown in Figure 3.2.

3.2.1.2 *Continuous random variables*

In the case of continuous random variables, the model assigns the variable under consideration a function $F(x)$ which is:

(i) continuous;
(ii) non-decreasing, i.e., if $x_1 < x_2$ then $F(x_1) \le F(x_2)$
(iii) and
(iv) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Such a function can serve as a **cumulative distribution function** (c.d.f.), for X.

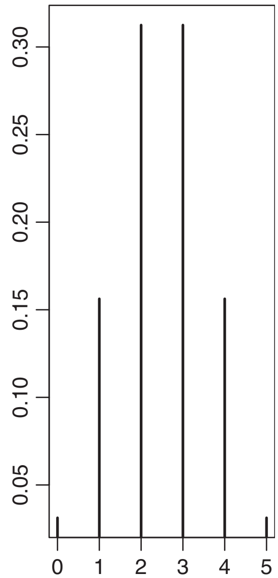

Figure 3.1 *The graph of the p.d.f. P(x) = equation random variable*

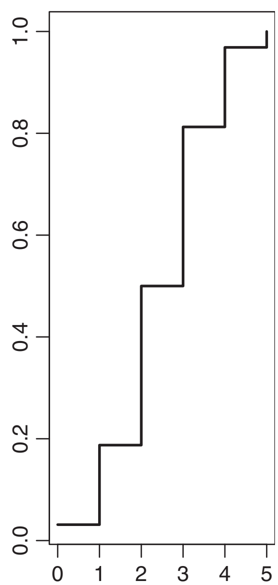

Figure 3.2 *The graph of the c.f.g. P(x) = equation random variable*

An example of a c.d.f. for a continuous random variable which assumes non-negative values, for example, the operation total time until a part fails, is

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

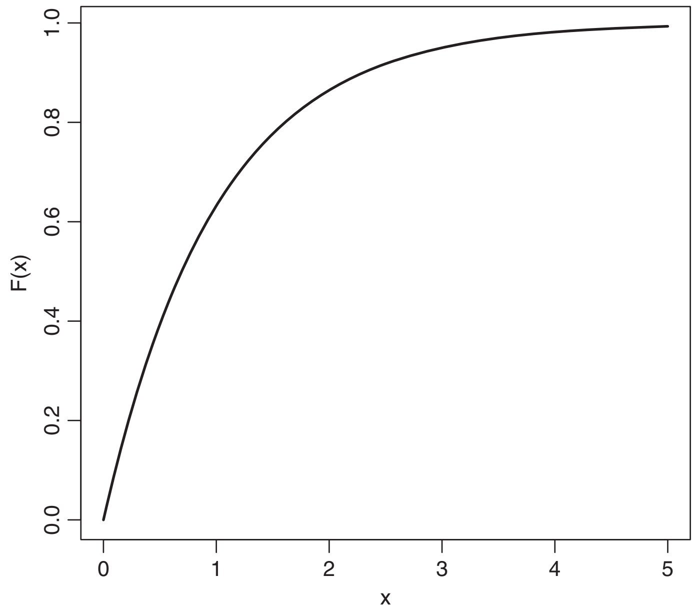

Figure 3.3 *c.d.f. of* $F(x) = 1 - e^{-x}$

This function (see Figure 3.3) is continuous, monotonically increasing, and $\lim_{x\to\infty} F(x) = 1 - \lim_{x\to\infty} e^{-x} = 1$. If the c.d.f. of a continuous random variable can be represented as

$$F(x) = \int_{-\infty}^{x} f(y)dy, \tag{3.16}$$

for some $f(y) \geq 0$, then we say that $F(x)$ is **absolutely continuous** and $f(x) = \frac{d}{dx}F(x)$. (The derivative $f(x)$ may not exist on a finite number of x values, in any finite interval.) The function $f(x)$ is called the **probability density function** (p.d.f.) of X.

In the above example of total operational time, the p.d.f. is

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ e^{-x}, & \text{if } x \geq 0. \end{cases}$$

Thus, as in the discrete case, we have $F(x) = \Pr\{X \leq x\}$. It is now possible to write

$$\Pr\{a \leq X < b\} = \int_{a}^{b} f(t)dt = F(b) - F(a) \tag{3.17}$$

or

$$\Pr\{X \geq b\} = \int_{b}^{\infty} f(t)dt = 1 - F(b). \tag{3.18}$$

Thus, if X has the exponential c.d.f.

$$\Pr\{1 \leq X \leq 2\} = F(2) - F(1) = e^{-1} - e^{-2} = 0.2325.$$

There are certain phenomena which require more complicated modeling. The random variables under consideration may not have purely discrete or purely absolutely continuous distribution. There are many random variables with c.d.f.'s which

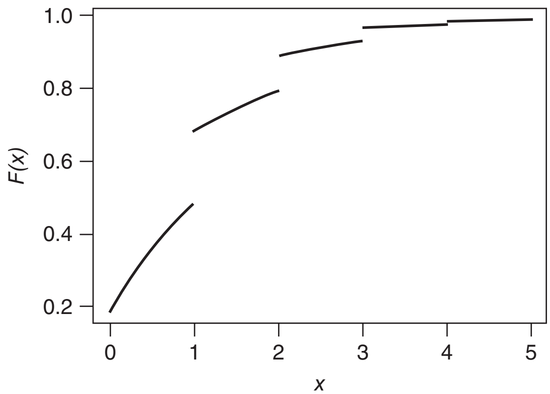

Figure 3.4 *c.d.f. of the mixture distribution* $F(x) = 0.5\,(1 - e^{-x}) + 0.5 \times e^{-1} \sum_{j=0}^{[x]} \frac{1}{j!}$

are absolutely continuous within certain intervals, and have jump points (points of discontinuity) at the end points of the intervals. Distributions of such random variables can be expressed as mixtures of purely discrete c.d.f., $F_d(x)$, and of absolutely continuous c.d.f., $F_{ac}(x)$, that is,

$$F(x) = pF_d(x) + (1 - p)F_{ac}(x), \quad -\infty < x < \infty, \tag{3.19}$$

where $0 \leq p \leq 1$ (see Figure 3.4).

Example 3.17. A distribution which is a mixture of discrete and continuous distributions is obtained, for example, when a measuring instrument is not sensitive enough to measure small quantities or large quantities which are outside its range. This could be the case for a weighing instrument which assigns the value 0 [mg] to any weight smaller than 1 [mg], the value 1 [g] to any weight greater than 1 gram, and the correct weight to values in between.

Another example is the total number of minutes, within a given working hour, that a service station is busy serving customers. In this case the c.d.f. has a jump at 0, of height p, which is the probability that the service station is idle at the beginning of the hour, and no customer arrives during that hour. In this case,

$$F(x) = p + (1 - p)G(x), \quad 0 \leq x < \infty,$$

where $G(x)$ is the c.d.f. of the total service time, $G(0) = 0$. ■

3.2.2 Expected values and moments of distributions

The **expected value** of a function $g(X)$, under the distribution $F(x)$, is

$$E_F\{g(X)\} = \begin{cases} \int_{-\infty}^{\infty} g(x)f(x)dx, & \text{if } X \text{ is continuous} \\ \sum_{j=0}^{k} g(x_j)p(x_j), & \text{if } X \text{ is discrete.} \end{cases}$$

In particular,

$$\mu_l(F) = E_F\{X^l\}, \quad l = 1, 2, \cdots \tag{3.20}$$

is called the l-th **moment** of $F(x)$. $\mu_1(F) = E_F\{X\}$ is the expected value of X, or the **population mean**, according to the model $F(x)$.

Moments around $\mu_1(F)$ are called **central moments**, which are

$$\mu_l^*(F) = E\{(X - \mu_1(F))^l\}, \quad l = 1, 2, 3, \cdots. \tag{3.21}$$

Obviously, $\mu_1^*(F) = 0$. The **second central moment** is called the **variance** of $F(x)$, $V_F\{X\}$.

In the following, the notation $\mu_l(F)$ will be simplified to μ_l, if there is no room for confusion.

Expected values of a function $g(X)$, and in particular the moments, may not exist, since an integral $\int_{-\infty}^{\infty} x^l f(x)dx$ may not be well defined. An example of such a case is the distribution, called the **Cauchy distribution**, with p.d.f.

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

Notice that under this model, moments do not exist for any $l = 1, 2, \cdots$. Indeed, the integral

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x}{1+x^2} dx$$

does not exist. If the second moment exists, then

$$V\{X\} = \mu_2 - \mu_1^2.$$

Example 3.18. Consider the random experiment of casting a die once. The random variable, X, is the face number. Thus $p(x) = \frac{1}{6}$, $x = 1, \cdots, 6$ and

$$\mu_1 = E\{X\} = \frac{1}{6} \sum_{j=1}^{6} j = \frac{6(6+1)}{2 \times 6} = \frac{7}{2} = 3.5$$

$$\mu_2 = \frac{1}{6} \sum_{j=1}^{6} j^2 = \frac{6(6+1)(2 \times 6 + 1)}{6 \times 6} = \frac{7 \times 13}{6} = \frac{91}{6} = 15.167.$$

The variance is

$$\begin{aligned} V\{X\} &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{182 - 147}{12} \\ &= \frac{35}{12}. \end{aligned}$$

■

Example 3.19. X has a continuous distribution with p.d.f.

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

Thus,

$$\mu_1 = \int_1^2 x dx = \frac{1}{2}\left(x^2 \Big|_1^2\right) = \frac{1}{2}(4-1) = 1.5$$

$$\mu_2 = \int_1^2 x^2 dx = \frac{1}{3}\left(x^3 \Big|_1^2\right) = \frac{7}{3}$$

$$V\{X\} = \mu_2 - \mu_1 = \frac{7}{3} - \frac{9}{4} = \frac{28-27}{12} = \frac{1}{12}.$$

■

The following is a useful formula when X assumes only positive values, i.e., $F(x) = 0$ for all $x \le 0$,

$$\mu_1 = \int_0^{\infty} (1 - F(x))dx, \tag{3.22}$$

for continuous c.d.f. $F(x)$. Indeed,

$$\mu_1 = \int_0^\infty xf(x)dx$$
$$= \int_0^\infty \left(\int_0^x dy\right) f(x)dx$$
$$= \int_0^\infty \left(\int_y^\infty f(x)dx\right) dy$$
$$= \int_0^\infty (1 - F(y))dy.$$

For example, suppose that $f(x) = \mu e^{-\mu x}$, for $x \geq 0$. Then $F(x) = 1 - e^{-\mu x}$ and

$$\int_0^\infty (1 - F(x))dx = \int_0^\infty e^{-\mu x}dx = \frac{1}{\mu}.$$

When X is discrete, assuming the values $\{1, 2, 3, \ldots\}$ then we have a similar formula

$$E\{X\} = 1 + \sum_{i=1}^{\infty}(1 - F(i)).$$

3.2.3 The standard deviation, quantiles, measures of skewness and kurtosis

The **standard deviation** of a distribution $F(x)$ is $\sigma = (V\{X\})^{1/2}$. The standard deviation is used as a measure of dispersion of a distribution. An important theorem in probability theory, called the **Chebychev Theorem**, relates the standard deviation to the probability of deviation from the mean. More formally, the theorem states that, if σ exists, then

$$\Pr\{|X - \mu_1| > \lambda\sigma\} \leq \frac{1}{\lambda^2}. \tag{3.23}$$

Thus, by this theorem, the probability that a random variable will deviate from its expected value by more than three standard deviations is less than 1/9, whatever the distribution is. This theorem has important implications, which will be highlighted later.

The p-th quantile of a distribution $F(x)$ is the smallest value of x, ξ_p such that $F(x) \geq p$. We also write $\xi_p = F^{-1}(p)$.

For example, if $F(x) = 1 - e^{-\lambda x}$, $0 \leq x < \infty$, where $0 < \lambda < \infty$, then ξ_p is such that

$$F(\xi_p) = 1 - e^{-\lambda\xi_p} = p.$$

Solving for ξ_p we get

$$\xi_p = -\frac{1}{\lambda} \cdot \ln\ (1 - p).$$

The **median** of $F(x)$ is $f^{-1}(.5) = \xi_{.5}$. Similarly $\xi_{.25}$ and $\xi_{.75}$ are the first and third quartiles of F.

A distribution $F(x)$ is **symmetric** about the mean $\mu_1(F)$ if

$$F(\mu_1 + \delta) = 1 - F(\mu_1 - \delta)$$

for all $\delta \geq 0$.

In particular, if F is symmetric, then $F(\mu_1) = 1 - F(\mu_1)$ or $\mu_1 = F^{-1}(.5) = \xi_{.5}$. Accordingly, the mean and median of a symmetric distribution coincide. In terms of the p.d.f., a distribution is symmetric about its mean if

$$f(\mu_1 + \delta) = f(\mu_1 - \delta), \quad \text{for all} \quad \delta \geq 0.$$

A commonly used index of **skewness** (asymmetry) is

$$\beta_3 = \frac{\mu_3^*}{\sigma^3}, \tag{3.24}$$

where μ_3^* is the third central moment of F. One can prove that **if** $F(x)$ **is symmetric, then** $\beta_3 = 0$. If $\beta_3 > 0$, we say that $F(x)$ is positively skewed, otherwise it is negatively skewed.

Example 3.20. Consider the **binomial** distribution, with p.d.f.

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x = 0, 1, \cdots, n.$$

In this case

$$\begin{aligned}\mu_1 &= \sum_{x=0}^{n} x \binom{n}{x} p^x(1-p)^{n-x} \\ &= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1}(1-p)^{n-1-(x-1)} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j(1-p)^{n-1-j} \\ &= np.\end{aligned}$$

Indeed,

$$\begin{aligned}x\binom{n}{x} &= x\frac{n!}{x!(n-x)!} = \frac{n!}{(x-1)!((n-1)-(x-1))!} \\ &= n\binom{n-1}{x-1}.\end{aligned}$$

Similarly, we can show that

$$\mu_2 = n^2p^2 + np(1-p),$$

and

$$\mu_3 = np[n(n-3)p^2 + 3(n-1)p + 1 + 2p^2].$$

The third central moment is

$$\begin{aligned}\mu_3^* &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \\ &= np(1-p)(1-2p).\end{aligned}$$

Furthermore,

$$\begin{aligned}V\{X\} &= \mu_2 - \mu_1^2 \\ &= np(1-p).\end{aligned}$$

Hence,

$$\sigma = \sqrt{np(1-p)}$$

and the index of asymmetry is

$$\begin{aligned}\beta_3 &= \frac{\mu_3^*}{\sigma_3} = \frac{np(1-p)(1-2p)}{(np(1-p))^{3/2}} \\ &= \frac{1-2p}{\sqrt{np(1-p)}}.\end{aligned}$$

Thus, if $p = \frac{1}{2}$, then $\beta_3 = 0$ and the distribution is symmetric. If $p < \frac{1}{2}$, the distribution is positively skewed, and it is negatively skewed if $p > \frac{1}{2}$. ■

In Chapter 2 we mentioned also the index of **kurtosis** (steepness). This is given by

$$\beta_4 = \frac{\mu_4^*}{\sigma^4}. \tag{3.25}$$

Example 3.21. Consider the exponential c.d.f.

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

The p.d.f. is $f(x) = e^{-x}$, $x \geq 0$. Thus, for this distribution

$$\begin{aligned} \mu_1 &= \int_0^\infty xe^{-x}dx = 1 \\ \mu_2 &= \int_0^\infty x^2e^{-x}dx = 2 \\ \mu_3 &= \int_0^\infty x^3e^{-x}dx = 6 \\ \mu_4 &= \int_0^\infty x^4e^{-x}dx = 24. \end{aligned}$$

Therefore,

$$\begin{aligned} V\{X\} &= \mu_2 - \mu_1^2 = 1, \\ \sigma &= 1 \\ \mu_4^* &= \mu_4 - 4\mu_3 \cdot \mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4 \\ &= 24 - 4 \times 6 \times 1 + 6 \times 2 \times 1 - 3 = 9. \end{aligned}$$

Finally, the index of kurtosis is

$$\beta_4 = 9.$$

■

3.2.4 Moment generating functions

The **moment generating function** (m.g.f.) of a distribution of X, is defined as a function of a real variable t,

$$M(t) = E\{e^{tX}\}. \tag{3.26}$$

$M(0) = 1$ for all distributions. $M(t)$, however, may not exist for some $t \neq 0$. To be useful, it is sufficient that $M(t)$ will exist in some interval containing $t = 0$.

For example, if X has a continuous distribution with p.d.f.

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b, \quad a < b \\ 0, & \text{otherwise} \end{cases}$$

then

$$M(t) = \frac{1}{b-a}\int_a^b e^{tx}dx = \frac{1}{t(b-a)}(e^{tb} - e^{ta}).$$

This is a differentiable function of t, for all t, $-\infty < t < \infty$.

On the other hand, if for $0 < \lambda < \infty$,

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 \le x < \infty \\ 0, & x < 0 \end{cases}$$

then

$$M(t) = \lambda \int_0^{\infty} e^{tx - \lambda x} dx = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

This m.g.f. exists only for $t < \lambda$. The m.g.f. $M(t)$ is a transform of the distribution $F(x)$, and the correspondence between $M(t)$ and $F(x)$ is one-to-one. In the above example, $M(t)$ is the Laplace transform of the p.d.f. $\lambda e^{-\lambda x}$. This correspondence is often useful in identifying the distributions of some statistics, as will be shown later.

Another useful property of the m.g.f. $M(t)$ is that often we can obtain the moments of $F(x)$ by differentiating $M(t)$. More specifically, consider the r-th order derivative of $M(t)$. Assuming that this derivative exists, and differentiation can be interchanged with integration (or summation), then

$$M^{(r)}(t) = \frac{d^r}{dt^r} \int e^{tx} f(x) dx = \int \left(\frac{d^r}{dt^r} e^{tx} \right) f(x) dx = \int x^r e^{tx} f(x) dx.$$

Thus, if these operations are justified, then

$$M^{(r)}(t)|_{t=0} = \int x^r f(x) dx = \mu_r. \tag{3.27}$$

In the following sections we will illustrate the usefulness of the m.g.f.

3.3 Families of discrete distribution

In the present section we discuss several families of discrete distributions, and illustrate their possible application in modeling industrial phenomena.

3.3.1 The binomial distribution

Consider n **identical** independent trials. In each trial the probability of "success" is fixed at some value p, and successive events of "success" or "failure" are **independent**. Such trials are called **Bernoulli trials**. The distribution of the number of "successes," J_n, is binomial with p.d.f.

$$b(j; n, p) = \binom{n}{j} p^j (1 - p)^{n-j}, \quad j = 0, 1, \cdots, n. \tag{3.28}$$

This p.d.f. was derived in Example 5.11 as a special case.

A binomial random variable, with parameters (n, p) will be designated as $B(n, p)$. n is a given integer and p belongs to the interval $(0, 1)$. The collection of all such binomial distributions is called the **binomial family**.

The binomial distribution is a proper model whenever we have a sequence of independent binary events $(0 - 1$, or "Success" and "Failure") with the same probability of "Success."

Example 3.22. We draw a random sample of $n = 10$ items from a mass production line of light bulbs. Each light bulb undergoes an inspection and if it complies with the production specifications, we say that the bulb is compliant (successful event). Let $X_i = 1$ if the i-th bulb is compliant and $X_i = 0$ otherwise. If we can assume that the probability of $\{X_i = 1\}$

is the same, p, for all bulbs and if the n events are mutually independent, then the number of bulbs in the sample which comply with the specifications, i.e., $J_n = \sum_{i=1}^{n} X_i$, has the binomial p.d.f. $b(i; n, p)$. Notice that if we draw a sample at random **with** replacement, RSWR, from a lot of size N, which contains M compliant units, then J_n is $B\left(n, \frac{M}{N}\right)$.

Indeed, if sampling is with replacement, the probability that the i-th item selected is compliant is $p = \frac{M}{N}$ for all $i = 1, \cdots, n$. Furthermore, selections are **independent** of each other. ■

The binomial c.d.f. will be denoted by $B(i; n, p)$. Recall that

$$B(i; n, p) = \sum_{j=0}^{i} b(j; n, p), \tag{3.29}$$

$i = 0, 1, \cdots, n$. The m.g.f. of $B(n, p)$ is

$$\begin{aligned} M(t) &= E\{e^{tX}\} \\ &= \sum_{j=0}^{n} \binom{n}{j} (pe^t)^j (1-p)^{n-j} \\ &= (pe^t + (1-p))^n, \quad -\infty < t < \infty. \end{aligned} \tag{3.30}$$

Notice that

$$M'(t) = n(pe^t + (1-p))^{n-1} pe^t$$

and

$$M''(t) = n(n-1)p^2 e^{2t}(pe^t + (1-p))^{n-2} + npe^t(pe^t + (1-p))^{n-1}.$$

The expected value and variance of $B(n, p)$ are

$$E\{J_n\} = np, \tag{3.31}$$

and

$$V\{J_n\} = np(1-p). \tag{3.32}$$

This was shown in Example 3.20 and can be verified directly by the above formulae of $M'(t)$ and $M''(t)$. To obtain the values of $b(i; n, p)$ we can use R or MINITAB. For example, suppose we wish to tabulate the values of the p.d.f. $b(i; n, p)$, and those of the c.d.f. $B(i; n, p)$ for $n = 30$ and $p = .60$. Below are R commands to generate a data frame with values as illustrated in Table 3.1.

```
< X <- data.frame(i=0:30,
                  b=dbinom(x=0:30, size=30, prob=0.6),
                  B=pbinom(q=0:30, size=30, prob=0.6))
< rm(X)
```

In MINITAB We first put in column $C1$ the integers $0, 1, \cdots, 30$ and put the value of $b(i; 30, .60)$ in column $C2$, and those of $B(i; 30, .60)$ in $C3$. To make MINITAB commands visible in the session window, go to Editor > Enable Commands. We then type the commands:

```
MTB> Set C1
DATA> 1(0:30/1)1
DATA> End.
MTB> PDF C1 C2;
SUBC> Binomial 30 0.60.
MTB> CDF C1 C3;
SUBC> Binomial 30 0.60.
```

In Table 3.1 we present these values.

Table 3.1 Values of the p.d.f. and c.d.f. of $B(30,.6)$

i	$b(i; 30, .6)$	$B(i; 30, .6)$
8	0.0002	0.0002
9	0.0006	0.0009
10	0.0020	0.0029
11	0.0054	0.0083
12	0.0129	0.0212
13	0.0269	0.0481
14	0.0489	0.0971
15	0.0783	0.1754
16	0.1101	0.2855
17	0.1360	0.4215
18	0.1474	0.5689
19	0.1396	0.7085
20	0.1152	0.8237
21	0.0823	0.9060
22	0.0505	0.9565
23	0.0263	0.9828
24	0.0115	0.9943
25	0.0041	0.9985
26	0.0012	0.9997
27	0.0003	1.0000

An alternative option is to use the pull-down window.

After tabulating the values of the c.d.f. we can obtain the quantiles (or fractiles) of the distribution. Recall that in the discrete case, the p-th quantile of a random variable X is

$$x_p = \text{smallest} \quad x \quad \text{such that} \quad F(x) \geq p.$$

Thus, from Table 3.1 we find that the lower quartile, the median and upper quartile of $B(30, .6)$ are $Q_1 = 16$, $M_e = 18$ and $Q_3 = 20$. These values can also be obtained directly with R code

```
> qbinom(p=0.5, size=30, prob=0.6)
```

or with MINITAB commands

```
MTB> InvCDF .5 k1;
SUBC> Binomial 30 .6.
```

The value of the median is stored in the constant $k1$.

In Figure 3.5 we present the p.d.f. of three binomial distributions, with $n = 50$ and $p = .25, .50$ and $.75$. We see that if $p = .25$, the p.d.f. is positively skewed. When $p = .5$, it is symmetric, and when $p = .75$, it is negatively skewed. This is in accordance with the index of skewness β_3, which was presented in Example 3.20.

3.3.2 The hypergeometric distribution

Let J_n denote the number of units, in a RSWOR of size n, from a population of size N, having a certain property. The number of population units before sampling having this property is M. The distribution of J_n is called the hypergeometric

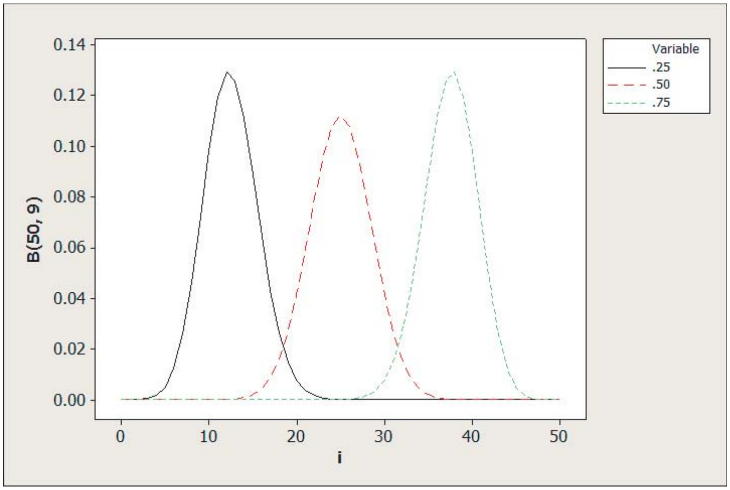

Figure 3.5 *p.d.f. of B(50, p), p = .25, .50, .75 (MINITAB)*

distribution. We denote a random variable having such a distribution by $H(N, M, n)$. The p.d.f. of J_n is

$$h(j; N, M, n) = \frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}}, \quad j = 0, \cdots, n. \tag{3.33}$$

This formula was shown already in Section 3.1.4.

The c.d.f. of $H(N, M, n)$ will be designated by $H(j; N, M, n)$. In Table 3.2 we present the p.d.f. and c.d.f. of $H(75, 15, 10)$.

In Figure 3.6 we show the p.d.f. of $H(500, 350, 100)$.

The expected value and variance of $H(N, M, n)$ are:

$$E\{J_n\} = n \cdot \frac{M}{N} \tag{3.34}$$

and

$$V\{J_n\} = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)\left(1 - \frac{n-1}{N-1}\right). \tag{3.35}$$

Notice that when $n = N$, the variance of J_n is $V\{J_N\} = 0$. Indeed, if $n = N$, $J_N = M$, which is not a random quantity. Derivations of these formulae are given in Section 5.2.2. There is no simple expression for the m.g.f.

Table 3.2 *The p.d.f. and c.d.f. of H(75, 15, 10)*

j	$h(j; 75, 15, 10)$	$H(j; 75, 15, 10)$
0	0.0910	0.0910
1	0.2675	0.3585
2	0.3241	0.6826
3	0.2120	0.8946
4	0.0824	0.9770
5	0.0198	0.9968
6	0.0029	0.9997
7	0.0003	1.0000

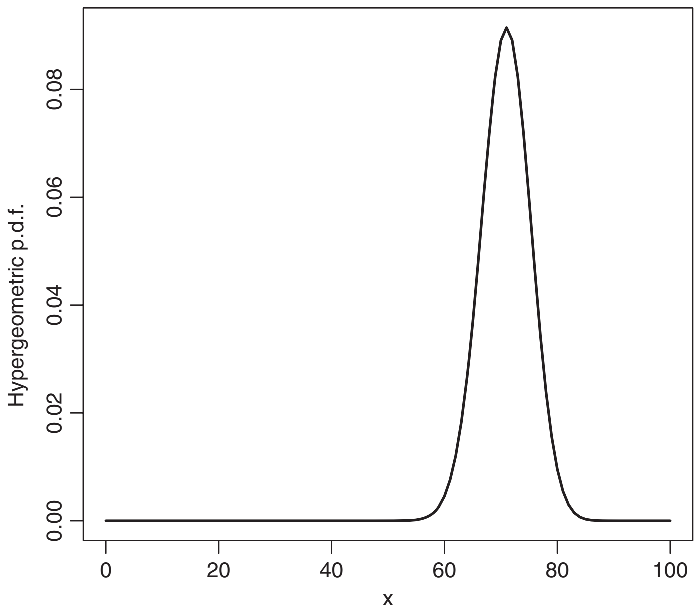

Figure 3.6 *The p.d.f. $h(i; 500, 350, 100)$*

Table 3.3 *The p.d.f. of $H(500, 350, 20)$ and $B(20, 0.7)$*

i	$h(i; 500, 350, 20)$	$b(i; 20, 0.7)$
5	0.00003	0.00004
6	0.00016	0.00022
7	0.00082	0.00102
8	0.00333	0.00386
9	0.01093	0.01202
10	0.02928	0.03082
11	0.06418	0.06537
12	0.11491	0.11440
13	0.16715	0.16426
14	0.19559	0.19164
15	0.18129	0.17886
16	0.12999	0.13042
17	0.06949	0.07160
18	0.02606	0.02785
19	0.00611	0.00684
20	0.00067	0.00080

If the sample size n is small relative to N, i.e., $n/N \ll 0.1$, the hypergeometric p.d.f. can be approximated by that of the binomial $B\left(n, \frac{M}{N}\right)$. In Table 3.3 we compare the p.d.f. of $H(500, 350, 20)$ to that of $B(20, 0.7)$.

The expected value and variance of the binomial and the hypergeometric distributions are compared in Table 3.4. We see that the expected values have the same formula, but that the variance formulae differ by the correction factor $(N - n)/(N - 1)$ which becomes 1 when $n = 1$ and 0 when $n = N$.

Table 3.4 *The expected value and variance of the hypergeometric and binomial distribution*

	$H(a;N,M,n)$ Hypergeometric	$B\left(n,\frac{M}{N}\right)$ Binomial
Expected Value	$n\dfrac{M}{N}$	$n\dfrac{M}{N}$
Variance	$n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{n-1}{N-1}\right)$	$n\frac{M}{N}\left(1-\frac{M}{N}\right)$

Example 3.23. At the end of a production day, printed circuit boards (PCB) soldered by the wave soldering process are subjected to sampling audit. A RSWOR of size n is drawn from the lot, which consists of all the PCB's produced on that day. If the sample has any defective PCB, another RSWOR of size $2n$ is drawn from the lot. If there are more than three defective boards in the combined sample, the lot is sent for rectification, in which every PCB is inspected. If the lot consists of $N = 100$ PCB's, and the number of defective ones is $M = 5$, what is the probability that the lot will be rectified, when $n = 10$?

Let J_1 be the number of defective items in the first sample. If $J_1 > 3$, then the lot is rectified without taking a second sample. If $J_1 = 1$, 2 or 3, a second sample is drawn. Thus, if R denotes the event "the lot is sent for rectification,"

$$\begin{aligned}\Pr\{R\} &= 1 - H(3;100,5,10)\\ &\quad + \sum_{i=1}^{3} h(i;100,5,10)\cdot[1 - H(3-i;90,5-i,20)\\ &= 0.00025 + 0.33939\times 0.03313\\ &\quad + 0.07022\times 0.12291\\ &\quad + 0.00638\times 0.397 = 0.0227.\end{aligned}$$

■

3.3.3 The Poisson distribution

A third discrete distribution that plays an important role in quality control is the Poisson distribution, denoted by $P(\lambda)$. It is sometimes called the distribution of rare events, since it is used as an approximation to the Binomial distribution when the sample size, n, is large and the proportion of defectives, p, is small. The parameter λ represents the "rate" at which defectives occur, that is, the expected number of defectives per time interval or per sample. The Poisson probability distribution function is given by the formula

$$p(j;\lambda) = \frac{e^{-\lambda}\lambda^j}{j!},\quad j = 0,1,2,\cdots \tag{3.36}$$

and the corresponding c.d.f. is

$$P(j;\lambda) = \sum_{i=0}^{j} p(i;\lambda),\quad j = 0,1,2,\cdots. \tag{3.37}$$

Example 3.24. Suppose that a machine produces aluminum pins for airplanes. The probability p that a single pin emerges defective is small, say, $p = .002$. In one hour, the machine makes $n = 1000$ pins (considered here to be a random sample of pins). The number of defective pins produced by the machine in one hour has a Binomial distribution with a mean of $\mu = np = 1000(.002) = 2$, so the rate of defective pins for the machine is $\lambda = 2$ pins per hour. In this case, the Binomial

Table 3.5 *Binomial distributions for np = 2 and the Poisson distribution with $\lambda = 2$*

	Binomial				Poisson
	$n = 20$	$n = 40$	$n = 100$	$n = 1000$	
k	$p = 0.1$	$p = 0.05$	$p = 0.02$	$p = .002$	$\lambda = 2$
0	0.121577	0.128512	0.132620	0.135065	0.135335
1	0.270170	0.270552	0.270652	0.270670	0.270671
2	0.285180	0.277672	0.273414	0.270942	0.270671
3	0.190120	0.185114	0.182276	0.180628	0.180447
4	0.089779	0.090122	0.090208	0.090223	0.090224
5	0.031921	0.034151	0.035347	0.036017	0.036089
6	0.008867	0.010485	0.011422	0.011970	0.012030
7	0.001970	0.002680	0.003130	0.003406	0.003437
8	0.000356	0.000582	0.000743	0.000847	0.000859
9	0.000053	0.000109	0.000155	0.000187	0.000191

probabilities are very close to the Poisson probabilities. This approximation is illustrated in Table 3.5, by considering processes which produce defective items at a rate of $\lambda = 2$ parts per hour, based on various sample sizes. In Exercise [3.46] the student is asked to prove that the binomial p.d.f. converges to that of the Poisson with mean λ when $n \to \infty$, $p \to 0$ but $np \to \lambda$.

The m.g.f. of the Poisson distribution is

$$\begin{aligned} M(t) &= e^{-\lambda} \sum_{j=0}^{\infty} e^{tj} \frac{\lambda^j}{j!} \\ &= e^{-\lambda} \cdot e^{\lambda e^t} = e^{-\lambda(1-e^t)}, \quad -\infty < t < \infty. \end{aligned} \tag{3.38}$$

Thus,

$$\begin{aligned} M'(t) &= \lambda M(t) e^t \\ M''(t) &= \lambda^2 M(t) e^{2t} + \lambda M(t) e^t \\ &= (\lambda^2 e^{2t} + \lambda e^t) M(t). \end{aligned}$$

Hence, the mean and variance of the Poisson distribution are

$$\mu = E\{X\} = \lambda$$

and (3.39)

$$\sigma^2 = V\{X\} = \lambda.$$

The Poisson distribution is used not only as an approximation to the Binomial. It is a useful model for describing the number of "events" occurring in a unit of time (or area, volume, etc.) when those events occur "at random." The rate at which these events occur is denoted by λ. An example of a Poisson random variable is the number of decaying atoms, from a radioactive substance, detected by a Geiger counter in a fixed period of time. If the rate of detection is 5 per second, then the number of atoms detected in a second has a Poisson distribution with mean $\lambda = 5$. The number detected in 5 seconds, however, will have a Poisson distribution with $\lambda = 25$. A rate of 5 per second equals a rate of 25 per 5 seconds. Other examples of Poisson random variables include:

1. The number of blemishes found in a unit area of a finished surface (ceramic plate).
2. The number of customers arriving at a store in one hour.
3. The number of defective soldering points found on a circuit board.

The p.d.f., c.d.f. and quantiles of the Poisson distribution can be computed using R, MINITAB or JMP. In Figure 3.7 we illustrate the p.d.f. for three values of λ.

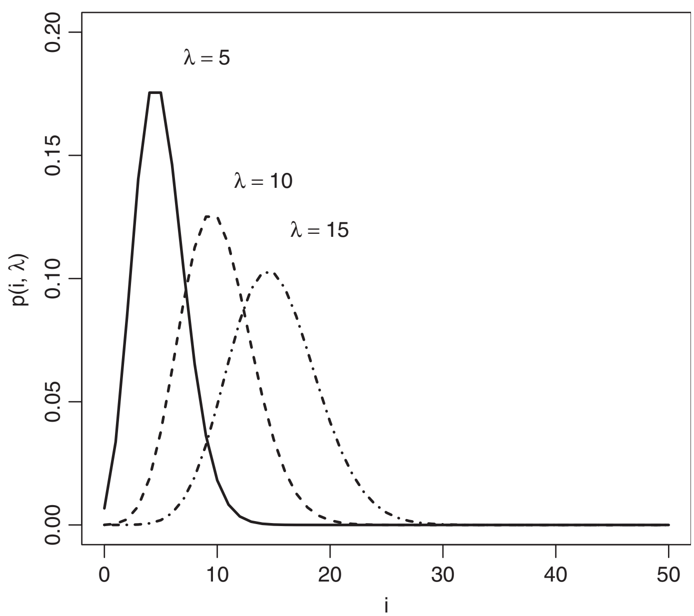

Figure 3.7 *Poisson p.d.f. $\lambda = 5, 10, 15$*

■

3.3.4 The geometric and negative binomial distributions

Consider a sequence of **independent** trials, each one having the same probability for "Success," say, p. Let N be a random variable which counts the number of trials until the first "Success" is realized, including the successful trial. N may assume positive integer values with probabilities

$$\Pr\{N = n\} = p(1-p)^{n-1}, \quad n = 1, 2, \cdots. \tag{3.40}$$

This probability function is the p.d.f. of the **geometric** distribution.

Let $g(n; p)$ designate the p.d.f. The corresponding c.d.f. is

$$G(n; p) = 1 - (1-p)^n, \quad n = 1, 2, \cdots.$$

From this we obtain that the α-quantile $(0 < \alpha < 1)$ is given by

$$N_\alpha = \left[\frac{\log\ (1-\alpha)}{\log\ (1-p)}\right] + 1,$$

where $[x]$ designates the integer part of x.

The expected value and variance of the geometric distribution are

$$E\{N\} = \frac{1}{p},$$

and (3.41)

$$V\{N\} = \frac{1-p}{p^2}.$$

Indeed, the m.g.f. of the geometric distribution is

$$\begin{aligned} M(t) &= pe^t \sum_{j=0}^{\infty} (e^t(1-p))^j \\ &= \frac{pe^t}{1-e^t(1-p)}, \quad \text{if } t < -\log\ (1-p). \end{aligned} \tag{3.42}$$

Thus, for $t < -\log\ (1-p)$,

$$M'(t) = \frac{pe^t}{(1-e^t(1-p))^2}$$

and

$$M''(t) = \frac{pe^t}{(1-e^t(1-p))^2} + \frac{2p(1-p)e^{2t}}{(1-e^t(1-p))^3}.$$

Hence

$$\begin{aligned} \mu_1 &= M'(0) = \frac{1}{p} \\ \mu_2 &= M''(0) = \frac{2-p}{p^2}, \end{aligned} \tag{3.43}$$

and the above formulae of $E\{X\}$ and $V\{X\}$ are obtained.

The geometric distribution is applicable in many problems. We illustrate one such application in the following example.

Example 3.25. An insertion machine stops automatically if there is a failure in handling a component during an insertion cycle. A cycle starts immediately after the insertion of a component and ends at the insertion of the next component. Suppose that the probability of stopping is $p = 10^{-3}$ per cycle. Let N be the number of cycles until the machine stops. It is assumed that events at different cycles are mutually independent. Thus N has a geometric distribution and $E\{N\} = 1,000$. We expect a run of 1,000 cycles between consecutive stopping. The number of cycles, N, however, is a random variable with standard deviation of $\sigma = \left(\frac{1-p}{p^2}\right)^{1/2} = 999.5$. This high value of σ indicates that we may see very short runs and also long ones. Indeed, for $\alpha = 0.5$ the quantiles of N are, $N_{.05} = 52$ and $N_{.95} = 2995$. ■

The number of failures until the first success, $N - 1$, has a shifted geometric distribution, which is a special case of the family of **Negative-Binomial** distribution.

We say that a non-negative integer valued random variable X has a negative-binomial distribution, with parameters (p, k), where $0 < p < 1$ and $k = 1, 2, \cdots$, if its p.d.f. is

$$g(j; p, k) = \binom{j+k-1}{k-1} p^k (1-p)^j, \tag{3.44}$$

$j = 0, 1, \cdots$. The shifted geometric distribution is the special case of $k = 1$.

A more general version of the negative-binomial distribution can be formulated, in which $k - 1$ is replaced by a positive real parameter. A random variable having the above negative-binomial will be designated by $NB(p, k)$. The $NB(p, k)$

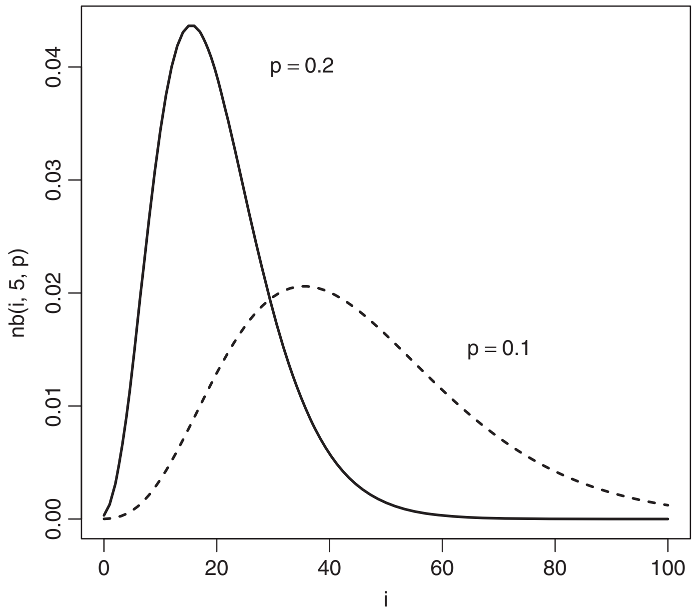

Figure 3.8 *p.d.f. of NB(p, 5) with p = 0.10, 0.20*

represents the number of failures observed until the k-th success. The expected value and variance of $NB(p,k)$ are:

$$E\{X\} = k\frac{1-p}{p},$$

and (3.45)

$$V\{X\} = k\frac{1-p}{p^2}.$$

In Figure 3.8 we present the p.d.f. of $NB(p,k)$. The negative binomial distributions have been applied as a model of the distribution for the periodic demand of parts in inventory theory.

3.4 Continuous distributions

3.4.1 The uniform distribution on the interval (a,b), $a < b$

We denote a random variable having this distribution by $U(a,b)$. The p.d.f. is given by

$$f(x;a,b) = \begin{cases} 1/(b-a), & a \le x \le b \\ \\ 0, & \text{elsewhere}, \end{cases} \tag{3.46}$$

and the c.d.f. is

$$F(x;a,b) = \begin{cases} 0, & \text{if } x < a \\ (x-a)/(b-a), & \text{if } a \le x < b \\ 1, & \text{if } b \le x \end{cases} \tag{3.47}$$

The expected value and variance of $U(a,b)$ are

$$\mu = (a+b)/2,$$

and (3.48)

$$\sigma^2 = (b-a)^2/12.$$

The p-th fractile is $x_p = a + p(b-a)$.

To verify the formula for μ, we set

$$\begin{aligned}\mu &= \frac{1}{b-a}\int_a^b x dx = \frac{1}{b-a}\left|\frac{1}{2}x^2\right|_a^b = \frac{1}{2(b-a)}(b^2-a^2)\\ &= \frac{a+b}{2}.\end{aligned}$$

Similarly,

$$\begin{aligned}\mu_2 &= \frac{1}{b-a}\int_a^b x^2 dx = \frac{1}{b-a}\left|\frac{1}{3}x^3\right|_a^b\\ &= \frac{1}{3(b-a)}(b^3-a^3) = \frac{1}{3}(a^2+ab+b^2).\end{aligned}$$

Thus,

$$\begin{aligned}\sigma^2 = \mu_2 - \mu_1^2 &= \frac{1}{3}(a^2+ab+b^2) - \frac{1}{4}(a^2+2ab+b^2)\\ &= \frac{1}{12}(4a^2+4ab+4b^2-3a^2-6ab-3b^2)\\ &= \frac{1}{12}(b-a)^2.\end{aligned}$$

We can get these moments also from the m.g.f., which is

$$M(t) = \frac{1}{t(b-a)}(e^{tb}-e^{ta}), \quad -\infty < t < \infty.$$

Moreover, for values of t close to 0

$$M(t) = 1 + \frac{1}{2}t(b+a) + \frac{1}{6}t^2(b^2+ab+a^2) + \cdots.$$

3.4.2 The normal and log-normal distributions

3.4.2.1 The normal distribution

The Normal or Gaussian distribution denoted by $N(\mu,\sigma)$, occupies a central role in statistical theory. Its density function (p.d.f.) is given by the formula

$$n(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}. \tag{3.49}$$

This p.d.f. is symmetric around the location parameter, μ. σ is a scale parameter. The m.g.f. of $N(0,1)$ is

$$\begin{aligned}M(t) &= \frac{1}{\sqrt{2\pi}}e^{tx-\frac{1}{2}x^2}dx\\ &= \frac{e^{t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx+t^2)}dx\\ &= e^{t^2/2}.\end{aligned} \tag{3.50}$$

Indeed, $\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-t)^2\right\}$ is the p.d.f. of $N(t, 1)$. Furthermore,

$$\begin{aligned}
M'(t) &= tM(t) \\
M''(t) &= t^2M(t) + M(t) = (1+t^2)M(t) \\
M'''(t) &= (t+t^3)M(t) + 2tM(t) \\
&= (3t+t^3)M(t) \\
M^{(4)}(t) &= (3+6t^2+t^4)M(t).
\end{aligned}$$

Thus, by substituting $t = 0$ we obtain that

$$\begin{aligned}
E\{N(0,1)\} &= 0, \\
V\{N(0,1)\} &= 1, \\
\mu_3^* &= 0, \\
\mu_4^* &= 3.
\end{aligned} \tag{3.51}$$

To obtain the moments in the general case of $N(\mu, \sigma^2)$, we write $X = \mu + \sigma N(0, 1)$. Then

$$\begin{aligned}
E\{X\} &= E\{\mu + \sigma N(0,1)\} \\
&= \mu + \sigma E\{N(0,1)\} = \mu \\
V\{X\} &= E\{(X-\mu)^2\} = \sigma^2 E\{N^2(0,1)\} = \sigma^2 \\
\mu_3^* &= E\{(X-\mu)^3\} = \sigma^3 E\{N^3(0,1)\} = 0 \\
\mu_4^* &= E\{(X-\mu)^4\} = \sigma^4 E\{N^4(0,1)\} = 3\sigma^4.
\end{aligned}$$

Thus, the index of kurtosis in the normal case is $\beta_4 = 3$.

The graph of the p.d.f. $n(x; \mu, \sigma)$ is a symmetric bell-shaped curve that is centered at μ (shown in Figure 3.9). The spread of the density is determined by the variance σ^2 in the sense that most of the area under the curve (in fact, 99.7% of the area) lies between $\mu - 3\sigma$ and $\mu + 3\sigma$. Thus, if X has a normal distribution with mean $\mu = 25$ and standard deviation $\sigma = 2$, the probability is .997 that the observed value of X will fall between 19 and 31.

Areas (that is, probabilities) under the normal p.d.f. are found in practice using a table or appropriate software like MINITAB. Since it is not practical to have a table for each pair of parameters μ and σ, we use the standardized form of the normal random variable. A random variable Z is said to have a **standard normal distribution** if it has a normal distribution with mean zero and variance one. The standard normal density function is $\phi(x) = n(x; 0, 1)$ and the standard cumulative distribution function is denoted by $\Phi(x)$. This function is also called the **standard normal** integral, that is,

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \tag{3.52}$$

The c.d.f., $\Phi(x)$, represents the area over the x-axis under the standard normal p.d.f. to the left of the value x (see Figure 3.10).

If we wish to determine the probability that a standard normal random variable is less than 1.5, for example, we use R code

```
> pnorm(q=1.5, mean=0, sd=1)
```

or the MINITAB commands

```
MTB> CDF 1.5;
SUBC> NORMAL 0 1.
```

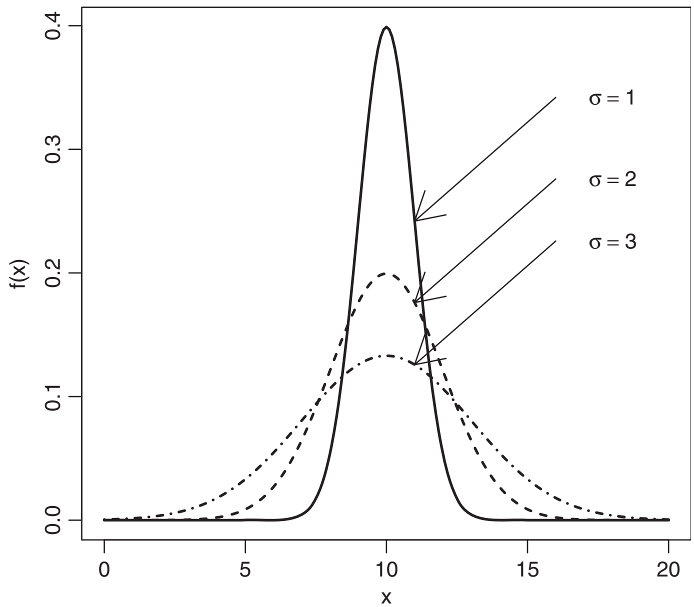

Figure 3.9 *The p.d.f. of* $N(\mu, \sigma)$, $\mu = 10$, $\sigma = 1, 2, 3$

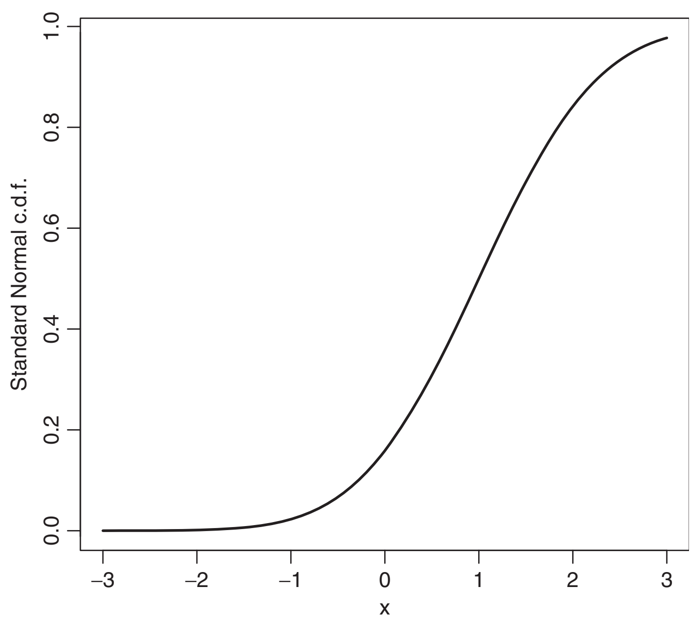

Figure 3.10 *Standard normal c.d.f.*

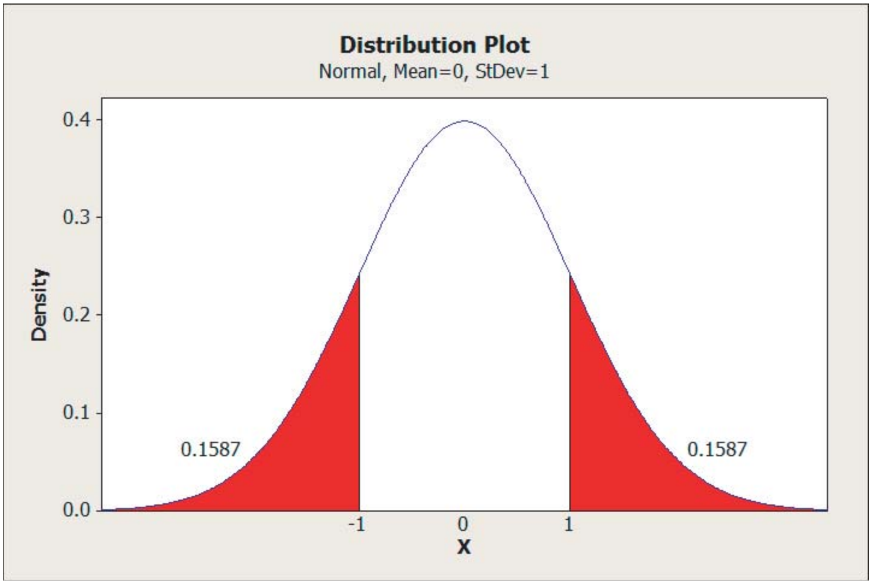

Figure 3.11 *The symmetry of the normal distribution (MINITAB)*

We find that $\Pr\{Z \le 1.5\} = \Phi(1.5) = 0.9332$. To obtain the probability that Z lies between .5 and 1.5 we first find the probability that Z is less than 1.5, then subtract from this number the probability that Z is less than .5. This yields

$$\begin{aligned}\Pr\{.5 < Z < 1.5\} &= \Pr\{Z < 1.5\} - \Pr\{Z < .5\}\\ &= \Phi(1.5) - \Phi(.5) = .9332 - .6915 = .2417.\end{aligned}$$

Many tables of the normal distribution do not list values of $\Phi(x)$ for $x < 0$. This is because the normal density is symmetric about $x = 0$, and we have the relation

$$\Phi(-x) = 1 - \Phi(x), \quad \text{for all} \quad x. \tag{3.53}$$

Thus, to compute the probability that Z is less than -1, for example, we write

$$\Pr\{Z < -1\} = \Phi(-1) = 1 - \Phi(1) = 1 - .8413 = .1587 \text{ (see Figure 3.11).}$$

The **p-th quantile (percentile of fractile)** of the standard normal distribution is the number z_p that satisfies the statement

$$\Phi(z_p) = \Pr\{Z \le z_p\} = p. \tag{3.54}$$

If X has a normal distribution with mean μ and standard deviation σ, we denote the p-th fractile of the distribution by x_p. We can show that x_p is related to the standard normal quantile by

$$x_p = \mu + z_p\sigma.$$

The p-th fractile of the normal distribution can be obtained by using R code

```
> qnorm(p=0.95, mean=0, sd=1)
```

or the MINITAB command

```
MTB> InvCDF 0.95;
SUBC> Normal 0.0 1.0.
```

In this command we used $p = 0.95$. The printed result is $z_{.95} = 1.6449$. We can use any value of μ and σ in the subcommand. Thus, for $\mu = 10$ and $\sigma = 1.5$

$$x_{.95} = 10 + z_{.95} \times \sigma = 12.4673.$$

Now suppose that X is a random variable having a normal distribution with mean μ and variance σ^2. That is, X has a $N(\mu, \sigma)$ distribution. We define the **standardized form** of X as

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

By subtracting the mean from X and then dividing by the standard deviation, we transform X to a standard normal random variable. (That is, Z has expected value zero and standard deviation one.) This will allow us to use the standard normal table to compute probabilities involving X. Thus, to compute the probability that X is less than a, we write

$$\begin{aligned}\Pr\{X \leq a\} &= \Pr\left\{\frac{X - \mu}{\sigma} < \frac{a - \mu}{\sigma}\right\} \\ &= \Pr\left\{Z < \frac{a - \mu}{\sigma}\right\} = \Phi\left(\frac{a - \mu}{\sigma}\right).\end{aligned}$$

Example 3.26. Let X represent the length (with cap) of a randomly selected aluminum pin. Suppose we know that X has a normal distribution with mean $\mu = 60.02$ and standard deviation $\sigma = 0.048$ [mm]. What is the probability that the length with cap of a randomly selected pin will be less than 60.1 [mm]? Using R

```
> pnorm(q=60.1, mean=60.02, sd=0.048, lower.tail=TRUE)
```

or the MINITAB command

```
MTB> CDF 60.1;
SUBC> Normal 60.02 0.048.
```

we obtain $\Pr\{X \leq 60.1\} = 0.9522$. If we have to use the table of $\Phi(Z)$, we write

$$\begin{aligned}\Pr\{X \leq 60.1\} &= \Phi\left(\frac{60.1 - 60.02}{0.048}\right) \\ &= \Phi(1.667) = 0.9522.\end{aligned}$$

Continuing with the example, consider the following question: If a pin is considered "acceptable" when its length is between 59.9 and 60.1 mm, what proportion of pins is expected to be rejected? To answer this question, we first compute the probability of accepting a single pin. This is the probability that X lies between 59.9 and 60.1, that is,

$$\begin{aligned}\Pr\{50.9 < X < 60.1\} &= \Phi\left(\frac{60.1 - 60.02}{0.048}\right) - \Phi\left(\frac{59.9 - 60.02}{0.048}\right) \\ &= \Phi(1.667) - \Phi(-2.5) \\ &= .9522 - 0.0062 = 0.946.\end{aligned}$$

Thus, we expect that 94.6% of the pins will be accepted, and that 5.4% of them will be rejected. ■

3.4.2.2 *The log-normal distribution*

A random variable X is said to have a **log-normal distribution**, $LN\ (\mu, \sigma^2)$, if $Y = \log\ X$ has the normal distribution $N(\mu, \sigma^2)$.

The log-normal distribution has been applied to model distributions of strength variables, like the tensile strength of fibers (see Chapter 2), the compressive strength of concrete cubes, etc. It has also been used for random quantities of pollutants in water or air, and other phenomena with skewed distributions.

The p.d.f. of $LN\ (\mu, \sigma)$ is given by the formula

$$f(x; \mu, \sigma^2) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\dfrac{1}{2\sigma^2}(\ln\ x - \mu)^2\right\}, & 0 < x < \infty \\ 0, & x \leq 0. \end{cases} \tag{3.55}$$

The c.d.f. is expressed in terms of the standard normal integral as

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \Phi\left(\dfrac{\ln\ x - \mu}{\sigma}\right), & 0 < x < \infty. \end{cases} \tag{3.56}$$

The expected value and variance of $LN\ (\mu, \sigma)$ are

$$E\{X\} = e^{\mu + \sigma^2/2}$$

and (3.57)

$$V\{X\} = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$

One can show that the third central moment of $LN\ (\mu, \sigma^2)$ is

$$\mu_3^* = e^{3\mu + \frac{3}{2}\sigma^2}(e^{3\sigma^2} - 3e^{\sigma^2} + 2).$$

Hence, the **index of skewness** of this distribution is

$$\beta_3 = \frac{\mu_3^*}{\sigma^3} = \frac{e^{3\sigma^2} - 3e^{\sigma^2} + 2}{(e^{\sigma^2} - 1)^{3/2}}. \tag{3.58}$$

It is interesting that the index of skewness does not depend on μ, and is positive for all $\sigma^2 > 0$. This index of skewness grows very fast as σ^2 increases. This is shown in Figure 3.12.

3.4.3 The exponential distribution

We designate this distribution by $E(\beta)$. The p.d.f. of $E(\beta)$ is given by the formula

$$f(x; \beta) = \begin{cases} 0, & \text{if } x < 0 \\ (1/\beta)e^{-x/\beta}, & \text{if } x \geq 0, \end{cases} \tag{3.59}$$

where β is a positive parameter, that is, $0 < \beta < \infty$. In Figure 3.13 we present these p.d.f.'s for various values of β.

The corresponding c.d.f. is

$$F(x; \beta) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-x/\beta}, & \text{if } x \geq 0. \end{cases} \tag{3.60}$$

The expected value and the variance of $E(\beta)$ are

$$\mu = \beta,$$

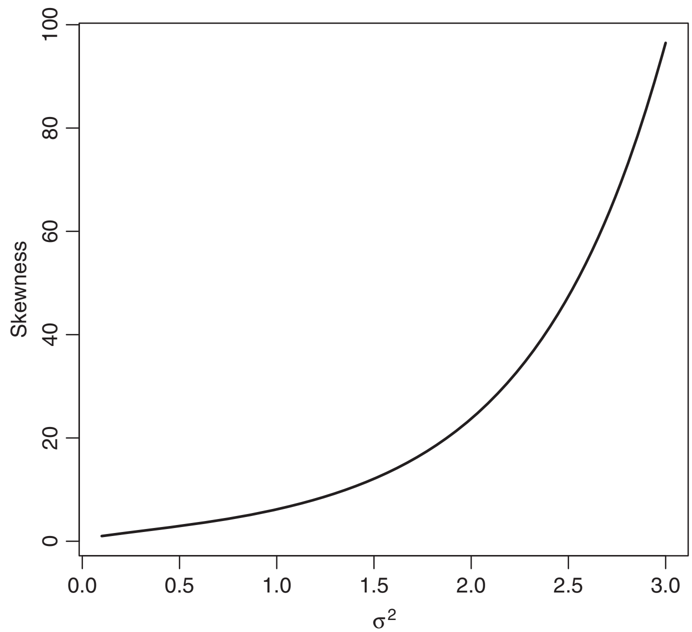

Figure 3.12 *The index of skewness of LN* (μ, σ)

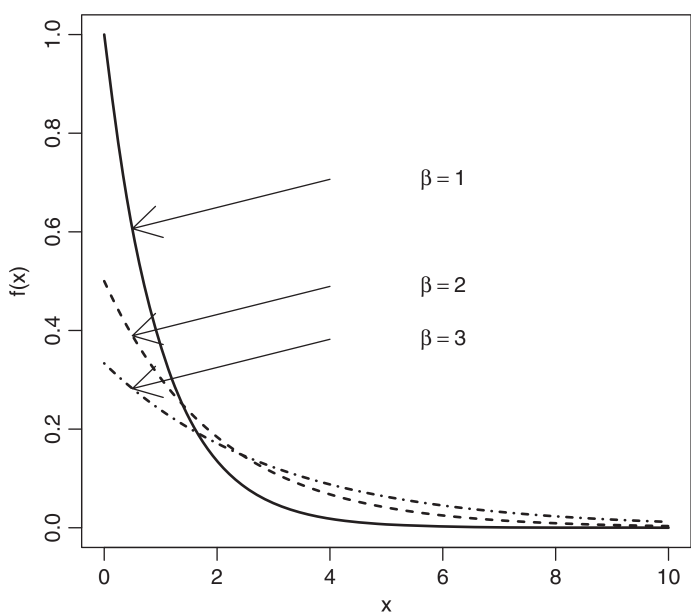

Figure 3.13 *The p.d.f. of E(*β*),* $\beta = 1, 2, 3$

and

$$\sigma^2 = \beta^2.$$

Indeed,

$$\mu = \frac{1}{\beta}\int_0^\infty xe^{-x/\beta}dx.$$

Making the change of variable to $y = x/\beta$, $dx = \beta dy$, we obtain

$$\begin{aligned}\mu &= \beta \int_0^\infty ye^{-y}dy \\ &= \beta.\end{aligned}$$

Similarly

$$\begin{aligned}\mu_2 &= \frac{1}{\beta}\int_0^\infty x^2e^{-x/\beta}dx = \beta^2\int_0^\infty y^2e^{-y}dy \\ &= 2\beta^2.\end{aligned}$$

Hence,

$$\sigma^2 = \beta^2.$$

The p-th quantile is $x_p = -\beta \ln\ (1 - p)$.

The exponential distribution is related to the Poisson model in the following way: If the number of events occurring in a period of time follows a Poisson distribution with rate λ, then the time between occurrences of events has an exponential distribution with parameter $\beta = 1/\lambda$. The exponential model can also be used to describe the lifetime (i.e. time to failure) of certain electronic systems. For example, if the mean life of a system is 200 hours, then the probability that it will work at least 300 hours without failure is

$$\begin{aligned}\Pr\{X \geq 300\} &= 1 - \Pr\{X < 300\} \\ &= 1 - F(300) = 1 - (1 - e^{-300/200}) = 0.223.\end{aligned}$$

The exponential distribution is positively skewed, and its index of skewness is

$$\beta_3 = \frac{\mu_3^*}{\sigma^3} = 2,$$

irrespective of the value of β. We have seen before that the kurtosis index is $\beta_4 = 9$.

3.4.4 The gamma and Weibull distributions

Two important distributions for studying the reliability and failure rates of systems are the gamma and the Weibull distributions. We will need these distributions in our study of reliability methods (Chapter 14). These distributions are discussed here as further examples of continuous distributions.

Suppose we use in a manufacturing process a machine which mass-produces a particular part. In a random manner, it produces defective parts at a rate of λ per hour. The number of defective parts produced by this machine in a time period $[0, t]$ is a random variable $X(t)$ having a Poisson distribution with mean λt, that is,

$$\Pr\{X(t) = j\} = (\lambda t)^j e^{-(\lambda t)}/j!, \quad j = 0, 1, 2, \cdots. \tag{3.61}$$

Suppose we wish to study the distribution of the time until the k-th defective part is produced. Call this continuous random variable Y_k. We use the fact that the k-th defect will occur before time t (i.e., $Y_k \leq t$) if and only if at least k defects occur up to time t (i.e., $X(t) \geq k$). Thus the c.d.f. for Y_k is

$$\begin{aligned} G(t; k, \lambda) &= \Pr\{Y_k \leq t\} \\ &= \Pr\{X(t) \geq k\} \\ &= 1 - \sum_{j=0}^{k-1} (\lambda t)^j e^{-\lambda t}/j!. \end{aligned} \tag{3.62}$$

The corresponding p.d.f. for Y_k is

$$g(t; k, \lambda) = \frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t}, \quad \text{for} \quad t \geq 0. \tag{3.63}$$

This p.d.f. is a member of a general family of distributions which depend on two parameters, ν and β, and are called the **gamma** distributions $G(\nu, \beta)$. The p.d.f. of a gamma distribution $G(\nu, \beta)$ is

$$g(x; \nu, \beta) = \begin{cases} \frac{1}{\beta^\nu \Gamma(\nu)} x^{\nu-1} e^{-x/\beta}, & x \geq 0, \\ \\ 0, & x < 0 \end{cases} \tag{3.64}$$

In R function `pgamma` computes c.d.f of a gamma distribution having ν = shape and β = scale

```
> pgamma(q=1, shape=1, scale=1)

[1] 0.6321206
```

where $0 < \nu, \beta < \infty$, $\Gamma(\nu)$ is called the **gamma function** of ν and is defined as the integral

$$\Gamma(\nu) = \int_0^\infty x^{\nu-1} e^{-x} dx, \quad \nu > 0. \tag{3.65}$$

The gamma function satisfies the relationship

$$\Gamma(\nu) = (\nu - 1)\Gamma(\nu - 1), \quad \text{for all } \nu > 1. \tag{3.66}$$

Hence, for every positive integer k, $\Gamma(k) = (k-1)!$. Also, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. We note also that the exponential distribution, $E(\beta)$, is a special case of the gamma distribution with $\nu = 1$. Some gamma p.d.f.'s are presented in Figure 3.14. The value of $\Gamma(\nu)$ can be computed in R by the following commands which compute $\Gamma(5)$. Generally, replace 5 in line 2 by ν.

```
> gamma(5)

[1] 24
```

The expected value and variance of the gamma distribution $G(\nu, \beta)$ are, respectively,

$$\mu = \nu\beta,$$

and (3.67)

$$\sigma^2 = \nu\beta^2.$$

To verify these formulae we write

$$\begin{aligned} \mu &= \frac{1}{\beta^\nu \Gamma(\nu)} \int_0^\infty x \cdot x^{\nu-1} e^{-x/\beta} dx \\ &= \frac{\beta^{\nu+1}}{\beta^\nu \Gamma(\nu)} \int_0^\infty y^\nu e^{-y} dy \\ &= \beta \frac{\Gamma(\nu + 1)}{\Gamma(\nu)} = \nu\beta. \end{aligned}$$

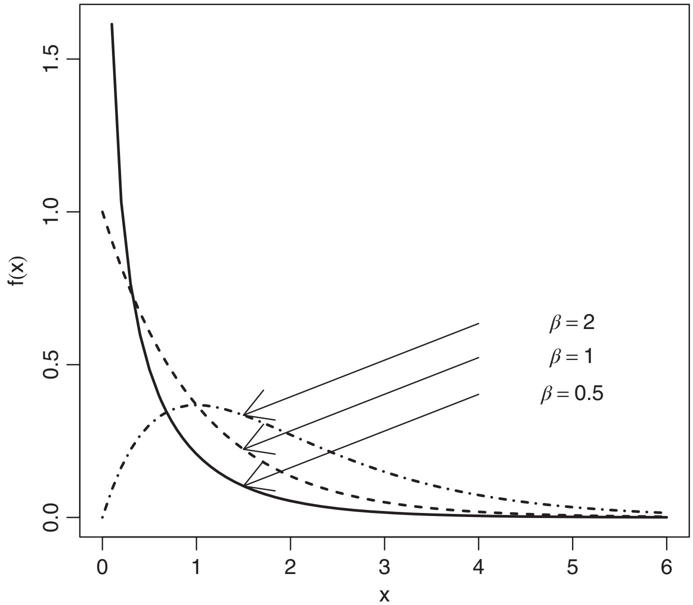

Figure 3.14 *The gamma densities, with $\beta = 1$ and $\nu = .5,\ 1,\ 2$*

Similarly,

$$\mu_2 = \frac{1}{\beta^\nu \Gamma(\nu)} \int_0^\infty x^2 \cdot x^{\nu-1} e^{-x/\beta} dx$$
$$= \frac{\beta^{\nu+2}}{\beta^\nu \Gamma(\nu)} \int_0^\infty y^{\nu+1} e^{-y} dy$$
$$= \beta^2 \frac{\Gamma(\nu+2)}{\Gamma(\nu)} = (\nu+1)\nu\beta^2.$$

Hence,

$$\sigma^2 = \mu_2 - \mu_1^2 = \nu\beta^2.$$

An alternative way is to differentiate the m.g.f.

$$M(t) = (1 - t\beta)^{-\nu}, \quad t < \frac{1}{\beta}. \tag{3.68}$$

Weibull distributions are often used in reliability models in which the system either "ages" with time or becomes "younger" (see Chapter 14). The Weibull family of distributions will be denoted by $W(\alpha, \beta)$. The parameters α and β, $\alpha, \beta > 0$, are called the shape and the scale parameters, respectively. The p.d.f. of $W(\alpha, \beta)$ is given by

$$w(t; \alpha, \beta) = \begin{cases} \frac{\alpha t^{\alpha-1}}{\beta^\alpha} e^{-(t/\beta)^\alpha}, & t \geq 0, \\ 0, & t < 0. \end{cases} \tag{3.69}$$

The corresponding c.d.f. is

$$W(t; \alpha, \beta) = \begin{cases} 1 - e^{-(t/\beta)^\alpha}, & t \geq 0 \\ 0, & t < 0. \end{cases} \tag{3.70}$$

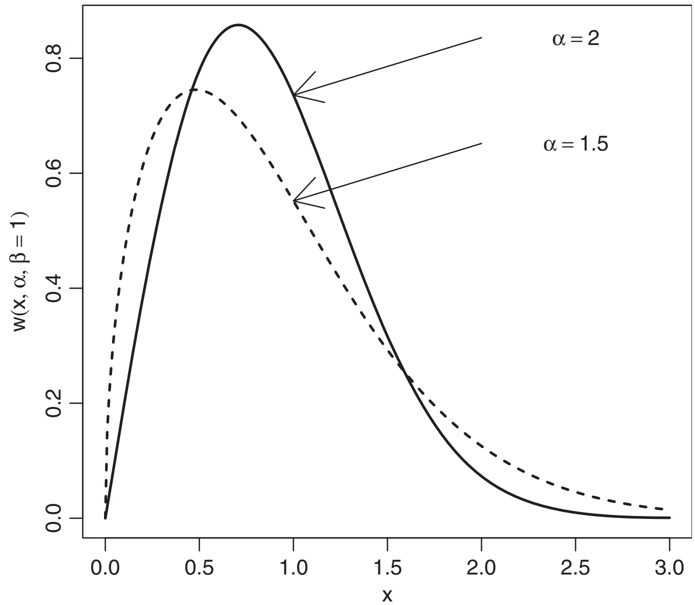

Figure 3.15 *Weibull density functions, $\alpha = 1.5, 2$*

Notice that $W(1, \beta) = E(\beta)$. The mean and variance of this distribution are

$$\mu = \beta \cdot \Gamma\left(1 + \frac{1}{\alpha}\right) \tag{3.71}$$

and

$$\sigma^2 = \beta^2 \left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right\} \tag{3.72}$$

respectively. The values of $\Gamma(1 + (1/\alpha))$ and $\Gamma(1 + (2/\alpha))$ can be computed by R, MINITAB or JMP. If, for example, $\alpha = 2$ then

$$\mu = \beta\sqrt{\pi}/2 = .8862\beta$$
$$\sigma^2 = \beta^2(1 - \pi/4) = .2145\beta^2,$$

since

$$\Gamma\left(1 + \frac{1}{2}\right) = \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi},$$

and

$$\Gamma\left(1 + \frac{2}{2}\right) = \Gamma(2) = 1.$$

In Figure 3.15 we present three p.d.f. of $W(\alpha, \beta)$ for $\alpha = 1.5, 2.0$ and $\beta = 1$.

3.4.5 The Beta distributions

Distributions having p.d.f. of the form

$$f(x; \nu_1, \nu_2) = \begin{cases} \frac{1}{B(\nu_1, \nu_2)} x^{\nu_1 - 1}(1 - x)^{\nu_2 - 1}, & 0 < x < 1, \\ 0, & \text{otherwise} \end{cases} \tag{3.73}$$

where, for ν_1, ν_2 positive,

$$B(\nu_1, \nu_2) = \int_0^1 x^{\nu_1-1}(1-x)^{\nu_2-1}dx \tag{3.74}$$

are called Beta distributions. The function $B(\nu_1, \nu_2)$ is called the Beta integral. One can prove that

$$B(\nu_1, \nu_2) = \frac{\Gamma(\nu_1)\Gamma(\nu_2)}{\Gamma(\nu_1+\nu_2)}. \tag{3.75}$$

The parameters ν_1 and ν_2 are shape parameters. Notice that when $\nu_1 = 1$ and $\nu_2 = 1$, the Beta reduces to $U(0, 1)$. We designate distributions of this family by Beta(ν_1, ν_2). The c.d.f. of Beta(ν_1, ν_2) is denoted also by $I_x(\nu_1, \nu_2)$, which is known as the **incomplete beta function ratio**, that is,

$$I_x(\nu_1, \nu_2) = \frac{1}{B(\nu_1, \nu_2)} \int_0^x u^{\nu_1-1}(1-u)^{\nu_2-1}du, \tag{3.76}$$

for $0 \le x \le 1$. Notice that $I_x(\nu_1, \nu_2) = 1 - I_{1-x}(\nu_2, \nu_1)$. The density functions of the p.d.f. Beta(2.5, 5.0) and Beta(2.5, 2.5) are plotted in Figure 3.16. Notice that if $\nu_1 = \nu_2$ then the p.d.f. is symmetric around $\mu = \frac{1}{2}$. There is no simple formula for the m.g.f. of Beta(ν_1, ν_2). However, the m-th moment is equal to

$$\begin{aligned}\mu_m &= \frac{1}{\beta(\nu_1, \nu_2)} \int_0^1 u^{m+\nu_1-1}(1-u)^{\nu_2-1}du \\ &= \frac{B(\nu_1+m, \nu_2)}{B(\nu_1, \nu_2)} \\ &= \frac{\nu_1(\nu_1+1)\cdots(\nu_1+m-1)}{(\nu_1+\nu_2)(\nu_1+\nu_2+1)\cdots(\nu_1+\nu_2+m-1)}.\end{aligned} \tag{3.77}$$

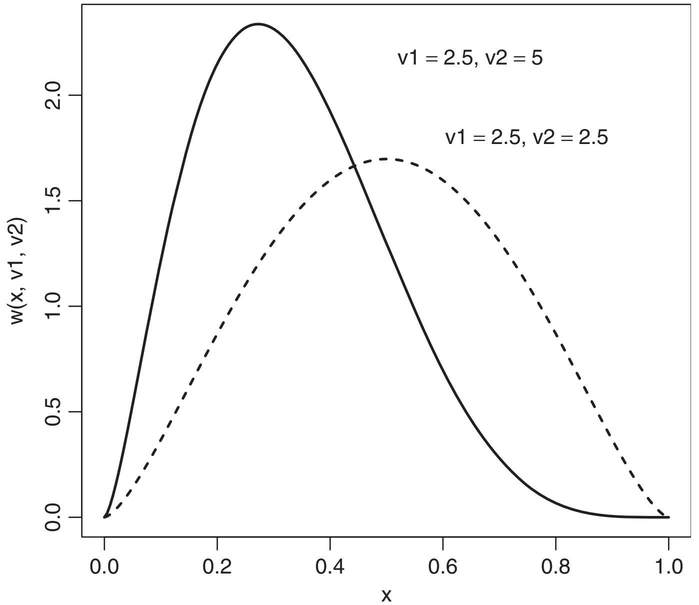

Figure 3.16 *Beta densities, $\nu_1 = 2.5$, $\nu_2 = 2.5$; $\nu_1 = 2.5$, $\nu_2 = 5.00$*

Hence,

$$E\{\mathrm{Beta}(\nu_1,\nu_2)\} = \frac{\nu_1}{\nu_1+\nu_2}$$
$$V\{\mathrm{Beta}(\nu_1,\nu_2)\} = \frac{\nu_1\nu_2}{(\nu_1+\nu_2)^2(\nu_1+\nu_2+1)}. \tag{3.78}$$

The beta distribution has an important role in the theory of statistics. As will be seen later, many methods of statistical inference are based on the order statistics (see Section 3.7). The distribution of the order statistics is related to the beta distribution. Moreover, since the beta distribution can have a variety of shapes, it has been applied in many cases in which the variable has a distribution on a finite domain. By introducing a location and a scale parameter, one can fit a shifted-scaled beta distribution to various frequency distributions.

3.5 Joint, marginal and conditional distributions

3.5.1 Joint and marginal distributions

Let $X_1, \cdots, X_k$ be random variables which are jointly observed at the same experiments. In Chapter 3 we presented various examples of bivariate and multivariate frequency distributions. In the present section we present only the fundamentals of the theory, mainly for future reference. We make the presentation here, focusing on continuous random variables. The theory holds generally for discrete or for a mixture of continuous and discrete random variables.

A function $F(x_1, \ldots, x_k)$ is called the joint c.d.f. of $X_1, \ldots, X_k$ if

$$F(x_1, \ldots, x_k) = \Pr\{X_1 \le x_1, \ldots, X_k \le x_k\} \tag{3.79}$$

for all $(x_1, \ldots, x_k) \in \mathbb{R}^k$ (the Euclidean k-space). By letting one or more variables tend to infinity, we obtain the joint c.d.f. of the remaining variables. For example,

$$\begin{aligned} F(x_1, \infty) &= \Pr\{X_1 \le x_1, X_2 \le \infty\} \\ &= \Pr\{X_1 \le x_1\} = F_1(x_1). \end{aligned} \tag{3.80}$$

The c.d.f.'s of the individual variables, are called the **marginal** distributions. $F_1(x_1)$ is the marginal c.d.f. of X_1.

A non-negative function $f(x_1, \cdots, x_k)$ is called the **joint p.d.f**. of $X_1, \cdots, X_k$, if

(i)
$$f(x_1, \cdots, x_k) \ge 0 \quad \text{for all} \quad (x_1, \cdots, x_k), \quad \text{where} \quad -\infty < x_i < \infty \qquad (i = 1, \cdots, k)$$

(ii)
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_k) dx_1, \cdots, dx_k = 1.$$

and

(iii)
$$F(x_1, \ldots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(y_1, \ldots, y_k) dy_1 \ldots dy_k.$$

The **marginal p.d.f**. of X_i $(i = 1, \cdots, k)$ can be obtained from the joint p.d.f. $f(x_1, \cdots, x_k)$, by integrating the joint p.d.f. with respect to all $x_j, j \ne i$. For example, if $k = 2, f(x_1, x_2)$ is the joint p.d.f. of X_1, X_2. The marginal p.d.f. of X_1 is

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2.$$

Similarly, the marginal p.d.f. of X_2 is

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1.$$

Indeed, the marginal c.d.f. of X_i is

$$F(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2.$$

Differentiating $F(x_1)$ with respect to x_1 we obtain the marginal p.d.f. of X_1, i.e.,

$$\begin{aligned} f(x_1) &= \frac{d}{dx_1} \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 \\ &= \int_{-\infty}^{\infty} f(x_1, y_2) dy_2. \end{aligned}$$

If $k = 3$, we can obtain the marginal joint p.d.f. of a pair of random variables by integrating with respect to the third variable. For example, the joint marginal p.d.f. of (X_1, X_2), can be obtained from that of (X_1, X_2, X_3) as

$$f_{1,2}(x_1, x_2) = \int_{-\infty}^{\infty} f(x_1, x_2, x_3) dx_3.$$

Similarly,

$$f_{1,3}(x_1, x_3) = \int_{-\infty}^{\infty} f(x_1, x_2, x_3) dx_2,$$

and

$$f_{2,3}(x_2, x_3) = \int_{-\infty}^{\infty} f(x_1, x_2, x_3) dx_1.$$

Example 3.27. The present example is theoretical and is designed to illustrate the above concepts.

Let (X, Y) be a pair of random variables having a joint uniform distribution on the region

$$T = \{(x, y) : 0 \le x, y, \quad x + y \le 1\}.$$

T is a triangle in the (x, y)-plane with vertices at $(0, 0)$, $(1, 0)$ and $(0, 1)$. According to the assumption of uniform distribution the joint p.d.f. of (X, Y) is

$$f(x, y) = \begin{cases} 2, & \text{if } (x, y) \in T \\ 0, & \text{otherwise.} \end{cases}$$

The marginal p.d.f. of X is obtained as

$$f_1(x) = 2 \int_0^{1-x} dy = 2(1 - x), \quad 0 \le x \le 1.$$

Obviously, $f_1(x) = 0$ for x outside the interval $[0, 1]$. Similarly, the marginal p.d.f. of Y is

$$f_2(y) = \begin{cases} 2(1 - y), & 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Both X and Y have the same marginal Beta $(1, 2)$ distribution. Thus,

$$E\{X\} = E\{Y\} = \frac{1}{3}$$

and

$$V\{X\} = V\{Y\} = \frac{1}{18}.$$

■

3.5.2 Covariance and correlation

Given any two random variables (X_1, X_2) having a joint distribution with p.d.f. $f(x_1, x_2)$, the **covariance** of X_1 and X_2 is defined as

$$\text{Cov}(X_1, X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 - \mu_1)(x_2 - \mu_2)f(x_1, x_2)dx_1dx_2, \tag{3.81}$$

where

$$\mu_i = \int_{-\infty}^{\infty} xf_i(x)dx, \quad i = 1, 2,$$

is the expected value of X_i. Notice that

$$\begin{aligned}\text{Cov}(X_1, X_2) &= E\{(X_1 - \mu_1)(X_2 - \mu_2)\}\\ &= E\{X_1X_2\} - \mu_1\mu_2.\end{aligned}$$

The **correlation** between X_1 and X_2 is defined as

$$\rho_{12} = \frac{\text{Cov}(X_1, X_2)}{\sigma_1\sigma_2}, \tag{3.82}$$

where σ_i $(i = 1, 2)$ is the standard deviation of X_i.

Example 3.28. In continuation of the previous example, we compute $\text{Cov}(X, Y)$.
We have seen that $E\{X\} = E\{Y\} = \frac{1}{3}$. We compute now the expected value of their product

$$\begin{aligned}E\{XY\} &= 2\int_0^1 x\int_0^{1-x} ydy\\ &= 2\int_0^1 x \cdot \frac{1}{2}(1-x)^2dx\\ &= B(2,3) = \frac{\Gamma(2)\Gamma(3)}{\Gamma(5)} = \frac{1}{12}.\end{aligned}$$

Hence,

$$\begin{aligned}\text{Cov}(X, Y) = E\{XY\} - \mu_1\mu_2 &= \frac{1}{12} - \frac{1}{9}\\ &= -\frac{1}{36}.\end{aligned}$$

Finally, the correlation between X, Y is

$$\rho_{XY} = -\frac{1/36}{1/18} = -\frac{1}{2}.$$

■

The following are some properties of the covariance

(i)
$$|\text{Cov}(X_1, X_2)| \le \sigma_1\sigma_2,$$
where σ_1 and σ_2 are the standard deviations of X_1 and X_2, respectively.

(ii) If c is any constant, then
$$\text{Cov}(X, c) = 0. \tag{3.83}$$

(iii) For any constants a_1 and a_2,
$$\text{Cov}(a_1X_1, a_2X_2) = a_1a_2\text{Cov}(X_1, X_2). \tag{3.84}$$

(iv) For any constants a, b, c, and d,

$$\begin{aligned}\mathrm{Cov}(aX_1 + bX_2, cX_3 + dX_4) &= ac\,\mathrm{Cov}(X_1, X_3) + ad\,\mathrm{Cov}(X_1, X_4)\\ &\quad + bc\,\mathrm{Cov}(X_2, X_3) + bd\,\mathrm{Cov}(X_2, X_4).\end{aligned}$$

Property (iv) can be generalized to be

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j\,\mathrm{Cov}(X_i, Y_j). \tag{3.85}$$

From property (i) above we deduce that $-1 \leq \rho_{12} \leq 1$. The correlation obtains the values ± 1 only if the two variables are linearly dependent.

Definition of Independence.

Random variables $X_1, \cdots, X_k$ are said to be **mutually independent** if, for every $(x_1, \cdots, x_k)$,

$$f(x_1, \cdots, x_k) = \prod_{i=1}^{k} f_i(x_i), \tag{3.86}$$

where $f_i(x_i)$ is the marginal p.d.f. of X_i. The variables X, Y of Example 3.26 are dependent, since $f(x, y) \neq f_1(x) f_2(y)$.

If two random variables are independent, then their correlation (or covariance) is zero. The converse is generally not true. Zero correlation **does not** imply independence.

We illustrate this in the following example.

Example 3.29. Let (X, Y) be discrete random variables having the following joint p.d.f.

$$p(x, y) = \begin{cases} \frac{1}{3}, & \text{if } X = -1, Y = 0 \text{ or } X = 0, Y = 0 \text{ or } X = 1, Y = 1 \\ 0, & \text{elsewhere.} \end{cases}$$

In this case the marginal p.d.f. are

$$p_1(x) = \begin{cases} \frac{1}{3}, & x = -1, 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

$$p_2(y) = \begin{cases} \frac{1}{3}, & y = 0 \\ \frac{2}{3}, & y = 1. \end{cases}$$

$p(x, y) \neq p_1(x) p_2(y)$ if $X = 1, Y = 1$ for example. Thus, X and Y are dependent. On the other hand, $E\{X\} = 0$ and $E\{XY\} = 0$. Hence, $\mathrm{Cov}(X, Y) = 0$. ■

The following result is very important for independent random variables.

If $X_1, X_2, \ldots, X_k$ **are mutually independent then, for any integrable functions** $g_1(X_1), \ldots, g_k(X_k)$,

$$E\left\{\prod_{i=1}^{k} g_i(X_i)\right\} = \prod_{i=1}^{k} E\{g_i(X_i)\}. \tag{3.87}$$

Indeed,

$$
\begin{aligned}
E\left\{\prod_{i=1}^{k} g_i(X_i)\right\} &= \int \cdots \int g_1(x_1)\cdots g_k(x_k)\cdot \\
f(x_1,\cdots,x_k)dx_1,\cdots,dx_k &= \int \cdots \int g_1(x_1)\cdots g_k(x_k)f_1(x_1)\cdots f_k(x_k)dx_1\cdots dx_k \\
&= \int g_1(x_1)f_1(x_1)dx_1 \cdot \int g_2(x_2)f_2(x_2)dx_2 \cdots \int g_k(x_k)f_k(x_k)dx_k \\
&= \prod_{i=1}^{k} E\{g_i(X_i)\}.
\end{aligned}
$$

3.5.3 Conditional distributions

If (X_1, X_2) are two random variables having a joint p.d.f. $f(x_1, x_2)$ and marginals ones, $f_1(\cdot)$, and $f_2(\cdot)$, respectively, then the **conditional** p.d.f. of X_2, given $\{X_1 = x_1\}$, where $f_1(x_1) > 0$, is defined to be

$$f_{2\cdot 1}(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}. \tag{3.88}$$

Notice that $f_{2\cdot 1}(x_2 \mid x_1)$ is a p.d.f. Indeed, $f_{2\cdot 1}(x_2 \mid x_1) \geq 0$ for all x_2, and

$$
\begin{aligned}
\int_{-\infty}^{\infty} f_{2\cdot 1}(x_2 \mid x_1)dx_2 &= \frac{\int_{-\infty}^{\infty} f(x_1, x_2)dx_2}{f_1(x_1)} \\
&= \frac{f_1(x_1)}{f_1(x_1)} = 1.
\end{aligned}
$$

The **conditional expectation** of X_2, given $\{X_1 = x_1\}$ such that $f_1(x_1) > 0$, is the expected value of X_2 with respect to the conditional p.d.f. $f_{2\cdot 1}(x_2 \mid x_1)$, that is,

$$E\{X_2 \mid X_1 = x_1\} = \int_{-\infty}^{\infty} x f_{2\cdot 1}(x \mid x_1)dx.$$

Similarly, we can define the **conditional variance** of X_2, given $\{X_1 = x_1\}$, as the variance of X_2, with respect to the conditional p.d.f. $f_{2\cdot 1}(x_2 \mid x_1)$. If X_1 and X_2 are independent, then by substituting $f(x_1, x_2) = f_1(x_1)f_2(x_2)$ we obtain

$$f_{2\cdot 1}(x_2 \mid x_1) = f_2(x_2),$$

and

$$f_{1\cdot 2}(x_1 \mid x_2) = f_1(x_1).$$

Example 3.30. Returning to Example 3.26, we compute the conditional distribution of Y, given $\{X = x\}$, for $0 < x < 1$.

According to the above definition, the conditional p.d.f. of Y, given $\{X = x\}$, for $0 < x < 1$, is

$$
f_{Y|X}(y \mid x) = \begin{cases} \dfrac{1}{1-x}, & \text{if } 0 < y < (1-x) \\ \\ 0, & \text{otherwise.} \end{cases}
$$

Notice that this is a uniform distribution over $(0, 1-x)$, $0 < x < 1$. If $x \notin (0, 1)$, then the conditional p.d.f. does not exist. This is, however, an event of probability zero. From the above result, the conditional expectation of Y, given $X = x$, $0 < x < 1$, is

$$E\{Y \mid X = x\} = \frac{1-x}{2}.$$

The conditional variance is

$$V\{Y \mid X = x\} = \frac{(1-x)^2}{12}.$$

In a similar fashion we show that the conditional distribution of X, given $Y = y$, $0 < y < 1$, is uniform on $(0, 1 - y)$. ■

One can immediately prove that if X_1 and X_2 are independent, then the conditional distribution of X_1 given $\{X_2 = x_2\}$, when $f_2(x_2) > 0$, is just the marginal distribution of X_1. Thus, X_1 and X_2 are independent if, and only if,

$$f_{2\cdot 1}(x_2 \mid x_1) = f_2(x_2) \quad \text{for all} \quad x_2$$

and

$$f_{1\cdot 2}(x_1 \mid x_2) = f_1(x_1) \quad \text{for all} \quad x_1,$$

provided that the conditional p.d.f. are well defined.

Notice that for a pair of random variables (X, Y), $E\{Y \mid X = x\}$ changes with x, as shown in Example 3.29, if X and Y are dependent. Thus, we can consider $E\{Y \mid X\}$ to be a random variable, which is a function of X. It is interesting to compute the expected value of this function of X, that is,

$$\begin{aligned} E\{E\{Y \mid X\}\} &= \int E\{Y \mid X = x\} f_1(x)dx \\ &= \int \left\{ \int y f_{Y\cdot X}(y \mid x)dy \right\} f_1(x)dx \\ &= \int \int y \frac{f(x,y)}{f_1(x)} f_1(x)dydx. \end{aligned}$$

If we can interchange the order of integration (whenever $\int |y| f_2(y)dy < \infty$), then

$$\begin{aligned} E\{E\{Y \mid X\}\} &= \int y \left\{ \int f(x,y)dx \right\} dy \\ &= \int y f_2(y)dy \qquad (3.89) \\ &= E\{Y\}. \end{aligned}$$

This result, known as **the law of the iterated expectation**, is often very useful. An example of the use of the law of the iterated expectation is the following.

Example 3.31. Let (J, N) be a pair of random variables. The conditional distribution of J, given $\{N = n\}$, is the binomial $B(n, p)$. The marginal distribution of N is Poisson with mean λ. What is the expected value of J?

By the law of the iterated expectation,

$$\begin{aligned} E\{J\} &= E\{E\{J \mid N\}\} \\ &= E\{Np\} = pE\{N\} = p\lambda. \end{aligned}$$

One can show that the marginal distribution of J is Poisson, with mean $p\lambda$. ■

Another important result relates variances and conditional variances. That is, if (X, Y) is a pair of random variables, having finite variances then

$$V\{Y\} = E\{V\{Y \mid X\}\} + V\{E\{Y \mid X\}\}. \qquad (3.90)$$

We call this relationship the **law of total variance**.

Example 3.32. Let (X, Y) be a pair of independent random variables having finite variances σ_X^2 and σ_Y^2 and expected values μ_X, μ_Y. Determine the variance of $W = XY$. By the law of total variance,

$$V\{W\} = E\{V\{W \mid X\}\} + V\{E\{W \mid X\}\}.$$

Since X and Y are independent

$$\begin{aligned} V\{W \mid X\} &= V\{XY \mid X\} = X^2 V\{Y \mid X\} \\ &= X^2 \sigma_Y^2. \end{aligned}$$

Similarly,

$$E\{W \mid X\} = X\mu_Y.$$

Hence,

$$\begin{aligned} V\{W\} &= \sigma_Y^2 E\{X^2\} + \mu_Y^2 \sigma_X^2 \\ &= \sigma_Y^2(\sigma_X^2 + \mu_X^2) + \mu_Y^2 \sigma_X^2 \\ &= \sigma_X^2 \sigma_Y^2 + \mu_X^2 \sigma_Y^2 + \mu_Y^2 \sigma_X^2. \end{aligned}$$

■

3.6 Some multivariate distributions

3.6.1 The multinomial distribution

The multinomial distribution is a generalization of the binomial distribution to cases of n **independent** trials in which the results are classified to k possible categories (e.g. Excellent, Good, Average, Poor). The random variables $(J_1, J_2, \cdots, J_k)$ are the number of trials yielding results in each one of the k categories. These random variables are dependent, since $J_1 + J_2 + \cdots + J_k = n$. Furthermore, let $p_1, p_2, \cdots, p_k$; $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$, be the probabilities of the k categories. The binomial distribution is the special case of $k = 2$. Since $J_k = n - (J_1 + \cdots + J_{k-1})$, the joint probability function is written as a function of $k - 1$ arguments, and its formula is

$$p(j_1, \cdots, j_{k-1}) = \binom{n}{j_1, \cdots, j_{k-1}} p_1^{j_1} \cdots p_{k-1}^{j_{k-1}} p_k^{j_k} \tag{3.91}$$

for $j_1, \cdots, j_{k-1} \geq 0$ such that $\sum_{i=1}^{k-1} j_i \leq n$. In this formula,

$$\binom{n}{j_1, \cdots, j_{k-1}} = \frac{n!}{j_1! j_2! \cdots j_k!}, \tag{3.92}$$

and $j_k = n - (j_1 + \cdots + j_{k-1})$. For example, if $n = 10$, $k = 3$, $p_1 = .3$, $p_2 = .4$, $p_3 = .3$,

$$\begin{aligned} p(5, 2) &= \frac{10!}{5!2!3!}(0.3)^5(0.4)^2(0.3)^3 \\ &= 0.02645. \end{aligned}$$

The marginal distribution of each one of the k variables is binomial, with parameters n and p_i $(i = 1, \cdots, k)$. The joint marginal distribution of (J_1, J_2) is trinomial, with parameters n, p_1, p_2 and $(1 - p_1 - p_2)$. Finally, the conditional distribution of $(J_1, \cdots, J_r)$, $1 \leq r < k$, given $\{J_{r+1} = j_{r+1}, \cdots, J_k = j_k\}$ is $(r + 1)$-nomial, with parameters $n_r = n - (j_{r+1} + \cdots + j_k)$ and $p'_1, \cdots, p'_r, p'_{r+1}$, where

$$p'_i = \frac{p_i}{(1 - p_{r+1} - \cdots - p_k)}, \quad i = 1, \cdots, r$$

and

$$p'_{r+1} = 1 - \sum_{i=1}^{r} p'_i.$$

Finally, we can show that, for $i \neq j$,

$$\text{Cov}(J_i, J_j) = -np_ip_j. \tag{3.93}$$

Example 3.33. An insertion machine is designed to insert components into computer printed circuit boards. Every component inserted on a board is scanned optically. An insertion is either error-free or its error is classified to the following two main categories: misinsertion (broken lead, off pad, etc.) or wrong component. Thus, we have altogether three general categories. Let

$$J_1 = \# \text{ of error-free components;}$$

$$J_2 = \# \text{ of misinsertion;}$$

$$J_3 = \# \text{ of wrong components.}$$

The probabilities that an insertion belongs to one of these categories is $p_1 = 0.995$, $p_2 = 0.001$ and $p_2 = 0.004$.

The insertion rate of this machine is $n = 3500$ components per hour of operation. Thus, we expect during one hour of operation $n \times (p_2 + p_3) = 175$ insertion errors.

Given that there are 16 insertion errors during a particular hour of operation, the conditional distribution of the number of misinsertions is binomial $B\left(16, \frac{0.01}{0.05}\right)$.

Thus,

$$E\{J_2 \mid J_2 + J_3 = 16\} = 16 \times 0.2 = 3.2.$$

On the other hand,

$$E\{J_2\} = 3500 \times 0.001 = 3.5.$$

We see that the information concerning the total number of insertion errors makes a difference.

Finally

$$\begin{aligned}\text{Cov}(J_2, J_3) &= -3500 \times 0.001 \times 0.004 \\ &= -0.014 \\ V\{J_2\} &= 3500 \times 0.001 \times 0.999 = 3.4965\end{aligned}$$

and

$$V\{J_3\} = 3500 \times 0.004 \times 0.996 = 13.944.$$

Hence, the correlation between J_2 and J_3 is

$$\rho_{2,3} = \frac{-0.014}{\sqrt{3.4965 \times 13.944}} = -0.0020.$$

This correlation is quite small. ■

3.6.2 The multi-hypergeometric distribution

Suppose that we draw from a population of size N a RSWOR of size n. Each one of the n units in the sample is classified to one of k categories. Let $J_1, J_2, \cdots, J_k$ be the number of sample units belonging to each one of these categories. $J_1 + \cdots + J_k = n$. The distribution of $J_1, \cdots, J_k$ is k-variate hypergeometric. If $M_1, \cdots, M_k$ are the number of units in the population in these categories, before the sample is drawn, then the joint p.d.f. of $J_1, \cdots, J_k$ is

$$p(j_1, \cdots, j_{k-1}) = \frac{\binom{M_1}{j_1}\binom{M_2}{j_2}\cdots\binom{M_k}{j_k}}{\binom{N}{n}}, \tag{3.94}$$

where $j_k = n - (j_1 + \cdots + j_{k-1})$. This distribution is a generalization of the hypergeometric distribution $H(N, M, n)$. The hypergeometric distribution $H(N, M_i, n)$ is the marginal distribution of J_i $(i = 1, \cdots, k)$. Thus,

$$E\{J_i\} = n\frac{M_i}{N}, \quad i = 1, \cdots, k$$
$$V\{J_i\} = n\frac{M_i}{N}\left(1 - \frac{M_i}{N}\right)\left(1 - \frac{n-1}{N-1}\right), \quad i = 1, \cdots, k \tag{3.95}$$

and for $i \neq j$

$$\mathrm{Cov}(J_i, J_j) = -n\frac{M_i}{N} \cdot \frac{M_j}{N}\left(1 - \frac{n-1}{N-1}\right).$$

Example 3.34. A lot of 100 spark plugs contains 20 plugs from vendor V_1, 50 plugs from vendor V_2 and 30 plugs from vendor V_3.

A random sample of $n = 20$ plugs is drawn from the lot without replacement.

Let J_i be the number of plugs in the sample from the vendor V_i, $i = 1, 2, 3$. Accordingly,

$$\Pr\{J_1 = 5, J_2 = 10\} = \frac{\binom{20}{5}\binom{50}{10}\binom{30}{5}}{\binom{100}{20}} = 0.00096.$$

If we are told that 5 out of the 20 plugs in the sample are from vendor V_3, then the conditional distribution of J_1 is

$$\Pr\{J_1 = j_1 \mid J_3 = 5\} = \frac{\binom{20}{j_1}\binom{50}{15-j_1}}{\binom{70}{15}}, \quad j_1 = 0, \cdots, 15.$$

Indeed, given $J_3 = 5$, then J_1 can assume only the values $0, 1, \cdots, 15$. The conditional probability that j_1 out of the 15 remaining plugs in the sample are from vendor V_1, is the same as that of choosing a RSWOR of size 15 from a lot of size $70 = 20 + 50$, with 20 plugs from vendor V_1. ■

3.6.3 The bivariate normal distribution

The bivariate normal distribution is the joint distribution of two continuous random variables (X, Y) having a joint p.d.f.

$$f(x, y; \mu, \eta, \sigma_X, \sigma_Y, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu}{\sigma_x}\right)^2 - 2\rho\frac{x-\mu}{\sigma_Y} \cdot \frac{y-\eta}{\sigma_Y} + \left(\frac{y-\eta}{\sigma_Y}\right)^2\right]\right\},$$
$$-\infty < x, y < \infty. \tag{3.96}$$

μ, η, σ_X, σ_Y and ρ are parameters of this distribution.

Integration of y yields that the marginal distribution of X is $N(\mu, \sigma_x^2)$. Similarly, the marginal distribution of Y is $N(\eta, \sigma_Y^2)$. Furthermore, ρ is the correlation between X and Y. Notice that if $\rho = 0$, then the joint p.d.f. becomes the product of the two marginal ones, that is,

$$f(x, y); \mu, \eta, \sigma_X, \sigma_Y, 0) = \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_X}\right)^2\right\} \cdot \frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\frac{1}{2}\left(\frac{y-\eta}{\sigma_Y}\right)^2\right\}, \quad \text{for all } -\infty < x, y < \infty.$$

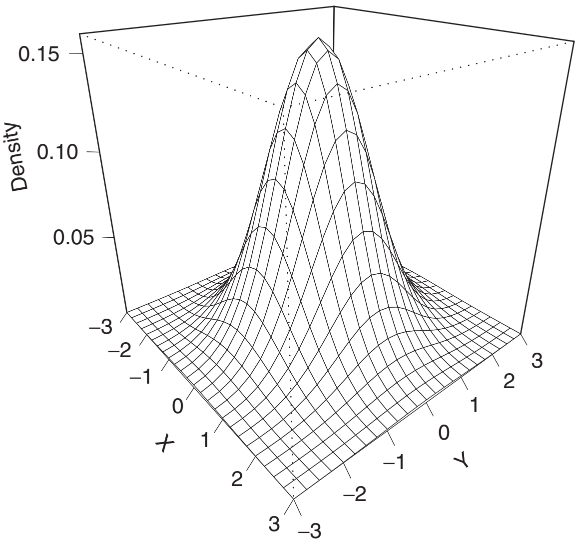

Figure 3.17 *Bivariate normal p.d.f.*

Hence, if $\rho = 0$, then X and Y are independent. On the other hand, if $\rho \neq 0$, then $f(x, y; \mu, \eta, \sigma_X, \sigma_Y, \rho) \neq f_1(x; \mu, \sigma_X) f_2(y; \eta, \sigma_Y)$, and the two random variables are dependent.

In Figure 3.17 we present the bivariate p.d.f. for $\mu = \eta = 0$, $\sigma_X = \sigma_Y = 1$ and $\rho = 0.5$.

One can verify also that the conditional distribution of Y, given $\{X = x\}$ is normal with mean

$$\mu_{Y \cdot x} = \eta + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu) \tag{3.97}$$

and variance

$$\sigma^2_{Y \cdot x} = \sigma^2_Y (1 - \rho^2). \tag{3.98}$$

It is interesting to see that $\mu_{Y \cdot x}$ is a linear function of x. We can say that $\mu_{Y \cdot x} = E\{Y \mid X = x\}$ is, in the bivariate normal case, the theoretical (linear) regression of Y on X (see Chapter 5). Similarly,

$$\mu_{X \cdot y} = \mu + \rho \frac{\sigma_X}{\sigma_Y}(y - \eta),$$

and

$$\sigma^2_{X \cdot y} = \sigma^2_X (1 - \rho^2).$$

If $\mu = \eta = 0$ and $\sigma_X = \sigma_Y = 1$, we have the **standard** bivariate normal distribution. The joint c.d.f. in the standard case is denoted by $\Phi_2(x, y; \rho)$ and its formula is

$$\begin{aligned} \Phi_2(x, y; \rho) &= \frac{1}{2\pi\sqrt{1 - \rho^2}} \int_{-\infty}^{x} \int_{-\infty}^{y} \exp\left\{ -\frac{1}{2(1 - \rho^2)} (z_1^2 - 2\rho z_1 z_2 + z^2) \right\} dz_1 dz_2 \\ &= \int_{-\infty}^{x} \phi(z_1) \Phi\left(\frac{y - \rho z_1}{\sqrt{1 - \rho^2}} \right) dz_1 \end{aligned} \tag{3.99}$$

values of $\Phi_2(x, y; \rho)$ can be obtained by numerical integration. If one has to compute the bivariate c.d.f. in the general case, the following formula is useful

$$F(x, y; \mu, \eta, \sigma_X, \sigma_Y, \rho) = \Phi_2\left(\frac{x - \mu}{\sigma_X}, \frac{y - \eta}{\sigma_Y}; \rho \right).$$

For computing $\Pr\{a \le X \le b, c \le Y \le d\}$ we use the formula,

$$\Pr\{a \le X \le b, c \le Y \le d\} = F(b, d;\ -)$$
$$- F(a, d;\ -) - F(b, c;\ -) + F(a, c;\ -).$$

Example 3.35. Suppose that (X, Y) deviations in components placement on PCB by an automatic machine have a bivariate normal distribution with means $\mu = \eta = 0$, standard deviations $\sigma_X = 0.00075$ and $\sigma_Y = 0.00046$ [Inch] and $\rho = 0.160$. The placement errors are within the specifications if $|X| < 0.001$ [Inch] and $|Y| < 0.001$ [Inch]. What proportion of components are expected to have X, Y deviations compliant with the specifications? The standardized version of the spec limits are $Z_1 = \frac{0.001}{0.00075} = 1.33$ and $Z_2 = \frac{0.001}{0.00046} = 2.174$. We compute

$$\begin{aligned}\Pr\{|X| < 0.001, |Y| < 0.001\} &= \Phi_2(1.33, 2.174,\ .16) - \Phi_2(-1.33, 2.174,\ .16)\\ &\quad - \Phi_2(1.33, -2.174;.16) + \Phi_2(-1.33, -2.174;.16)\\ &= 0.793.\end{aligned}$$

This is the expected proportion of good placements. ∎

3.7 Distribution of order statistics

As defined in Chapter 2, the order statistics of the sample are the sorted data. More specifically, let $X_1, \cdots, X_n$ be identically distributed independent (i.i.d.) random variables. The order statistics are $X_{(i)}$, $i = 1, \cdots, n$, where

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

In the present section we discuss the distributions of these order statistics, when $F(x)$ is (absolutely) continuous, having a p.d.f. $f(x)$.

We start with the extremal statistics $X_{(1)}$ and $X_{(n)}$.

Since the random variables X_i $(i = 1, \cdots, n)$ are i.i.d., the c.d.f. of $X_{(1)}$ is

$$\begin{aligned}F_{(1)}(x) &= \Pr\{X_{(1)} \le x\}\\ &= 1 - \Pr\{X_{(1)} \ge x\} = 1 - \prod_{i=1}^{n} \Pr\{X_i \ge x\}\\ &= 1 - (1 - F(x))^n.\end{aligned}$$

By differentiation we obtain that the p.d.f. of $X_{(1)}$ is

$$f_{(1)}(x) = nf(x)[1 - F(x)]^{n-1}. \tag{3.100}$$

Similarly, the c.d.f. of the sample maximum $X_{(n)}$ is

$$\begin{aligned}F_{(n)}(x) &= \prod_{i=1}^{n} \Pr\{X_i \le x\}\\ &= (F(x))^n.\end{aligned}$$

The p.d.f. of $X_{(n)}$ is

$$f_{(n)}(x) = nf(x)(F(x))^{n-1}. \tag{3.101}$$

Example 3.36. (i) A switching circuit consists of n modules, which operate independently and which are connected in **series** (see Figure 3.18). Let X_i be the time till failure of the i-th module. The system fails when any module fails. Thus, the time till failure of the system is $X_{(1)}$. If all X_i are exponentially distributed with mean life β, then the c.d.f. of $X_{(1)}$ is

$$F_{(1)}(x) = 1 - e^{-nx/\beta}, \quad x \geq 0.$$

Thus, $X_{(1)}$ is distributed like $E\left(\frac{\beta}{n}\right)$. It follows that the expected time till failure of the circuit is $E\{X_{(1)}\} = \frac{\beta}{n}$.

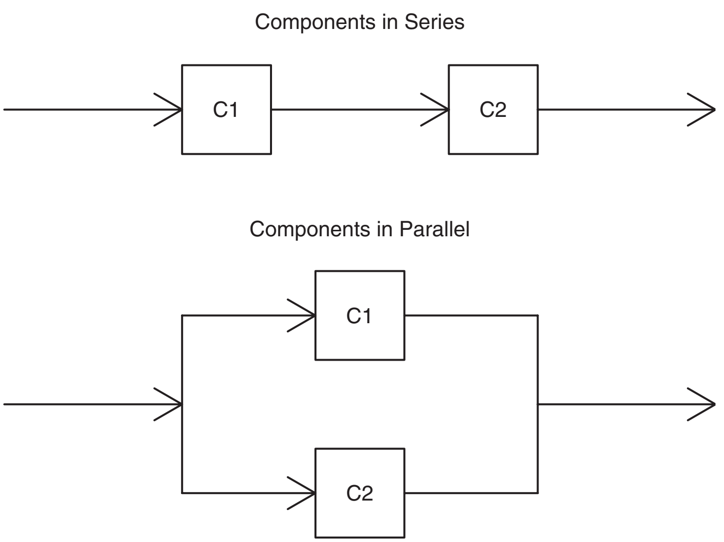

Figure 3.18 *Series and parallel systems*

(ii) If the modules are connected in parallel, then the circuit fails at the instant the last of the n modules fail, which is $X_{(n)}$. Thus, if X_i is $E(\beta)$, the c.d.f. of $X_{(n)}$ is

$$F_{(n)}(x) = (1 - e^{-(x/\beta)})^n.$$

The expected value of $X_{(n)}$ is

$$\begin{aligned} E\{X_{(n)}\} &= \frac{n}{\beta}\int_0^\infty xe^{-x/\beta}(1 - e^{-x/\beta})^{n-1}dx \\ &= n\beta\int_0^\infty ye^{-y}(1 - e^{-y})^{n-1}dy \\ &= n\beta\sum_{j=0}^{n-1}(-1)^j\binom{n-1}{j}\int_0^\infty ye^{-(1+j)y}dy \\ &= n\beta\sum_{j=1}^{n}(-1)^{j-1}\binom{n-1}{j-1}\frac{1}{j^2}. \end{aligned}$$

Furthermore, since $n\binom{n-1}{j-1} = j\binom{n}{j}$, we obtain that

$$E\{X_{(n)}\} = \beta\sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\frac{1}{j}.$$

One can also show that this formula is equivalent to

$$E\{X_{(n)}\} = \beta \sum_{j=1}^{n} \frac{1}{j}.$$

Accordingly, if the parallel circuit consists of 3 modules, and the time till failure of each module is exponential with $\beta = 1,000$ [hr], the expected time till failure of the system is 1,833.3 [hr]. ■

Generally, the distribution of $X_{(i)}$ $(i = 1, \cdots, n)$ can be obtained by the following argument. The event $\{X_{(i)} \leq x\}$ is equivalent to the event that the number of X_i values in the random example which are smaller or equal to x is at least i.

Consider n independent and identical trials, in which "success" is that $\{X_i \leq x\}$ $(i = 1, \cdots, n)$. The probability of "success" is $F(x)$. The distribution of the number of successes is $B(n, F(x))$. Thus, the c.d.f. of $X_{(i)}$ is

$$\begin{aligned} F_{(i)}(x) = \Pr\{X_{(i)} \leq x\} &= 1 - B(i-1; n, F(x)) \\ &= \sum_{j=i}^{n} \binom{n}{j} (F(x))^j (1 - F(x))^{n-j}. \end{aligned}$$

Differentiating this c.d.f. with respect to x yields the p.d.f. of $X_{(i)}$, namely:

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} f(x)(F(x))^{i-1}(1 - F(x))^{n-i}. \tag{3.102}$$

Notice that if X has a uniform distribution on $(0, 1)$, then the distribution of $X_{(i)}$ is like that of Beta$(i, n - i + 1)$, $i = 1, \cdots, n$. In a similar manner one can derive the joint p.d.f. of $(X_{(i)}, X_{(j)})$, $1 \leq i < j \leq n$, etc. This joint p.d.f. is given by

$$\begin{aligned} f_{(i),(j)}(x, y) = & \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(x)f(y) \cdot \\ & \cdot (F(x))^{i-1}[F(y) - F(x)]^{j-i-1}(1 - F(y))^{n-j}, \end{aligned} \tag{3.103}$$

for $-\infty < x < y < \infty$.

3.8 Linear combinations of random variables

Let $X_1, X_2, \cdots, X_n$ be random variables having a joint distribution, with joint p.d.f. $f(x_1, \cdots, x_n)$. Let $\alpha_1, \cdots, \alpha_n$ be given constants. Then

$$W = \sum_{i=1}^{n} \alpha_i X_i$$

is a linear combination of the X's. The p.d.f. of W can generally be derived, using various methods. We discuss in the present section only the formulae of the expected value and variance of W.

It is straightforward to show that

$$E\{W\} = \sum_{i=1}^{n} \alpha_i E\{X_i\}. \tag{3.104}$$

That is, the expected value of a linear combination is the same linear combination of the expectations.

The formula for the variance is somewhat more complicated, and is given by

$$V\{W\} = \sum_{i=1}^{n} \alpha_i^2 V\{X_i\} + \sum\sum_{i \neq j} \alpha_i \alpha_j \text{cov}(X_i, X_j). \tag{3.105}$$

Example 3.37. Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables, with common expectations μ and common finite variances σ^2. The sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a particular linear combination, with

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = \frac{1}{n}.$$

Hence,

$$E\{\overline{X}_n\} = \frac{1}{n}\sum_{i=1}^{n} E\{X_i\} = \mu$$

and, since $X_1, X_2, \cdots, X_n$ are mutually independent, $\text{cov}(X_i, X_j) = 0$, all $i \neq j$. Hence

$$V\{\overline{X}_n\} = \frac{1}{n^2}\sum_{i=1}^{n} V\{X_i\} = \frac{\sigma^2}{n}.$$

Thus, we have shown that in a random sample of n i.i.d. random variables, the sample mean has the same expectation as that of the individual variables, but its sample variance is reduced by a factor of $1/n$.

Moreover, from Chebychev's inequality, for any $\epsilon > 0$

$$\Pr\{|\overline{X}_n - \mu| > \epsilon\} < \frac{\sigma^2}{n\epsilon^2}.$$

Therefore, since $\lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$,

$$\lim_{n\to\infty} \Pr\{|\overline{X}_n - \mu| > \epsilon\} = 0.$$

This property is called the **convergence in probability** of $\overline{X}_n$ to μ. ■

Example 3.38. Let U_1, U_2, U_3 be three i.i.d. random variables having uniform distributions on $(0, 1)$. We consider the statistic

$$W = \frac{1}{4}U_{(1)} + \frac{1}{2}U_{(2)} + \frac{1}{4}U_{(3)},$$

where $0 < U_{(1)} < U_{(2)} < U_{(3)} < 1$ are the order statistics. We have seen in Section 4.7 that the distribution of $U_{(i)}$ is like that of Beta$(i, n-i+1)$. Hence

$$E\{U_{(1)}\} = E\{\text{Beta}(1,3)\} = \frac{1}{4}$$

$$E\{U_{(2)}\} = E\{\text{Beta}(2,2)\} = \frac{1}{2}$$

$$E\{U_{(3)}\} = E\{\text{Beta}(3,1)\} = \frac{3}{4}.$$

It follows that

$$E\{W\} = \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{2}.$$

To find the variance of W we need more derivations.

First

$$V\{U_{(1)}\} = V\{\text{Beta}(1,3)\} = \frac{3}{4^2\times 5} = \frac{3}{80}$$

$$V\{U_{(2)}\} = V\{\text{Beta}(2,2)\} = \frac{4}{4^2\times 5} = \frac{1}{20}$$

$$V\{U_{(3)}\} = V\{\text{Beta}(3,1)\} = \frac{3}{4^2\times 5} = \frac{3}{80}.$$

We need to find $\text{Cov}(U_{(1)}, U_{(2)})$, $\text{Cov}(U_{(1)}, U_{(3)})$ and $\text{Cov}(U_{(2)}, U_{(3)})$. From the joint p.d.f. formula of order statistics, the joint p.d.f. of $(U_{(1)}, U_{(2)})$ is

$$f_{(1),(2)}(x, y) = 6(1 - y), \quad 0 < x \leq y < 1.$$

Hence

$$\begin{aligned} E\{U_{(1)}U_{(2)}\} &= 6\int_0^1 x\left(\int_0^1 y(1-y)dy\right)dx \\ &= \frac{6}{40}. \end{aligned}$$

Thus,

$$\begin{aligned} \text{Cov}(U_{(1)}, U_{(2)}) &= \frac{6}{40} - \frac{1}{4}\cdot\frac{1}{2} \\ &= \frac{1}{40}. \end{aligned}$$

Similarly, the p.d.f. of $(U_{(1)}, U_{(3)})$ is

$$f_{(1),(3)}(x, y) = 6(y - x), \quad 0 < x \leq y < 1.$$

Thus,

$$\begin{aligned} E\{U_{(1)}U_{(3)}\} &= 6\int_0^1 x\left(\int_x^1 y(y-x)dy\right)dx \\ &= 6\int_0^1 x\left(\frac{1}{3}\left(1-x^3\right) - \frac{x}{2}(1-x^2)\right)dx \\ &= \frac{1}{5}, \end{aligned}$$

and

$$\text{Cov}(U_{(1)}, U_{(3)}) = \frac{1}{5} - \frac{1}{4}\cdot\frac{3}{4} = \frac{1}{80}.$$

The p.d.f. of $(U_{(2)}, U_{(3)})$ is

$$f_{(2),(3)}(x, y) = 6x, \quad 0 < x \leq y \leq 1,$$

and

$$\text{Cov}(U_{(2)}, U_{(3)}) = \frac{1}{40}.$$

Finally,

$$\begin{aligned} V\{W\} &= \frac{1}{16}\cdot\frac{3}{80} + \frac{1}{4}\cdot\frac{1}{20} + \frac{1}{16}\cdot\frac{3}{80} \\ &\quad + 2\cdot\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{40} + 2\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{80} \\ &\quad + 2\cdot\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{40} \\ &= \frac{1}{32} = 0.03125. \end{aligned}$$

■

The following is a useful result:

If $X_1, X_2, \cdots, X_n$ are mutually independent, then the m.g.f. of $T_n = \sum_{i=1}^{n} X_i$ is

$$M_{T_n}(t) = \prod_{i=1}^{n} M_{X_i}(t). \tag{3.106}$$

Indeed, as shown in Section 4.5.2, when $X_1, \ldots, X_n$ are independent, the expected value of the product of functions is the product of their expectations.

Therefore,

$$\begin{aligned} M_{T_n}(t) &= E\{e^{t\sum_{i=1}^{n} X_i}\} \\ &= E\left\{\prod_{i=1}^{n} e^{tX_i}\right\} \\ &= \prod_{i=1}^{n} E\{e^{tX_i}\} \\ &= \prod_{i=1}^{n} M_{X_i}(t). \end{aligned}$$

The expected value of the product is equal to the product of the expectations, since $X_1, \cdots, X_n$ are mutually independent.

Example 3.39. In the present example we illustrate some applications of the last result.

(i) Let $X_1, X_2, \cdots, X_k$ be independent random variables having binomial distributions like $B(n_i, p)$, $i = 1, \cdots, k$, then their sum T_k has the binomial distribution. To show this,

$$\begin{aligned} M_{T_k}(t) &= \prod_{i=1}^{k} M_{X_i}(t) \\ &= [e^t p + (1-p)]^{\sum_{i=1}^{k} n_i}. \end{aligned}$$

That is, T_k is distributed like $B\left(\sum_{i=1}^{k} n_i, p\right)$. This result is intuitively clear.

(ii) If $X_1, \cdots, X_n$ are independent random variables, having Poisson distributions with parameters λ_i $(i = 1, \cdots, n)$ then the distribution of $T_n = \sum_{i=1}^{n} X_i$ is Poisson with parameter $\mu_n = \sum_{i=1}^{n} \lambda_i$. Indeed,

$$\begin{aligned} M_{T_n}(t) &= \prod_{j=1}^{n} \exp\ \{-\lambda_j(1-e^t)\} \\ &= \exp\left\{-\sum_{j=1}^{n} \lambda_j(1-e^t)\right\} \\ &= \exp\ \{-\mu_n(1-e^t)\}. \end{aligned}$$

(iii) Suppose $X_1, \cdots, X_n$ are independent random variables, and the distribution of X_i is normal $N(\mu_i, \sigma_i^2)$, then the distribution of $W = \sum_{i=1}^{n} \alpha_i X_i$ is normal like that of

$$N\left(\sum_{i=1}^{n} \alpha_i \mu_i, \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2\right).$$

To verify this we recall that $X_i = \mu_i + \sigma_i Z_i$, where Z_i is $N(0, 1)$ $(i = 1, \cdots, n)$. Thus

$$\begin{aligned} M_{\alpha_i X_i}(t) &= E\{e^{t(\alpha_i\mu_i + \alpha_i\sigma_i Z_i}\} \\ &= e^{t\alpha_i\mu_i} M_{Z_i}(\alpha_i\sigma_i t). \end{aligned}$$

We derived before that $M_{Z_i}(u) = e^{u^2/2}$. Hence,

$$M_{\alpha_i X_i}(t) = \exp\left\{\alpha_i \mu_i t + \frac{\alpha_i^2 \sigma_i^2}{2} t^2\right\}.$$

Finally

$$\begin{aligned} M_W(t) &= \prod_{i=1}^{n} M_{\alpha_i X_i}(t) \\ &= \exp\left\{\left(\sum_{i=1}^{n} \alpha_i \mu_i\right) t + \frac{\sum_{i=1}^{n} \alpha_i^2 \sigma_i^2}{2} t^2\right\}. \end{aligned}$$

This implies that the distribution of W is normal, with

$$E\{W\} = \sum_{i=1}^{n} \alpha_i \mu_i$$

and

$$V\{W\} = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2.$$

(iv) If $X_1, X_2, \cdots, X_n$ are independent random variables, having gamma distribution like $G(\nu_i, \beta)$, respectively, $i = 1, \cdots, n$, then the distribution of $T_n = \sum_{i=1}^{n} X_i$ is gamma, like that of $G\left(\sum_{i=1}^{n} \nu_i, \beta\right)$. Indeed,

$$\begin{aligned} M_{T_n}(t) &= \prod_{i=1}^{n} (1 - t\beta)^{-\nu_i} \\ &= (1 - t\beta)^{-\sum_{i=1}^{n} \nu_i}. \end{aligned}$$

■

3.9 Large sample approximations

3.9.1 The law of large numbers

We have shown in Example 3.36 that the mean of a random sample, $\overline{X}_n$, converges in probability to the expected value of X, μ (the population mean). This is the **law of large numbers** (L.L.N.) which states that, if $X_1, X_2, \cdots$ are i.i.d. random variables and $E\{|X_1|\} < \infty$, then for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\{|\overline{X}_n - \mu| > \epsilon\} = 0.$$

We also write

$$\lim_{n\to\infty} \overline{X}_n = \mu, \quad \text{in probability.}$$

This is known as the **weak** L.L.N. There is a stronger law, which states that, under the above conditions,

$$\Pr\{\lim_{n\to\infty} \overline{X}_n = \mu\} = 1.$$

It is beyond the scope of the book to discuss the meaning of the strong L.L.N.

3.9.2 The Central Limit Theorem

The **Central Limit Theorem**, C.L.T., is one of the most important theorems in probability theory. We formulate here the simplest version of this theorem, which is often sufficient for applications. The theorem states that if $\overline{X}_n$ is the sample mean of n i.i.d. random variables, then if the population variance σ^2 is positive and finite, the sampling distribution of $\overline{X}_n$ is approximately normal, as $n \to \infty$. More precisely,

If $X_1, X_2, \cdots$ is a sequence of i.i.d. random variables, with $E\{X_1\} = \mu$ and $V\{X_1\} = \sigma^2$, $0 < \sigma^2 < \infty$, then

$$\lim_{n\to\infty} \Pr\left\{ \frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma} \leq z \right\} = \Phi(z), \tag{3.107}$$

where $\Phi(z)$ is the c.d.f. of $N(0, 1)$.

The proof of this basic version of the C.L.T. is based on a result in probability theory, stating that if $X_1, X_2, \cdots$ is a sequence of random variables having m.g.f.'s, $M_n(T)$, $n = 1, 2, \cdots$ and if $\lim_{n\to\infty} M_n(t) = M(t)$ is the m.g.f. of a random variable X^*, having a c.d.f. $F^*(x)$, then $\lim_{n\to\infty} F_n(x) = F^*(x)$, where $F_n(x)$ is the c.d.f. of X_n.

The m.g.f. of

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma},$$

can be written as

$$\begin{aligned} M_{Z_n}(t) &= E\left\{ \exp\left\{ \frac{t}{\sqrt{n}\sigma} \sum_{i=1}^{n} (X_i - \mu) \right\} \right\} \\ &= \left(E\left\{ \exp\left\{ \frac{t}{\sqrt{n}\sigma} (X_1 - \mu) \right\} \right\} \right)^n, \end{aligned}$$

since the random variables are independent. Furthermore, Taylor expansion of $\exp\left\{ \frac{t}{\sqrt{n}\sigma}(X_1 - \mu) \right\}$ is

$$1 + \frac{t}{\sqrt{n}\sigma}(X_1 - \mu) + \frac{t^2}{2n\sigma^2}(X_1 - \mu)^2 + o\left(\frac{1}{n}\right),$$

for n large. Hence, as $n \to \infty$

$$E\left\{ \exp\left\{ \frac{t}{\sqrt{n}\sigma}(X_1 - \mu) \right\} \right\} = 1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

Hence,

$$\begin{aligned} \lim_{n\to\infty} M_{Z_n}(t) &= \lim_{n\to\infty} \left(1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right)^n \\ &= e^{t^2/2}, \end{aligned}$$

which is the m.g.f. of $N(0, 1)$. This is a sketch of the proof. For rigorous proofs and extensions, see textbooks on probability theory.

3.9.3 Some normal approximations

The C.L.T. can be applied to provide an approximation to the distribution of the sum of n i.i.d. random variables, by a standard normal distribution, when n is large. We list below a few such useful approximations.

(i) **Binomial Distribution**
When n is large, then the c.d.f. of $B(n,p)$ can be approximated by

$$B(k;n,p) \cong \Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right). \tag{3.108}$$

We add $\frac{1}{2}$ to k, in the argument of $\Phi(\cdot)$ to obtain a better approximation when n is not too large. This modification is called, "correction for discontinuity."
How large should n be to get a "good" approximation? A general rule is

$$n > \frac{9}{p(1-p)}. \tag{3.109}$$

(ii) **Poisson Distribution**
The c.d.f. of Poisson with parameter λ can be approximated by

$$P(k;\lambda) \cong \Phi\left(\frac{k+\frac{1}{2}-\lambda}{\sqrt{\lambda}}\right), \tag{3.110}$$

if λ is large (greater than 30).
(iii) **Gamma Distribution**
The c.d.f. of $G(\nu,\beta)$ can be approximated by

$$G(x;\nu,\beta) \cong \Phi\left(\frac{x-\nu\beta}{\beta\sqrt{\nu}}\right), \tag{3.111}$$

for large values of ν.

Example 3.40. (i) A lot consists of $n = 10,000$ screws. The probability that a screw is defective is $p = 0.01$. What is the probability that there are more than 120 defective screws in the lot?

The number of defective screws in the lot, J_n, has a distribution like $B(10000, 0.01)$. Hence,

$$\begin{aligned}\Pr\{J_{10000} > 120\} &= 1 - B(120; 10000,\ .01)\\ &\cong 1 - \Phi\left(\frac{120.5-100}{\sqrt{99}}\right)\\ &= 1 - \Phi(2.06) = 0.0197.\end{aligned}$$

(ii) In the production of industrial film, we find on the average 1 defect per 100 $[\text{ft}]^2$ of film. What is the probability that fewer than 100 defects will be found on 12,000 $[\text{ft}]^2$ of film?

We assume that the number of defects per unit area of film is a Poisson random variable. Thus, our model is that the number of defects, X, per 12,000 $[\text{ft}]^2$ has a Poisson distribution with parameter $\lambda = 120$. Thus,

$$\begin{aligned}\Pr\{X < 100\} &\cong \Phi\left(\frac{99.5-120}{\sqrt{120}}\right)\\ &= 0.0306.\end{aligned}$$

(iii) The time till failure, T, of radar equipment is exponentially distributed with mean time till failure (M.T.T.F.) of $\beta = 100$ [hr].

A sample of $n = 50$ units is put on test. Let $\overline{T}_{50}$ be the sample mean. What is the probability that $\overline{T}_{50}$ will fail in the interval (95, 105) [hr]?

We have seen that $\sum_{i=1}^{50} T_i$ is distributed like $G(50, 100)$, since $E(\beta)$ is distributed like $G(1, \beta)$. Hence $\overline{T}_{50}$ is distributed like $\frac{1}{50}G(50, 100)$ which is $G(50, 2)$. By the normal approximation

$$\Pr\{95 < \overline{T}_{50} < 105\} \cong \Phi\left(\frac{105-100}{2\sqrt{50}}\right) - \Phi\left(\frac{95-100}{2\sqrt{50}}\right) = 2\Phi(0.3536) - 1 = 0.2763.$$

■

3.10 Additional distributions of statistics of normal samples

In the present section we assume that $X_1, X_2, \cdots, X_n$ are i.i.d. $N(\mu, \sigma^2)$ random variables. In the subsections 3.10.1–3.10.3 we present the chi-squared, t- and F-distributions which play an important role in the theory of statistical inference (Chapter 4).

3.10.1 Distribution of the sample variance

Writing $X_i = \mu + \sigma Z_i$, where $Z_1, \cdots, Z_n$ are i.i.d. $N(0, 1)$, we obtain that the sample variance S^2 is distributed like

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(\mu + \sigma Z_i - (\mu + \sigma\overline{Z}_n))^2 \\ &= \frac{\sigma^2}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)^2. \end{aligned}$$

One can show that $\sum_{i=1}^{n}(Z_i - \overline{Z}_n)^2$ is distributed like $\chi^2[n-1]$, where $\chi^2[\nu]$ is called a **chi-squared random variable with ν degrees of freedom**. Moreover, $\chi^2[\nu]$ is distributed like $G\left(\frac{\nu}{2}, 2\right)$.

The α-th quantile of $\chi^2[\nu]$ is denoted by $\chi^2_\alpha[\nu]$. Accordingly, the c.d.f. of the sample variance is

$$\begin{aligned} H_{S^2}(x; \sigma^2) &= \Pr\left\{\frac{\sigma^2}{n-1}\chi^2[n-1] \le x\right\} \\ &= \Pr\left\{\chi^2[n-1] \le \frac{(n-1)x}{\sigma^2}\right\} \\ &= \Pr\left\{G\left(\frac{n-1}{2}, 2\right) \le \frac{(n-1)x}{\sigma^2}\right\} \\ &= G\left(\frac{(n-1)x}{2\sigma^2}; \frac{n-1}{2}, 1\right). \end{aligned} \tag{3.112}$$

The probability values of the distribution of $\chi^2[\nu]$, as well as the α-quantiles, can be computed by R, MINITAB or JMP, or read from appropriate tables.

The expected value and variance of the sample variance are

$$E\{S^2\} = \frac{\sigma^2}{n-1}E\{\chi^2[n-1]\}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{n-1} E\left\{G\left(\frac{n-1}{2}, 2\right)\right\} \\
&= \frac{\sigma^2}{n-1} \cdot (n-1) = \sigma^2.
\end{aligned}
$$

Similarly

$$
\begin{aligned}
V\{S^2\} &= \frac{\sigma^4}{(n-1)^2} V\{\chi^2[n-1]\} \\
&= \frac{\sigma^4}{(n-1)^2} V\left\{G\left(\frac{n-1}{2}, 2\right)\right\} \\
&= \frac{\sigma^4}{(n-1)^2} \cdot 2(n-1) \\
&= \frac{2\sigma^4}{n-1}.
\end{aligned}
\tag{3.113}
$$

Thus applying the Chebychev's inequality, for any given $\epsilon > 0$,

$$
\Pr\{|S^2 - \sigma^2| > \epsilon\} < \frac{2\sigma^4}{(n-1)\epsilon^2}.
$$

Hence, S^2 converges in probability to σ^2. Moreover,

$$
\lim_{n\to\infty} \Pr\left\{\frac{(S^2 - \sigma^2)}{\sigma^2\sqrt{2}}\sqrt{n-1} \le z\right\} = \Phi(z). \tag{3.114}
$$

That is, the distribution of S^2 can be approximated by the normal distributions in large samples.

3.10.2 The "Student" *t*-statistic

We have seen that

$$
Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}
$$

has a $N(0, 1)$ distribution. As we will see in Chapter 6, when σ is unknown, we test hypotheses concerning μ by the statistic

$$
t = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S},
$$

where S is the sample standard deviation. If $X_1, \cdots, X_n$ are i.i.d. like $N(\mu_0, \sigma^2)$ then the distribution of t is called the **Student *t*-distribution with $\nu = n - 1$ degrees of freedom**. The corresponding random variable is denoted by $t[\nu]$.

The p.d.f. of $t[\nu]$ is symmetric about 0 (see Figure 3.19). Thus,

$$
E\{t[\nu]\} = 0, \quad \text{for} \quad \nu \ge 2 \tag{3.115}
$$

and

$$
V\{t[\nu]\} = \frac{\nu}{\nu - 2}, \quad \nu \ge 3. \tag{3.116}
$$

The α-quantile of $t[\nu]$ is denoted by $t_\alpha[\nu]$. It can be read from a table, or determined by R, JMP or MINITAB.

3.10.3 Distribution of the variance ratio

$$
F = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}.
$$

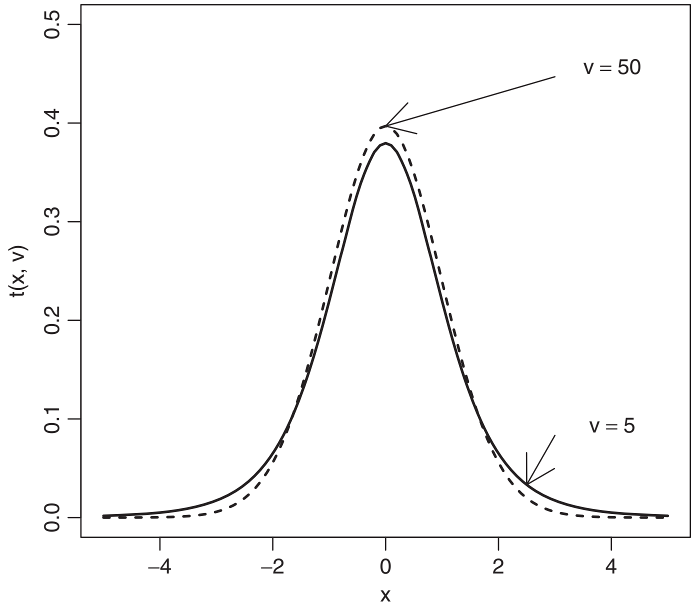

Figure 3.19 *Density Functions of* $t[\nu]$, $\nu = 5, 50$

Consider now two independent samples of size n_1 and n_2, respectively, which have been taken from normal populations having variances σ_1^2 and σ_2^2. Let

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \overline{X}_1)^2$$

and

$$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \overline{X}_2)^2$$

be the variances of the two samples where $\overline{X}_1$ and $\overline{X}_2$ are the corresponding sample means. The F-ratio has a distribution denoted by $F[\nu_1, \nu_2]$, with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. This distribution is called the **F-distribution with ν_1 and ν_2 degrees of freedom**. A graph of the densities of $F[\nu_1, \nu_2]$ is given in Figure 3.20.

The expected value and the variance of $F[\nu_1, \nu_2]$ are:

$$E\{F[\nu_1, \nu_2]\} = \nu_2/(\nu_2 - 2), \quad \nu_2 > 2, \tag{3.117}$$

and

$$V\{F[\nu_1, \nu_2]\} = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \quad \nu_2 > 4. \tag{3.118}$$

The $(1 - \alpha)$th quantile of $F[\nu_1, \nu_2]$, i.e. $F_{1-\alpha}[\nu_1, \nu_2]$, can be computed by MINITAB. If we wish to obtain the α-th fractile $F_\alpha[\nu_1, \nu_2]$ for values of $\alpha < .5$, we can apply the relationship:

$$F_{1-\alpha}[\nu_1, \nu_2] = \frac{1}{F_\alpha[\nu_2, \nu_1]}. \tag{3.119}$$

Thus, for example, to compute $F_{.05}[15, 10]$, we write

$$F_{.05}[15, 10] = 1/F_{.95}[10, 15] = 1/2.54 = .3937.$$

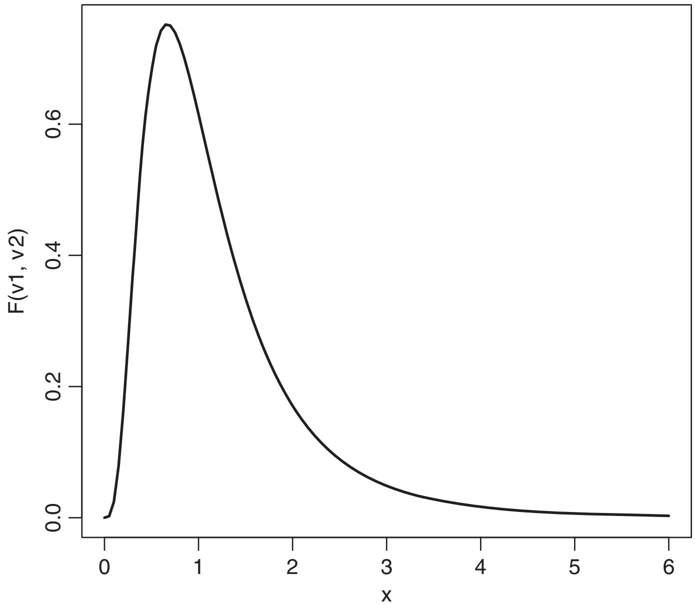

Figure 3.20 *Density function of* $F(\nu_1, \nu_2)$

3.11 Chapter highlights

The chapter provides the basics of probability theory and of the theory of distribution functions. The probability model for random sampling is discussed. This is fundamental for the sampling procedures to be discussed in Chapter 7. Bayes theorem has important ramifications in statistical inference, as will be discussed in Chapter 4. The concepts and definitions introduced are:

- Sample Space
- Elementary Events
- Operations with Events
- Disjoint Events
- Probability of Events
- Random Sampling With Replacement (RSWR)
- Random Sampling Without Replacement (RSWOR)
- Conditional Probabilities
- Independent Events
- Bayes' Formula
- Prior Probability
- Posterior Probability
- Probability Distribution Function (p.d.f.)
- Discrete Random Variable
- Continuous Random Variable
- Cumulative Distribution Function
- Central Moments
- Expected Value
- Standard Deviation
- Chebychev's Inequality

- Moment Generating Function
- Skewness
- Kurtosis
- Independent Trials
- *P*-th Quantile
- Joint Distribution
- Marginal Distribution
- Conditional Distribution
- Mutual Independence
- Conditional Independence
- Law of Total Variance
- Law of Iterated Expectation
- Order Statistics
- Convergence in Probability
- Central Limit Theorem
- Law of Large Numbers

3.12 Exercises

3.1 An experiment consists of making 20 observations on the quality of chips. Each observation is recorded as G or D.
(i) What is the sample space, S, corresponding to this experiment?
(ii) How many elementary events in S?
(iii) Let A_n, $n = 0, \cdots, 20$, be the event that exactly n G observations are made. Write the events A_n formally. How many elementary events belong to A_n?

3.2 An experiment consists of 10 measurements $w_1, \cdots, w_{10}$ of the weights of packages. All packages under consideration have weights between 10 and 20 pounds. What is the sample space S? Let $A = \{(w_1, w_2, \cdots, w_{10}) : w_1 + w_2 = 25\}$. Let $B = \{(w_1, \cdots, w_{10}) : w_1 + w_2 \leq 25\}$. Describe the events A and B graphically. Show that $A \subset B$.

3.3 Strings of 30 binary (0, 1) signals are transmitted.
(i) Describe the sample space, S.
(ii) Let A_{10} be the event that the first 10 signals transmitted are all 1's. How many elementary events belong to A_{10}?
(iii) Let B_{10} be the event that exactly 10 signals, out of 30 transmitted, are 1's. How many elementary events belong to B_{10}? Does $A_{10} \subset B_{10}$?

3.4 Prove DeMorgan laws
(i) $(A \cup B)^c = A^c \cap B^c$.
(ii) $(A \cap B)^c = A^c \cup B^c$.

3.5 Consider Exercise [3.3] Show that the events $A_0, A_1, \cdots, A_{20}$ are a partition of the sample space S.

3.6 Let $A_1, \cdots, A_n$ be a partition of S. Let B be an event. Show that $B = \bigcup_{i=1}^{n} A_i B$, where $A_i B = A_i \cap B$, is a union of disjoint events.

3.7 Develop a formula for the probability $\Pr\{A \cup B \cup C\}$, where A, B, C are arbitrary events.

3.8 Show that if $A_1, \cdots, A_n$ is a partition, then for any event B, $P\{B\} = \sum_{i=1}^{n} P\{A_i B\}$. [Use the result of [3.6].]

3.9 An unbiased die has the numbers $1, 2, \cdots, 6$ written on its faces. The die is thrown twice. What is the probability that the two numbers shown on its upper face sum up to 10?

3.10 The time till failure, T, of electronic equipment is a random quantity. The event $A_t = \{T > t\}$ is assigned the probability $\Pr\{A_t\} = \exp\ \{-t/200\}$, $t \geq 0$. What is the probability of the event $B = \{150 < T < 280\}$?

3.11 A box contains 40 parts, 10 of type A, 10 of type B, 15 of type C and 5 of type D. A random sample of 8 parts is drawn without replacement. What is the probability of finding two parts of each type in the sample?

3.12 How many samples of size $n = 5$ can be drawn from a population of size $N = 100$,
(i) with replacement?
(ii) without replacement?

3.13 A lot of 1,000 items contain $M = 900$ "good" ones, and 100 "defective" ones. A random sample of size $n = 10$ is drawn from the lot. What is the probability of observing in the sample at least 8 good items,
(i) when sampling is with replacement?
(ii) when sampling is without replacement?

3.14 In continuation of the previous exercise, what is the probability of observing in an RSWR at least one defective item?

3.15 Consider the problem of Exercise [3.10]. What is the conditional probability $\Pr\{T > 300 \mid T > 200\}$.

3.16 A point (X, Y) is chosen at random within the unit square, i.e.

$$S = \{(x, y) : 0 \le x, y \le 1\}.$$

Any set A contained in S having area given by

$$\text{Area}\{A\} = \int\int_A dxdy$$

is an event, whose probability is the area of A. Define the events

$$B = \left\{(x, y) : x > \frac{1}{2}\right\}$$
$$C = \{(x, y) : x^2 + y^2 \le 1\}$$
$$D = \{(x, y) : (x + y) \le 1\}.$$

(i) Compute the conditional probability $\Pr\{D \mid B\}$.
(ii) Compute the conditional probability $\Pr\{C \mid D\}$.

3.17 Show that if A and B are independent events, then A^c and B^c are also independent events.

3.18 Show that if A and B are disjoint events, then A and B are dependent events.

3.19 Show that if A and B are independent events, then

$$\begin{aligned}\Pr\{A \cup B\} &= \Pr\{A\}(1 - \Pr\{B\}) + \Pr\{B\}\\ &= \Pr\{A\} + \Pr\{B\}(1 - \Pr\{A\}).\end{aligned}$$

3.20 A machine which tests whether a part is defective, D, or good, G, may err. The probabilities of errors are given by

$$\Pr\{A \mid G\} = .95,$$
$$\Pr\{A \mid D\} = .10,$$

where A is the event "the part is considered G after testing." If $\Pr\{G\} = .99$, what is the probability of D given A?

Additional problems in combinatorial and geometric probabilities

3.21 Assuming 365 days in a year, if there are 10 people in a party, what is the probability that their birthdays fall on different days? Show that if there are more than 22 people in the party, the probability is greater than 1/2 that at least 2 will have birthdays on the same day.

3.22 A number is constructed at random by choosing 10 digits from $\{0, \ldots, 9\}$ with replacement. We allow the digit 0 at any position. What is the probability that the number does not contain 3 specific digits?

3.23 A caller remembers all the 7 digits of a telephone number, but is uncertain about the order of the last four. He keeps dialing the last four digits at random, without repeating the same number, until he reaches the right number. What is the probability that he will dial at least ten wrong numbers?

3.24 One hundred lottery tickets are sold. There are four prizes and ten consolation prizes. If you buy 5 tickets, what is the probability that you win:

(i) one prize?
(ii) a prize and a consolation prize?
(iii) something?

3.25 Ten PCB's are in a bin, two of these are defectives. The boards are chosen at random, one by one, without replacement. What is the probability that exactly five good boards will be found between the drawing of the first and second defective PCB?

3.26 A random sample of 11 integers is drawn without replacement from the set $\{1, 2, \ldots, 20\}$. What is the probability that the sample median, Me, is equal to the integer k? $6 \le k \le 15$.

3.27 A stick is broken at random into three pieces. What is the probability that these pieces can be the sides of a triangle?

3.28 A particle is moving at a uniform speed on a circle of unit radius and is released at a random point on the circumference. Draw a line segment of length $2h$ ($h < 1$) centered at a point A of distance $a > 1$ from the center of the circle, O. Moreover the line segment is perpendicular to the line connecting O with A. What is the probability that the particle will hit the line segment? [The particle flies along a straight line tangential to the circle.]

3.29 A block of 100 bits is transmitted over a binary channel, with probability $p = 10^{-3}$ of bit error. Errors occur independently. Find the probability that the block contains at least three errors.

3.30 A coin is tossed repeatedly until 2 "heads" occur. What is the probability that 4 tosses are required?

3.31 Consider the sample space S of all sequences of 10 binary numbers (0-1 signals). Define on this sample space two random variables and derive their probability distribution function, assuming the model that all sequences are equally probable.

3.32 The number of blemishes on a ceramic plate is a discrete random variable. Assume the probability model, with p.d.f.

$$p(x) = e^{-5} \frac{5^x}{x!}, \quad x = 0, 1, \cdots$$

(i) Show that $\sum_{x=0}^{\infty} p(x) = 1$
(ii) What is the probability of at most 1 blemish on a plate?
(iii) What is the probability of no more than 7 blemishes on a plate?

3.33 Consider a distribution function of a mixed type with c.d.f.

$$F_x(x) = \begin{cases} 0, & \text{if } x < -1 \\ .3 + .2(x+1), & \text{if } -1 \le x < 0 \\ .7 + .3x, & \text{if } 0 \le x < 1 \\ 1, & \text{if } 1 \le x. \end{cases}$$

(i) What is $\Pr\{X = -1\}$?
(ii) What is $\Pr\{-.5 < X < 0\}$?
(iii) What is $\Pr\{0 \le X < .75\}$?
(iv) What is $\Pr\{X = 1\}$?
(v) Compute the expected value, $E\{X\}$ and variance, $V\{X\}$.

3.34 A random variable has the Rayleigh distribution, with c.d.f.

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x^2/2\sigma^2}, & x \ge 0 \end{cases}$$

where σ^2 is a positive parameter. Find the expected value $E\{X\}$.

3.35 A random variable X has a discrete distribution over the integers $\{1, 2, \ldots, N\}$ with equal probabilities. Find $E\{X\}$ and $V\{X\}$.

3.36 A random variable has expectation $\mu = 10$ and standard deviation $\sigma = 0.5$. Use Chebychev's inequality to find a lower bound to the probability

$$\Pr\{8 < X < 12\}.$$

3.37 Consider the random variable X with c.d.f.

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x), \quad -\infty < x < \infty.$$

Find the .25th, .50th and .75th quantiles of this distribution.

3.38 Show that the central moments μ_l^* relate to the moments μ_l around the origin, by the formula

$$\mu_l^* = \sum_{j=0}^{l-2} (-1)^j \binom{l}{j} \mu_{l-j}\mu_1^j + (-1)^{l-1}(l-1)\mu_1^l.$$

3.39 Find the expected value μ_1 and the second moment μ_2 of the random variable whose c.d.f. is given in Exercise [3.33].

3.40 A random variable X has a continuous uniform distribution over the interval (a, b), i.e.,

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

Find the moment generating function of X. Find the mean and variance by differentiating the m.g.f.

3.41 Consider the moment generating function, m.g.f. of the exponential distribution, that is,

$$M(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

(i) Find the first four moments of the distribution, by differentiating $M(t)$.
(ii) Convert the moments to central moments.
(iii) What is the index of kurtosis β_4?

3.42 Using R, MINITAB or JMP prepare a table of the p.d.f. and c.d.f. of the binomial distribution $B(20, .17)$.

3.43 What are the 1st quantile, Q_1, median, Me, and 3rd quantile, Q_3, of $B(20, .17)$?

3.44 Compute the mean $E\{X\}$ and standard deviation, σ, of $B(45, .35)$.

3.45 A PCB is populated by 50 chips which are randomly chosen from a lot. The probability that an individual chip is non-defective is p. What should be the value of p so that no defective chip is installed on the board is $\gamma = .99$? [The answer to this question shows why the industry standards are so stringent.]

3.46 Let $b(j; n, p)$ be the p.d.f. of the binomial distribution. Show that as $n \to \infty$, $p \to 0$ so that $np \to \lambda$, $0 < \lambda < \infty$, then

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} b(j; n, p) = e^{-\lambda}\frac{\lambda^j}{j!}, \quad j = 0, 1, \ldots.$$

3.47 Use the result of the previous exercise to find the probability that a block of 1,000 bits, in a binary communication channel, will have less than 4 errors, when the probability of a bit error is $p = 10^{-3}$.

3.48 Compute $E\{X\}$ and $V\{X\}$ of the hypergeometric distribution $H(500, 350, 20)$.

3.49 A lot of size $N = 500$ items contains $M = 5$ defective ones. A random sample of size $n = 50$ is drawn from the lot without replacement (RSWOR). What is the probability of observing more than 1 defective item in the sample?

3.50 Consider Example 3.23. What is the probability that the lot will be rectified if $M = 10$ and $n = 20$?

3.51 Use the m.g.f. to compute the third and fourth central moments of the Poisson distribution $P(10)$. What is the index of skewness and kurtosis of this distribution?

3.52 The number of blemishes on ceramic plates has a Poisson distribution with mean $\lambda = 1.5$. What is the probability of observing more than 2 blemishes on a plate?

3.53 The error rate of an insertion machine is 380 PPM (per 10^6 parts inserted). What is the probability of observing more than 6 insertion errors in 2 hours of operation, when the insertion rate is 4,000 parts per hour?

3.54 In continuation of the previous exercise, let N be the number of parts inserted until an error occurs. What is the distribution of N? Compute the expected value and the standard deviation of N.

3.55 What are Q_1, Me and Q_3 of the negative binomial N.B. (p, k) with $p = 0.01$ and $k = 3$?

3.56 Derive the m.g.f. of N.B. (p, k).

3.57 Differentiate the m.g.f. of the geometric distribution, i.e.,

$$M(t) = \frac{pe^t}{(1 - e^t(1 - p))}, \quad t < -\log\ (1 - p),$$

to obtain its first four moments, and derive then the indices of skewness and kurtosis.

3.58 The proportion of defective RAM chips is $p = 0.002$. You have to install 50 chips on a board. Each chip is tested before its installation. How many chips should you order so that, with probability greater than $\gamma = .95$ you will have at least 50 good chips to install?

3.59 The random variable X assumes the values $\{1, 2, \ldots\}$ with probabilities of a geometric distribution, with parameter p, $0 < p < 1$. Prove the "memoryless" property of the geometric distribution, namely:

$$P[X > n + m \mid X > m] = P[X > n],$$

for all $n, m = 1, 2, \ldots$.

3.60 Let X be a random variable having a continuous c.d.f. $F(x)$. Let $Y = F(X)$. Show that Y has a uniform distribution on $(0, 1)$. Conversely, if U has a uniform distribution on $(0, 1)$ then $X = F^{-1}(U)$ has the c.d.f. $F(x)$.

3.61 Compute the expected value and the standard deviation of a uniform distribution $U(10, 50)$.

3.62 Show that if U is uniform on $(0, 1)$ then $X = -\log\ (U)$ has an exponential distribution $E(1)$.

3.63 Use R, MINITAB or JMP to compute the probabilities, for $N(100, 15)$, of
(i) $92 < X < 108$.
(ii) $X > 105$.
(iii) $2X + 5 < 200$.

3.64 The .9-quantile of $N(\mu, \sigma)$ is 15 and its .99-quantile is 20. Find the mean μ and standard deviation σ.

3.65 A communication channel accepts an arbitrary voltage input v and outputs a voltage $v + E$, where $E \sim N(0, 1)$. The channel is used to transmit binary information as follows:

$$\text{to transmit } 0, \text{input } -v$$

$$\text{to transmit } 1, \text{input } v$$

The receiver decides a 0 if the voltage Y is negative, and 1 otherwise. What should be the value of v so that the receiver's probability of bit error is $\alpha = .01$?

3.66 Aluminum pins manufactured for an aviation industry have a random diameter, whose distribution is (approximately) normal with mean of $\mu = 10$ [mm] and standard deviation $\sigma = 0.02$ [mm]. Holes are automatically drilled on aluminum plates, with diameters having a normal distribution with mean μ_d [mm] and $\sigma = 0.02$ [mm]. What should be the value of μ_d so that the probability that a pin will not enter a hole (too wide) is $\alpha = 0.01$?

3.67 Let $X_1, \ldots, X_n$ be a random sample (i.i.d.) from a normal distribution $N(\mu, \sigma^2)$. Find the expected value and variance of $Y = \sum_{i=1}^{n} iX_i$.

3.68 Concrete cubes have compressive strength with log-normal distribution LN(5, 1). Find the probability that the compressive strength X of a random concrete cube will be greater than 300 [kg/cm^2].

3.69 Using the m.g.f. of $N(\mu, \sigma)$, derive the expected value and variance of $LN\ (\mu, \sigma)$. [Recall that $X \sim e^{N(\mu,\sigma)}$.]

3.70 What are Q_1, Me and Q_3 of $E(\beta)$?

3.71 Show that if the life length of a chip is exponential $E(\beta)$, then only 36.7% of the chips will function longer than the mean time till failure β.

3.72 Show that the m.g.f. of $E(\beta)$ is $M(t) = (1 - \beta t)^{-1}$, for $t < \frac{1}{\beta}$.

3.73 Let X_1, X_2, X_3 be independent random variables having an identical exponential distribution $E(\beta)$. Compute $\Pr\{X_1 + X_2 + X_3 \geq 3\beta\}$.

3.74 Establish the formula

$$G(t; k, \frac{1}{\lambda}) = 1 - e^{-\lambda t} \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!},$$

by integrating in parts the p.d.f. of

$$G\left(k; \frac{1}{\lambda}\right).$$

3.75 Use R to compute $\Gamma(1.17)$, $\Gamma\left(\frac{1}{2}\right)$, $\Gamma\left(\frac{3}{2}\right)$.

3.76 Using m.g.f., show that the sum of k independent exponential random variables, $E(\beta)$, has the gamma distribution $G(k, \beta)$.

3.77 What is the expected value and variance of the Weibull distribution $W(2, 3.5)$?

3.78 The time till failure (days) of an electronic equipment has the Weibull distribution $W(1.5, 500)$. What is the probability that the failure time will not be before 600 days?

3.79 Compute the expected value and standard deviation of a random variable having the beta distribution Beta$\left(\frac{1}{2}, \frac{3}{2}\right)$.

3.80 Show that the index of kurtosis of Beta(ν, ν) is $\beta_2 = \frac{3(1+2\nu)}{3+2\nu}$.

3.81 The joint p.d.f. of two random variables (X, Y) is

$$f(x,y) = \begin{cases} \frac{1}{2}, & \text{if } (x,y) \in S \\ 0, & \text{otherwise} \end{cases}$$

where S is a square of area 2, whose vertices are $(1, 0)$, $(0, 1)$, $(-1, 0)$, $(0, -1)$.
(i) Find the marginal p.d.f. of X and of Y.
(ii) Find $E\{X\}$, $E\{Y\}$, $V\{X\}$, $V\{Y\}$.

3.82 Let (X, Y) have a joint p.d.f.

$$f(x,y) = \begin{cases} \frac{1}{y} \exp\left\{-y - \frac{x}{y}\right\}, & \text{if } 0 < x, y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

Find COV(X, Y) and the coefficient of correlation ρ_{XY}.

3.83 Show that the random variables (X, Y) whose joint distribution is defined in Example 3.26 are dependent. Find COV(X, Y).

3.84 Find the correlation coefficient of N and J of Example 3.30.

3.85 Let X and Y be independent random variables, $X \sim G(2, 100)$ and $W(1.5, 500)$. Find the variance of XY.

3.86 Consider the trinomial distribution of Example 3.32.
(i) What is the probability that during one hour of operation there will be no more than 20 errors?
(ii) What is the conditional distribution of wrong components, given that there are 15 misinsertions in a given hour of operation?
(iii) Approximating the conditional distribution of (ii) by a Poisson distribution, compute the conditional probability of no more than 15 wrong components.

3.87 In continuation of Example 3.33, compute the correlation between J_1 and J_2.

3.88 In a bivariate normal distribution, the conditional variance of Y given X is 150 and the variance of Y is 200. What is the correlation ρ_{XY}?

3.89 $n = 10$ electronic devices start to operate at the same time. The times till failure of these devices are independent random variables having an identical $E(100)$ distribution.

(i) What is the expected value of the first failure?
(ii) What is the expected value of the last failure?

3.90 A factory has $n = 10$ machines of a certain type. At each given day, the probability is $p = .95$ that a machine will be working. Let J denote the number of machines that work on a given day. The time it takes to produce an item on a given machine is $E(10)$, i.e., exponentially distributed with mean $\mu = 10$ [min]. The machines operate independently of each other. Let $X_{(1)}$ denote the minimal time for the first item to be produced. Determine
(i) $P[J = k, X_{(1)} \leq x]$, $k = 1, 2, \ldots$
(ii) $P[X_{(1)} \leq x \mid J \geq 1]$.
Notice that when $J = 0$, no machine is working. The probability of this event is $(0.05)^{10}$.

3.91 Let $X_1, X_2, \ldots, X_{11}$ be a random sample of exponentially distributed random variables with p.d.f. $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.
(i) What is the p.d.f. of the median $Me = X_{(6)}$?
(ii) What is the expected value of Me?

3.92 Let X and Y be independent random variables having an $E(\beta)$ distribution. Let $T = X + Y$ and $W = X - Y$. Compute the variance of $T + \frac{1}{2}W$.

3.93 Let X and Y be independent random variables having a common variance σ^2. What is the covariance $\mathrm{cov}(X, X + Y)$?

3.94 Let (X, Y) have a bivariate normal distribution. What is the variance of $\alpha X + \beta Y$?

3.95 Let X have a normal distribution $N(\mu, \sigma)$. Let $\Phi(z)$ be the standard normal c.d.f. Verify that $E\{\Phi(X)\} = P\{U < X\}$, where U is independent of X and $U \sim N(0, 1)$. Show that

$$E\{\Phi(X)\} = \Phi\left(\frac{\eta}{\sqrt{1+\sigma^2}}\right).$$

3.96 Let X have a normal distribution $N(\mu, \sigma)$. Show that

$$E\{\Phi^2(X)\} = \Phi_2\left(\frac{\mu}{\sqrt{1+\sigma^2}}, \frac{\mu}{\sqrt{1+\sigma^2}}; \frac{\sigma^2}{1+\sigma^2}\right).$$

3.97 X and Y are independent random variables having Poisson distributions, with means $\lambda_1 = 5$ and $\lambda_2 = 7$, respectively. Compute the probability that $X + Y$ is greater than 15.

3.98 Let X_1 and X_2 be independent random variables having continuous distributions with p.d.f. $f_1(x)$ and $f_2(x)$, respectively. Let $Y = X_1 + X_2$. Show that the p.d.f. of Y is

$$g(y) = \int_{-\infty}^{\infty} f_1(x) f_2(y - x) dx.$$

[This integral transform is called the convolution of $f_1(x)$ with $f_2(x)$. The convolution operation is denoted by $f_1 * f_2$.]

3.99 Let X_1 and X_2 be independent random variables having the uniform distributions on $(0, 1)$. Apply the convolution operation to find the p.d.f. of $Y = X_1 + X_2$.

3.100 Let X_1 and X_2 be independent random variables having a common exponential distribution $E(1)$. Determine the p.d.f. of $U = X_1 - X_2$. [The distribution of U is called bi-exponential or Laplace and its p.d.f. is $f(u) = \frac{1}{2}e^{-|u|}$.]

3.101 Apply the central limit theorem to approximate $P\{X_1 + \cdots + X_{20} \leq 50\}$, where $X_1, \cdots, X_{20}$ are independent random variables having a common mean $\mu = 2$ and a common standard deviation $\sigma = 10$.

3.102 Let X have a binomial distribution $B(200, .15)$. Find the normal approximation to $\Pr\{25 < X < 35\}$.

3.103 Let X have a Poisson distribution with mean $\lambda = 200$. Find, approximately, $\Pr\{190 < X < 210\}$.

3.104 $X_1, X_2, \cdots, X_{200}$ are 200 independent random variables have a common beta distribution $B(3, 5)$. Approximate the probability $\Pr\{|\overline{X}_{200} - .375| < 0.2282\}$, where

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad n = 200.$$

3.105 Use R, MINITAB or JMP to compute the .95-quantiles of $t[10]$, $t[15]$, $t[20]$.
3.106 Use R, MINITAB or JMP to compute the .95-quantiles of $F[10, 30]$, $F[15, 30]$, $F[20, 30]$.
3.107 Show that, for each $0 < \alpha < 1$, $t^2_{1-\alpha/2}[n] = F_{1-\alpha}[1, n]$.
3.108 Verify the relationship

$$F_{1-\alpha}[\nu_1, \nu_2] = \frac{1}{F_\alpha[\nu_2, \nu_1]}, \quad 0 < \alpha < 1,$$

$\nu_1, \nu_2 = 1, 2, \cdots$.
3.109 Verify the formula

$$V\{t[\nu]\} = \frac{\nu}{\nu - 2}, \quad \nu > 2.$$

3.110 Find the expected value and variance of $F[3, 10]$.