CHAPTER 1

Review of Probability Theory

1.1 INTRODUCTION

Probability theory provides a rational and efficient means of characterizing the uncertainty which is prevalent in geotechnical engineering. This chapter summarizes the background, fundamental axioms, and main results constituting modern probability theory. Common discrete and continuous distributions are discussed in the last sections of the chapter.

1.2 BASIC SET THEORY

1.2.1 Sample Spaces and Events

When a system is random and is to be modeled as such, the first step in the model is to decide what all of the possible states (outcomes) of the system are. For example, if the load on a retaining wall is being modeled as being random, the possible load can range anywhere from zero to infinity, at least conceptually (while a zero load is entirely possible, albeit unlikely, an infinite load is unlikely—we shall see shortly that the likelihood of an infinite load can be set to be appropriately small). Once the complete set of possible states has been decided on, interest is generally focused on probabilities associated with certain portions of the possible states. For example, it may be of interest to determine the probability that the load on the wall exceeds the sliding resistance of the wall base, so that the wall slides outward. This translates into determining the probability associated with some portion, or *subset*, of the total range of possible wall loads (we are assuming, for the time being, that the base sliding resistance is *known*). These ideas motivate the following definitions:

Definitions

Experiment: Any process that generates a set of data. The experiment may be, for example, the monitoring of the volume of water passing through an earth dam in a unit time. The volume recorded becomes the data set.

Sample Space: The set of all possible outcomes of an experiment. The sample space is represented by the symbol S.

Sample Point: An outcome in the sample space. For example, if the experiment consists of monitoring the volume of water passing through an earth dam per hour, a sample point would be the observation 1.2 m^3/h. Another would be the observation 1.41 m^3/h.

Event: A subset of a sample space. Events will be denoted using uppercase letters, such as $A, B, \ldots$. For example, we might define A to be the event that the flow rate through an earth dam is greater than 0.01 m^3/h.

Null Set: The empty set, having no elements, is used to represent the impossible "event" and is denoted $\emptyset$. For example, the event that the flow rate through an earth dam is both less than 1 and greater than 5 m^3/h is impossible and so the event is the null set.

These ideas will be illustrated with some simple examples.

Example 1.1 Suppose an experiment consists of observing the results of two static pile capacity tests. Each test is considered to be a success (1) if the pile capacity exceeds a certain design criterion and a failure (0) if not. This is an *experiment* since a set of data is derived from it. The actual data derived depend on what is of interest. For example:

1. Suppose that only the number of successful pile tests is of interest. The *sample space* would then be $S = \{0, 1, 2\}$. The elements 0, 1, and 2 of the set S are *sample points*. From this sample space, the following events (which may be of interest) can be defined; $\emptyset$, $\{0\}$, $\{1\}$, $\{2\}$, $\{0, 1\}$, $\{0, 2\}$, $\{1, 2\}$, and $S = \{0, 1, 2\}$ are possible events. The *null set* is used to denote all impossible events (for example, the event that the number of successful tests, out of two tests, is greater than 2).
2. Suppose that the order of occurrence of the successes and failures is of interest. The sample space would then be $S = \{11, 10, 01, 00\}$. Each outcome is a doublet depicting the sequence. Thus, the elements 11, 10, 01, and 00 of S are sample points. The possible events are $\emptyset$, $\{11\}$, $\{10\}$, $\{01\}$, $\{00\}$, $\{11, 10\}$, $\{11, 01\}$, $\{11, 00\}$, $\{10, 01\}$, $\{10, 00\}$, $\{01, 00\}$, $\{11, 10, 01\}$, $\{11, 10, 00\}$, $\{11, 01, 00\}$, $\{10, 01, 00\}$, and $\{11, 10, 01, 00\}$.

Note that the information in 1 could be recovered from that in 2, but not vice versa, so it is often useful to

define the experiment to be more general initially, when possible. Other types of events can then be derived after the experiment is completed.

Sample spaces may be either discrete or continuous:

Discrete Case: In this case, the sample space consists of a sequence of discrete values (e.g., $0, 1, \ldots$). For example, the number of blow counts in a standard penetration test (SPT). Conceptually, this could be any integer number from zero to infinity.

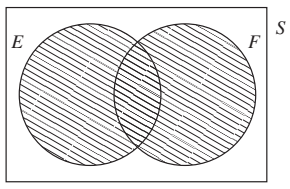
Continuous Case: In this case, the sample space is composed of a continuum of sample points and the number of sample points is effectively always infinite—for example, the elastic modulus of a soil sample. This could be any real number on the positive real line.

1.2.2 Basic Set Theory

The relationship between events and the corresponding sample space can often be illustrated graphically by means of a *Venn diagram*. In a Venn diagram the sample space is represented as a rectangle and events are (usually) drawn as circles inside the rectangle. For example, see Figure 1.1, where A_1, A_2, and A_3 are events in the sample space S.

We are often interested in probabilities associated with combinations of events; for example, the probability that a cone penetration test (CPT) sounding has tip resistance greater than x at the same time as the side friction is less that y. Such events will be formed as subsets of the sample space (and thus are sets themselves). We form these subsets using *set operators*. The union, intersection, and complement are set theory operators which are defined as follows:

The *union* of two events E and F is denoted $E \cup F$.



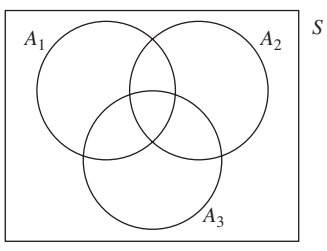
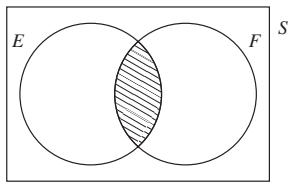
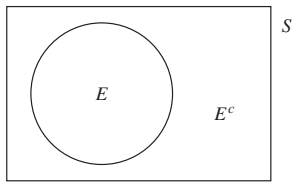


Figure 1.1 Simple Venn diagram.

The *intersection* of two events E and F is denoted $E \cap F$.



The *complement* of an event E is denoted E^c.



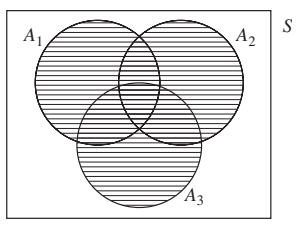
Two events E and F are said to be *mutually exclusive*, or *disjoint*, if $E \cap F = \emptyset$. For example, E and E^c are disjoint events.

Example 1.2 Three piles are being statically loaded to failure. Let A_i denote the event that the ith pile has a capacity exceeding specifications. Using only sets and set theory operators (i.e., using only $A_i, i = 1, 2, 3$, and $\cap$, $\cup$, and ^{c}), describe each of the following events. In each case, also draw a Venn diagram and shade the region corresponding to the event.

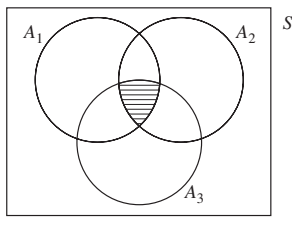
1. At least one pile has capacity exceeding the specification.
2. All three piles have capacities exceeding the specification.
3. Only the first pile has capacity exceeding the specification.
4. Exactly one pile has capacity exceeding the specification.
5. Either only the first pile or only both of the other piles have capacities exceeding the specification.

SOLUTION

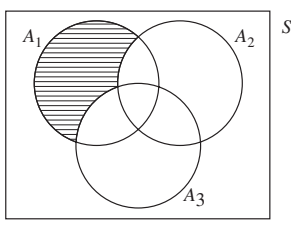
1. $A_1 \cup A_2 \cup A_3$



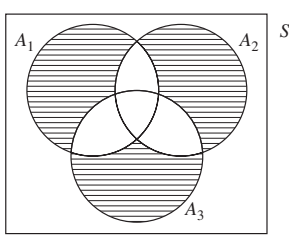
2. $A_1 \cap A_2 \cap A_3$



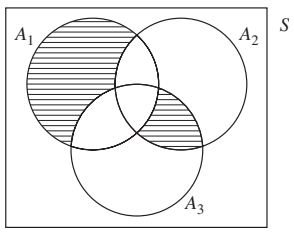
3. $A_1 \cap A_2^c \cap A_3^c$



4. $(A_1 \cap A_2^c \cap A_3^c)$
   $\cup (A_1^c \cap A_2 \cap A_3^c)$
   $\cup (A_1^c \cap A_2^c \cap A_3)$



5. $(A_1 \cap A_2^c \cap A_3^c)$
   $\cup (A_1^c \cap A_2 \cap A_3)$



It is clear from the Venn diagram that, for example, $A_1 \cap A_2^c \cap A_3^c$ and $A_1^c \cap A_2 \cap A_3$ are disjoint events, that is, $(A_1 \cap A_2^c \cap A_3^c) \cap (A_1^c \cap A_2 \cap A_3) = \emptyset$.

1.2.3 Counting Sample Points

Consider experiments which have a finite number of possible outcomes. For example, out of a group of piles, we could have three failing to meet specifications but cannot have 3.24 piles failing to meet specifications. That is, the sample space, in this case, consists of only whole numbers. Such sample spaces are called *discrete* sample spaces. We are often interested in computing the probability associated with each possible value in the sample space. For example, we may want to be able to compute the probability that exactly three piles fail to meet specifications at a site. While it is not generally easy to assign probabilities to something like the number of soft soil lenses at a site, some discrete sample spaces consist of *equi-likely* outcomes, where all possible outcomes have the same probability of occurrence. In this case, we only need to know the total number of possible outcomes in order to assign probabilities to individual outcomes (i.e., the probability of each outcome is equal to 1 over the total number of possible outcomes). Knowing the total number of possible outcomes is often useful, so some basic *counting* rules will be considered here.

Multiplication Rule The fundamental principle of counting, often referred to as the multiplication rule, is:

If an operation can be performed in n_1 ways, and if for each of these, a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 \times n_2$ different ways.

Example 1.3 How many possible outcomes are there when a soil's relative density is tested twice and the outcome of each test is either a pass or a fail? Assume that you are interested in the order in which the tests pass or fail.

SOLUTION On the first test, the test can proceed in any one of $n_1 = 2$ ways. For each of these, the second test can proceed in any one of $n_2 = 2$ ways. Therefore, by the multiplication rule, there are $n_1 \times n_2 = 2 \times 2 = 4$ possible test results. Consequently, there are four points in the sample space. These are (P,P), (P,F), (F,P), and (F,F) (see also Example 1.1).

The multiplication principle extends to k operations as follows:

If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed together in

$$n = n_1 \times n_2 \times \cdots \times n_k \tag{1.1}$$

different ways.

Example 1.4 Extending the previous example, suppose that a relative-density test classifies a soil into five possible states, ranging from "very loose" to "very dense." Then if four soil samples are tested, and the outcomes of the four tests are the ordered list of their states, how many possible ways can the tests proceed if the following conditions are assumed to hold?

1. The first sample is either very loose or loose, and all four tests are unique (i.e., all four tests result in different densities).
2. The first sample is either very loose or loose, and tests may yield the same results.
3. The first sample is anything but very loose, and tests may yield the same results.

SOLUTION

1. $2 \times 4 \times 3 \times 2 = 48$
2. $2 \times 5 \times 5 \times 5 = 250$
3. $4 \times 5 \times 5 \times 5 = 500$

Permutations Frequently, we are interested in sample spaces that contain, as elements, all possible orders or arrangements of a group of objects. For example, we may want to know the number of possible ways 6 CPT cones can be selected from a collection of 20 cones of various quality. Here are some examples demonstrating how this can be computed.

Example 1.5 Six piles are being driven to bedrock and the energy required to drive them will be recorded for each. That is, our experiment consists of recording the six measured energy levels. Suppose further that the pile results will be ranked from the one taking the highest energy to the one taking the lowest energy to drive. In how many different ways could this ranked list appear?

SOLUTION The counting process can be broken up into six simpler steps: (1) selecting the pile, out of the six, taking the highest energy to drive and placing it at the top of the list; (2) selecting the pile taking the next highest energy to drive from the remaining five piles and placing it next on the list, and so on for four more steps. Since we know in how many ways each of these operations can be done, we can apply the multiplication rule: $n = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$. Thus, there are 720 ways that the six piles could be ranked according to driving energy.

In the above example, the number of possible arrangements is 6!, where ! is the *factorial* operator. In general,

$$n! = n \times (n-1) \times \cdots \times 2 \times 1 \tag{1.2}$$

if n is a nonzero integer. Also $0! = 1$ by definition. The reasoning of the above example will always prevail when counting the number of possible ways of arranging all objects in a sequence.

Definition A *permutation* is an arrangement, that is, an ordered sequence, of all or part of a set of objects. If we are looking for the number of possible ordered sequences of an entire set, then

The number of permutations of n distinct objects is n!.

If only part of the set of objects is to be ordered, the reasoning is similar to that proposed in Example 1.5, except that now the number of "operations" is reduced. Consider the following example.

Example 1.6 A company has six nuclear density meters, labeled A through F. Because the company wants to keep track of the hours of usage for each, they must each be signed out. A particular job requires three of the meters to be signed out for differing periods of time. In how many ways can three of the meters be selected from the six if the first is to be used the longest, the second for an intermediate amount of time, and the third for the shortest time?

SOLUTION We note that since the three meters to be signed out will be used for differing amounts of time, it will make a difference if A is selected first, rather than second, and so on. That is, the order in which the meters are selected is important. In this case, there are six possibilities for the first meter selected. Once this is selected, the second meter is select from the remaining five meters, and so on. So in total we have $6 \times 5 \times 4 = 120$ ways.

The product $6 \times 5 \times 4$ can be written as

$$\frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1}$$

so that the solution to the above example can be written as

$$6 \times 5 \times 4 = \frac{6!}{(6-3)!}$$

In general, the number of permutations of r objects selected from n distinct objects, where order counts, is

$$P_r^n = \frac{n!}{(n-r)!} \tag{1.3}$$

Combinations In other cases, interest is in the number of ways of selecting r objects from n distinct objects *without regard to order*.

Definition A *combination* is the number of ways that objects can be selected without regard to order.

Question: If there is no regard to order, are there going to be more or less ways of doing things?

Example 1.7 In how many ways can I select two letters from A, B, and C if I do it (a) with regard to order and (b) without regard to order?

SOLUTION
In Figure 1.2, we see that there are *fewer combinations than permutations*. The number of combinations is reduced

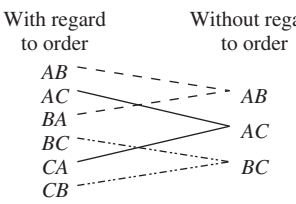


Figure 1.2 Selecting two letters from A, B, and C.

from the number of permutations by a factor of $2 \times 1 = 2$, which is the number of ways the two selected letters can be permuted among themselves.

In general we have:

The number of combinations of n distinct objects taken r at a time is written

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{1.4}$$

Example 1.8 A geotechnical engineering firm keeps a list of eight consultants. Not all consultants are asked to provide a quote on a given request. Determine the number of ways three consultants can be chosen from the list.

SOLUTION

$$\binom{8}{3} = \frac{8!}{3!5!} = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56$$

Sometimes, the multiplication rule, permutations, and/or combinations must be used together to count the number of points in a sample space.

Example 1.9 A company has seven employees specializing in laboratory testing and five employees specializing in field testing. A job requires two employees from each area of specialization. In how many ways can the team of four be formed?

SOLUTION

$$\binom{7}{2} \times \binom{5}{2} = 210$$

1.3 PROBABILITY

1.3.1 Event Probabilities

The probability of an event A, denoted by $\mathrm{P}[A]$, is a number satisfying

$$0 \le \mathrm{P}[A] \le 1$$

Also, we assume that

$$\mathrm{P}[\emptyset] = 0, \qquad \mathrm{P}[S] = 1$$

Probabilities can sometimes be obtained using the counting rules discussed in the previous section. For example, if an experiment can result in any one of N different but equally likely outcomes, and if exactly m of these outcomes correspond to event A, then the probability of event A is $\mathrm{P}[A] = m/N$.

Example 1.10 Sixty soil samples have been taken at a site, 5 of which were taken of a liquefiable soil. If 2 of the samples are selected at random from the 60 samples, what is the probability that neither sample will be of the liquefiable soil?

SOLUTION We could solve this by looking at the number of ways of selecting the 2 samples from the 55 nonliquefiable soil and dividing by the total number of ways of selecting the 2 samples,

$$\mathrm{P}\left[0 \textit{ liquefiable}\right] = \frac{\binom{55}{2}}{\binom{60}{2}} = \frac{99}{118}$$

Alternatively, we could solve this by considering the probability of selecting the "first" sample from the 55 nonliquefiable samples and of selecting the second sample from the remaining 54 nonliquefiable samples,

$$\mathrm{P}\left[0 \textit{ liquefiable}\right] = \frac{55}{60} \times \frac{54}{59} = \frac{99}{118}$$

Note, however, that we have introduced an "ordering" in the second solution that was not asked for in the original question. This ordering needs to be carefully taken account of if we were to ask about the probability of having one of the samples being of a liquefiable soil. See the next example.

Example 1.11 Sixty soil samples have been taken at a site, 5 of which were taken of a liquefiable soil. If 2 of the samples are selected at random from the 60 samples, what is the probability that exactly 1 sample will be of the liquefiable soil?

SOLUTION We could solve this by looking at the number of ways of selecting one sample from the 5 liquefiable samples and 1 sample from the 55 nonliquefiable samples and dividing by the total number of ways of selecting the two samples:

$$\mathrm{P}\left[1 \textit{ liquefiable}\right] = \frac{\binom{5}{1}\binom{55}{1}}{\binom{60}{2}} = 2\left(\frac{5}{60}\right)\left(\frac{55}{59}\right) = \frac{55}{354}$$

We could also solve it by considering the probability of selecting the first sample from the 5 liquefiable samples and the second from the 55 nonliquefiable samples. However, since the question is only looking for the probability of one of the samples being liquefiable, we need to add in the probability that the first sample is nonliquefiable and the second is liquefiable:

$$\begin{aligned}\mathrm{P}\left[1 \textit{ liquefiable}\right] &= \frac{5}{60} \times \frac{55}{59} + \frac{55}{60} \times \frac{5}{59} \\ &= 2\left(\frac{5}{60}\right)\left(\frac{55}{59}\right) = \frac{55}{354}\end{aligned}$$

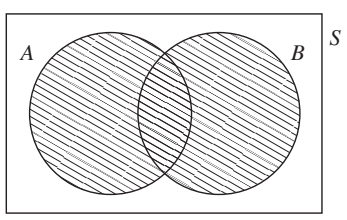


Figure 1.3 Venn diagram illustrating the union $A \cup B$.

1.3.2 Additive Rules

Often we must compute the probability of some event which is expressed in terms of other events. For example, if A is the event that company A requests your services and B is the event that company B requests your services, then the event that at least one of the two companies requests your services is $A \cup B$. The probability of this is given by the following relationship:

If A and B are any two events, then

$$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B] - \mathrm{P}[A \cap B] \tag{1.5}$$

This relationship can be illustrated by the Venn diagram in Figure 1.3. The desired quantity, $\mathrm{P}[A \cup B]$, is the area of $A \cup B$ which is shaded. If the shaded area is computed as the sum of the area of A, $\mathrm{P}[A]$, plus the area of B, $\mathrm{P}[B]$, then the intersection area, $\mathrm{P}[A \cap B]$, has been added twice. It must then be removed once to obtain the correct probability. Also,

If A and B are mutually exclusive, that is, are disjoint and so have no overlap, then

$$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B] \tag{1.6}$$

If $A_1, A_2, \ldots, A_n$ are mutually exclusive, then

$$\mathrm{P}[A_1 \cup \cdots \cup A_n] = \mathrm{P}[A_1] + \cdots + \mathrm{P}[A_n] \tag{1.7}$$

Definition We say that $A_1, A_2, \ldots, A_n$ is a *partition* of the sample space S if $A_1, A_2, \ldots, A_n$ are mutually exclusive and collectively exhaustive. Collectively exhaustive means that $A_1 \cup A_2 \cup \cdots\cdots \cup A_n = S$. If $A_1, A_2, \ldots, A_n$ is a partition of the sample space S, then

$$\mathrm{P}[A_1 \cup \cdots \cup A_n] = \mathrm{P}[A_1] + \cdots + \mathrm{P}[A_n] = \mathrm{P}[S] = 1 \tag{1.8}$$

The above ideas can be extended to the union of more than two events. For example:

For any three events A, B, and C, we have

$$\begin{aligned}\mathrm{P}[A \cup B \cup C] = &\mathrm{P}[A] + \mathrm{P}[B] + \mathrm{P}[C] - \mathrm{P}[A \cap B] \\ &- \mathrm{P}[A \cap C] - \mathrm{P}[B \cap C] \\ &+ \mathrm{P}[A \cap B \cap C]\end{aligned} \tag{1.9}$$

This can be seen by drawing a Venn diagram and keeping track of the areas which must be added and removed in order to get $\mathrm{P}[A \cup B \cup C]$. Example 1.2 illustrates the union of three events.

For the complementary events A and A^c, $\mathrm{P}[A] + \mathrm{P}[A^c] = 1$. This is often used to compute $\mathrm{P}[A^c] = 1 - \mathrm{P}[A]$.

Example 1.12 A data-logging system contains two identical batteries, A and B. If one battery fails, the system will still operate. However, because of the added strain, the remaining battery is now more likely to fail than was originally the case. Suppose that the design life of a battery is three years. If at least one battery fails before the end of the battery design life in 7% of all systems and both batteries fail during that three-year period in only 1% of all systems, what is the probability that battery A will fail during the battery design life?

SOLUTION Let F_A be the event that battery A fails and F_B be the event that battery B fails. Then we are given that

$$\mathrm{P}[F_A \cup F_B] = 0.07, \qquad \mathrm{P}[F_A \cap F_B] = 0.01,$$
$$\mathrm{P}[F_A] = \mathrm{P}[F_B]$$

and we are looking for $\mathrm{P}[F_A]$. The Venn diagram in Figure 1.4 fills in the remaining probabilities. From this diagram, the following result is straightforward: $\mathrm{P}[F_A] = 0.03 + 0.01 = 0.04$.

Example 1.13 Based upon past evidence, it has been determined that in a particular region 15% of CPT soundings encounter soft clay layers, 12% encounter boulders, and 8% encounter both. If a sounding is selected at random:

1. What is the probability that it has encountered both a soft clay layer and a boulder?
2. What is the probability that it has encountered at least one of these two conditions?

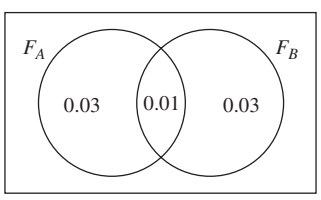


Figure 1.4 Venn diagram of battery failure events.

3. What is the probability that it has encountered neither of these two conditions?
4. What is the probability that it has *not* encountered a boulder?
5. What is the probability that it encounters a boulder but not a soft clay layer?

SOLUTION Let C be the event that the sounding encountered a soft clay layer. Let B be the event that the sounding encountered a boulder. We are given $\mathrm{P}[C] = 0.15$, $\mathrm{P}[B] = 0.12$, and $\mathrm{P}[C \cap B] = 0.08$, from which the Venn diagram in Figure 1.5 can be drawn:

1. $\mathrm{P}[C \cap B] = 0.08$
2. $\mathrm{P}[C \cup B] = \mathrm{P}[C] + \mathrm{P}[B] - \mathrm{P}[C \cap B]$
   $= 0.15 + 0.12 - 0.08$
   $= 0.19$
3. $\mathrm{P}\left[C^c \cap B^c\right] = \mathrm{P}\left[(C \cup B)^c\right]$
   $= 1 - \mathrm{P}[C \cup B]$
   $= 1 - 0.19$
   $= 0.81$
4. $\mathrm{P}[B^c] = 1 - \mathrm{P}[B] = 1 - 0.12 = 0.88$
5. $\mathrm{P}[B \cap C^c] = 0.04$ (see the Venn diagram)

1.4 CONDITIONAL PROBABILITY

The probability of an event is often affected by the occurrence of other events and/or the knowledge of information relevant to the event. Given two events, A and B, of an experiment, $\mathrm{P}[B\,|\,A]$ is called the conditional probability of B given that A has already occurred. It is defined by

$$\mathrm{P}[B\,|\,A] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[A]} \tag{1.10}$$

That is, if we are given that event A has occurred, then A becomes our sample space. The probability that B has also occurred within this new sample space will be the ratio of the "area" of B within A to the "area" of A.

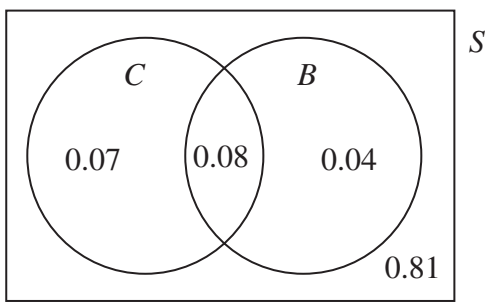


Figure 1.5 Venn diagram of CPT sounding events.

Example 1.14 Reconsidering Example 1.12, what is the probability that battery B will fail during the battery design life given that battery A has already failed?

SOLUTION We are told that F_A has occurred. This means that we are somewhere inside the F_A circle of Figure 1.4, which has "area" 0.04. We are asked to compute the conditional probability that F_B occurs given that F_A has occurred. This will be just the ratio of the area of F_B and F_A to the area of F_A,

$$\mathrm{P}[F_B|F_A] = \frac{\mathrm{P}[F_A \cap F_B]}{\mathrm{P}[F_A]} = \frac{0.01}{0.04} = 0.25$$

Example 1.15 A single soil sample is selected at random from a site. Three different toxic compounds, denoted A, B, and C, are known to occur in samples at this site with the following probabilities:

$$\mathrm{P}[A] = 0.01, \qquad \mathrm{P}[A \cap C] = 0.003,$$
$$\mathrm{P}[A \cap B] = 0.0025, \qquad \mathrm{P}[C] = 0.0075,$$
$$\mathrm{P}[A \cap B \cap C] = 0.001, \qquad \mathrm{P}[B \cap C] = 0.002,$$
$$\mathrm{P}[B] = 0.05$$

If both toxic compounds A and B occur in a soil sample, is the toxic compound C more likely to occur than if neither toxic compounds A nor B occur?

SOLUTION From the given information we can draw the Venn diagram in Figure 1.6.

We want to compare $\mathrm{P}[C|A \cap B]$ and $\mathrm{P}[C|A^c \cap B^c]$, where

$$\mathrm{P}[C\,|\,A \cap B] = \frac{\mathrm{P}[C \cap A \cap B]}{\mathrm{P}[A \cap B]} = \frac{0.001}{0.0025} = 0.4$$

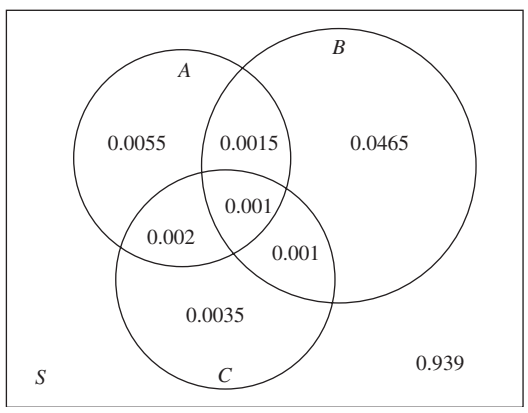


Figure 1.6 Venn diagram of toxic compound occurrence events.

$$\mathrm{P}\left[C \mid A^c \cap B^c\right] = \frac{\mathrm{P}[C \cap A^c \cap B^c]}{\mathrm{P}[A^c \cap B^c]} = \frac{0.0035}{0.939 + 0.0035} = 0.0037$$

so the answer to the question is, yes, if both toxic compounds A and B occur in a soil sample, then toxic compound C is much more likely to also occur.

Sometimes we know $\mathrm{P}[B \mid A]$ and wish to compute $\mathrm{P}[A \cap B]$. If the events A and B can both occur, then

$$\mathrm{P}[A \cap B] = \mathrm{P}[B \mid A]\,\mathrm{P}[A] \tag{1.11}$$

Example 1.16 A site is composed of 60% sand and 40% silt in separate layers and pockets. At this site, 10% of sand samples and 5% of silt samples are contaminated with trace amounts of arsenic. If a soil sample is selected at random, what is the probability that it is a sand sample and that it is contaminated with trace amounts of arsenic?

SOLUTION Let A be the event that the sample is sand. Let B be the event that the sample is silt. Let C be the event that the sample is contaminated with arsenic. Given $\mathrm{P}[A] = 0.6$, $\mathrm{P}[B] = 0.4$, $\mathrm{P}[C \mid A] = 0.1$, and $\mathrm{P}[C \mid B] = 0.05$. We want to find $\mathrm{P}[A \cap C]$:

$$\mathrm{P}[A \cap C] = \mathrm{P}[A]\,\mathrm{P}[C \mid A] = 0.6 \times 0.1 = 0.06$$

Two events A and B are independent *if and only if* $\mathrm{P}[A \cap B] = \mathrm{P}[A]\,\mathrm{P}[B]$. This also implies that $\mathrm{P}[A \mid B] = \mathrm{P}[A]$, that is, if the two events are independent, then they do not affect the probability of the other occurring. Note that *independent events are not disjoint* and *disjoint events are not independent*! In fact, if two events are disjoint, then if one occurs, the other cannot have occurred. Thus, the occurrence of one of two disjoint events has a severe impact on the probability of occurrence of the other event (its probability of occurrence drops to zero).

If, in an experiment, the events $A_1, A_2, \ldots, A_k$ can all occur, then

$$\begin{aligned}\mathrm{P}[A_1 \cap A_2 \cap \cdots \cap A_k] &= \mathrm{P}[A_1]\,\mathrm{P}[A_2 \mid A_1]\,\mathrm{P}[A_3 \mid A_1 \cap A_2] \\ &\quad \cdots \mathrm{P}\left[A_k \mid A_1 \cap \cdots \cap A_{k-1}\right] \\ &= \mathrm{P}[A_k]\,\mathrm{P}\left[A_{k-1} \mid A_k\right] \\ &\quad \cdots \mathrm{P}[A_1 \mid A_k \cap \cdots \cap A_2]\end{aligned} \tag{1.12}$$

On the right-hand side, we could have any ordering of the A's. If the events $A_1, A_2, \ldots, A_k$ are independent, then this simplifies to

$$\mathrm{P}[A_1 \cap A_2 \cap \cdots \cap A_k] = \mathrm{P}[A_1]\,\mathrm{P}[A_2] \cdots \mathrm{P}[A_k] \tag{1.13}$$

Example 1.17 Four retaining walls, A, B, C, and D, are constructed independently. If their probabilities of sliding failure are estimated to be $\mathrm{P}[A] = 0.01$, $\mathrm{P}[B] = 0.008$, $\mathrm{P}[C] = 0.005$, and $\mathrm{P}[D] = 0.015$, what is the probability that none of them will fail by sliding?

SOLUTION Let A be the event that wall A will fail. Let B be the event that wall B will fail. Let C be the event that wall C will fail. Let D be the event that wall D will fail. Given $\mathrm{P}[A] = 0.01$, $\mathrm{P}[B] = 0.008$, $\mathrm{P}[C] = 0.005$, $\mathrm{P}[D] = 0.015$, and that the events A, B, C, and D are independent. We want to find $\mathrm{P}[A^c \cap B^c \cap C^c \cap D^c]$:

$$\begin{aligned}&\mathrm{P}\left[A^c \cap B^c \cap C^c \cap D^c\right] \\ &\quad = \mathrm{P}\left[A^c\right]\mathrm{P}\left[B^c\right]\mathrm{P}\left[C^c\right]\mathrm{P}\left[D^c\right] \\ &\qquad \text{(since } A, B, C, \text{ and } D \text{ are independent)} \\ &\quad = (1 - \mathrm{P}[A])(1 - \mathrm{P}[B])(1 - \mathrm{P}[C])(1 - \mathrm{P}[D]) \\ &\quad = (1 - 0.01)(1 - 0.008)(1 - 0.005)(1 - 0.015) \\ &\quad = 0.9625\end{aligned}$$

1.4.1 Total Probability

Sometimes we know the probability of an event in terms of the occurrence of other events and want to compute the *unconditional* probability of the event. For example, when we want to compute the *total* probability of failure of a bridge, we can start by computing a series of simpler problems such as:

1. Probability of bridge failure given a maximum static load
2. Probability of bridge failure given a maximum dynamic traffic load
3. Probability of bridge failure given an earthquake
4. Probability of bridge failure given a flood

The *total probability theorem* can be used to combine the above probabilities into the unconditional probability of bridge failure. We need to know the above conditional probabilities along with the probabilities that the "conditions" occur (e.g., the probability that the maximum static load will occur during the design life).

Example 1.18 A company manufactures cone penetration testing equipment. Of the piezocones they use, 50% are

produced at plant A, 30% at plant B, and 20% at plant C. It is known that 1% of plant A's, 2% of plant B's, and 3% of plant C's output are defective. What is the probability that a piezocone chosen at random will be defective?

Setup

Let A be the event that the piezocone was produced at plant A. Let B be the event that the piezocone was produced at plant B. Let C be the event that the piezocone was produced at plant C. Let D be the event that the piezocone is defective. Given

$$\begin{aligned} \mathrm{P}[A] &= 0.50, \qquad \mathrm{P}[D \mid A] = 0.01, \\ \mathrm{P}[B] &= 0.30, \qquad \mathrm{P}[D \mid B] = 0.02, \\ \mathrm{P}[C] &= 0.20, \qquad \mathrm{P}[D \mid C] = 0.03 \end{aligned}$$

We want to find $\mathrm{P}[D]$. There are at least two possible approaches.

Approach 1

A Venn diagram of the sample space is given in Figure 1.7. The information given in the problem does not allow the Venn diagram to be easily filled in. It is easy to see the event of interest, though, as it has been shaded in. Then

$$\begin{aligned} \mathrm{P}[D] &= \mathrm{P}[(D \cap A) \cup (D \cap B) \cup (D \cap C)] \\ &= \mathrm{P}[D \cap A] + \mathrm{P}[D \cap B] + \mathrm{P}[D \cap C] \\ &\qquad \text{since } A \cap D,\ B \cap D, \text{ and } C \cap D \text{ are disjoint} \\ &= \mathrm{P}[D \mid A] \cdot \mathrm{P}[A] + \mathrm{P}[D \mid B] \cdot \mathrm{P}[B] \\ &\quad + \mathrm{P}[D \mid C] \cdot \mathrm{P}[C] \\ &= 0.01(0.5) + 0.02(0.3) + 0.03(0.2) \\ &= 0.017 \end{aligned}$$

Approach 2

Recall that when we only had probabilities like $\mathrm{P}[A]$, $\mathrm{P}[B]$, ..., that is, no conditional probabilities, we found it helpful to represent the probabilities in a Venn diagram. Unfortunately, there is no easy representation of the conditional probabilities in a Venn diagram: (In fact, conditional probabilities are ratios of probabilities that appear in the Venn diagram.) Conditional probabilities find a more natural home on *event trees*. Event trees must be constructed carefully and adhere to certain rules if they are going to be useful in calculations. Event trees consist of nodes and branches. There is a starting node from which two or more branches leave. At the end of each of these branches there is another node from which more branches may leave (and go to separate nodes). The idea is repeated from the newer nodes as often as required to completely depict all possibilities. A probability is associated with each branch and, for all branches except those leaving the starting node, the probabilities are conditional probabilities. Thus, the event tree is composed largely of conditional probabilities.

There is one other rule that event trees must obey: *Branches leaving any node must form a partition of the sample space.* That is, the events associated with each branch must be disjoint—you cannot be on more than one branch at a time—and must include all possibilities. The sum of probabilities of all branches leaving a node must be 1.0. Also keep in mind that an event tree will only be useful if all the branches can be filled with probabilities.

The event tree for this example is constructed as follows. The piezocone must *first be made at one of the three plants*, then *depending on where it was made*, it could be *defective* or not. The event tree for this problem is thus as given in Figure 1.8. Note that there are six "paths" on the tree. When a piezocone is selected at random, exactly one of these paths will have been followed—we will be on one of the branches. Recall that interest is in finding $\mathrm{P}[D]$. The event D will have occurred if either the first, third, or fifth path was followed. That is, the probability that the first, third, or fifth path was followed is sought. If the first path is followed, then the event $A \cap D$ has occurred. This has probability found by multiplying the probabilities along the path,

$$\mathrm{P}[A \cap D] = \mathrm{P}[D \mid A] \cdot \mathrm{P}[A] = 0.01(0.5) = 0.005$$

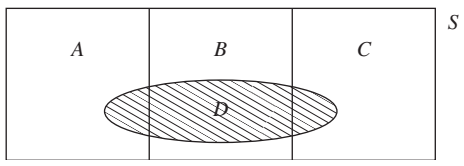


Figure 1.7 Venn diagram of piezocone events.

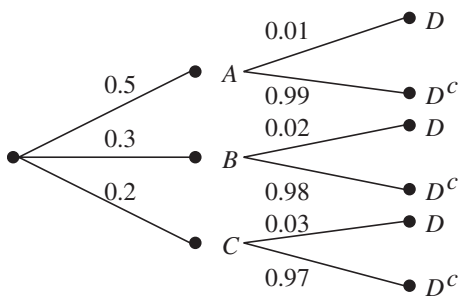


Figure 1.8 Event tree for piezocone events.

Looking back at the calculation performed in Approach 1, $\mathrm{P}[D]$ was computed as

$$\begin{aligned}\mathrm{P}[D] &= \mathrm{P}[D \mid A] \cdot \mathrm{P}[A] + \mathrm{P}[D \mid B] \cdot \mathrm{P}[B] \\ &\quad + \mathrm{P}[D \mid C] \cdot \mathrm{P}[C] \\ &= 0.01(0.5) + 0.02(0.3) + 0.03(0.2) \\ &= 0.017\end{aligned}$$

which, in terms of the event tree, is just the sum of all the paths that lead to the outcome that you desire, D. Event trees make "total probability" problems much simpler. They give a "picture" of what is going on and allow the computation of some of the desired probabilities directly.

The above is an application of the total probability theorem, which is stated generally as follows:

Total Probability Theorem If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S (i.e., are disjoint and collectively exhaustive), then for any event A in S

$$\mathrm{P}[A] = \sum_{i=1}^{k} \mathrm{P}[B_i \cap A] = \sum_{i=1}^{k} \mathrm{P}[A \mid B_i]\,\mathrm{P}[B_i] \qquad (1.14)$$

1.4.2 Bayes' Theorem

Sometimes we want to improve an estimate of a probability in light of additional information. Bayes' theorem allows us to do this. It arises from the observation that $\mathrm{P}[A \cap B]$ can be written in two ways:

$$\begin{aligned}\mathrm{P}[A \cap B] &= \mathrm{P}[A \mid B] \cdot \mathrm{P}[B] \\ &= \mathrm{P}[B \mid A] \cdot \mathrm{P}[A] \qquad (1.15)\end{aligned}$$

which implies that $\mathrm{P}[B \mid A] \cdot \mathrm{P}[A] = \mathrm{P}[A \mid B] \cdot \mathrm{P}[B]$, or

$$\mathrm{P}[B \mid A] = \frac{\mathrm{P}[A \mid B] \cdot \mathrm{P}[B]}{\mathrm{P}[A]} \qquad (1.16)$$

Example 1.19 Return to the manufacturer of piezocones from above (Example 1.18). If a piezocone is selected at random and found to be defective, what is the probability that it came from plant A?

Setup
Same as before, except now the probability of interest is $\mathrm{P}[A \mid D]$. Again, there are *two possible approaches*.

Approach 1
The relationship

$$\mathrm{P}[A \mid D] = \frac{\mathrm{P}[A \cap D]}{\mathrm{P}[D]}$$

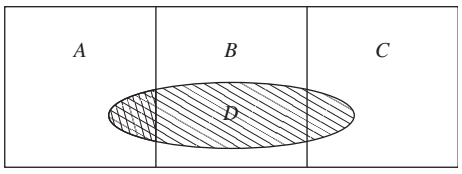


Figure 1.9 Venn diagram of conditional piezocone events.

can be seen as a ratio of areas in the Venn diagram in Figure 1.9, from which $\mathrm{P}[A \mid D]$ can be computed as follows:

$$\begin{aligned}&\mathrm{P}[A \mid D] \\ &= \frac{\mathrm{P}[A \cap D]}{\mathrm{P}[D]} \\ &= \frac{\mathrm{P}[A \cap D]}{\mathrm{P}[(A \cap D) \cup (B \cap D) \cup (C \cap D)]} \\ &= \frac{\mathrm{P}[A \cap D]}{\mathrm{P}[A \cap D] + \mathrm{P}[B \cap D] + \mathrm{P}[C \cap D]} \\ &\quad \text{since } A \cap D,\ B \cap D, \text{ and } C \cap D \text{ are disjoint} \\ &= \frac{\mathrm{P}[D \mid A]\,\mathrm{P}[A]}{\mathrm{P}[D \mid A]\,\mathrm{P}[A] + \mathrm{P}[D \mid B]\,\mathrm{P}[B] + \mathrm{P}[D \mid C]\,\mathrm{P}[C]} \\ &= \frac{0.01(0.5)}{(0.01)(0.5) + 0.02(0.3) + 0.03(0.2)} = \frac{0.005}{0.017} \\ &= 0.294\end{aligned}$$

Note that the denominator had already been calculated in the previous question; however the computations have been reproduced here for illustrative purposes.

Approach 2
The probability $\mathrm{P}[A \mid D]$ can be easily computed from the event tree. We are looking for the probability that A has occurred given that D has occurred. In terms of the paths on the tree, we know that (since D has occurred) one of the first, third, or fifth path has been taken. We want the probability that the first path was taken out of the three possible paths. Thus, we must compute the relative probability of taking path 1 out of the three paths:

$$\begin{aligned}&\mathrm{P}[A \mid D] \\ &= \frac{\mathrm{P}[D \mid A]\,\mathrm{P}[A]}{\mathrm{P}[D \mid A]\,\mathrm{P}[A] + \mathrm{P}[D \mid B]\,\mathrm{P}[B] + \mathrm{P}[D \mid C]\,\mathrm{P}[C]} \\ &= \frac{0.01(0.5)}{(0.01)(0.5) + 0.02(0.3) + 0.03(0.2)} = \frac{0.005}{0.017} \\ &= 0.294\end{aligned}$$

Event trees provide a simple graphical approach to solving problems involving conditional probabilities.

The above is an application of Bayes' Theorem, which is stated formally as follows.

Bayes' Theorem If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S (i.e., are disjoint and collectively exhaustive), then for any event A of S such that $\mathrm{P}[A] \neq 0$

$$\mathrm{P}\left[B_j \mid A\right] = \frac{\mathrm{P}\left[B_j \cap A\right]}{\sum_{i=1}^{k} \mathrm{P}\left[B_i \cap A\right]} = \frac{\mathrm{P}\left[A \mid B_j\right]\mathrm{P}\left[B_j\right]}{\sum_{i=1}^{k} \mathrm{P}\left[A \mid B_i\right]\mathrm{P}\left[B_i\right]} = \frac{\mathrm{P}\left[A \mid B_j\right]\mathrm{P}\left[B_j\right]}{\mathrm{P}[A]} \tag{1.17}$$

for any $j = 1, 2, \ldots, k$.

Bayes' theorem is useful for revising or updating probabilities as more data and information become available. In the previous example on piezocones, there was an *initial* probability that a piezocone would have been manufactured at plant A: $\mathrm{P}[A] = 0.5$. This probability is referred to as the *prior* probability of A. That is, in the absence of any other information, a piezocone chosen at random has a probability of having been manufactured at plant A of 0.5. However, if a piezocone chosen at random is found to be defective (so that there is now more information on the piezocone), then the probability that it was manufactured at plant A reduces from 0.5 to 0.294. This latter probability is referred to as the *posterior* probability of A. *Bayesian* updating of probabilities is a very powerful tool in engineering reliability-based design.

For problems involving conditional probabilities, event trees are usually the easiest way to proceed. However, event trees are not always easy to draw, and the purely mathematical approach is sometimes necessary. As an example of a tree which is not quite straightforward, see if you can draw the event tree and answer the questions in the following exercise. Remember that you must set up the tree in such a way that you can fill in most of the probabilities on the branches. If you are left with too many empty branches and no other given information, you are likely to have confused the order of the events; try reorganizing your tree.

Exercise When contracting out a site investigation, an engineer will check companies A, B, and C in that sequence and will hire the first company which is available to do the work. From past experience, the engineer knows that the probability that company A will be available is 0.2. However, if company A is not available, then the probability that company B will be available is only 0.04. If neither company A nor B is available, then the probability that company C will be available is 0.4. If none of the companies are available, the engineer is forced to delay the investigation to a later time.

(a) What is the probability that one of the companies A or B will be available?
(b) What is the probability that the site investigation will take place on time?
(c) If the site investigation takes place on time, what is the probability that it was not investigated by company C?

Example 1.20 At a particular site, experience has shown that piles have a 20% probability of encountering a soft clay layer. Of those which encounter this clay layer, 60% fail a static load test. Of the piles which do not encounter the clay layer, only 10% fail a static load test.

1. What is the probability that a pile selected at random will fail a static load test?
2. Supposing that a pile has failed a static load test, what is the updated probability that it encountered the soft clay layer?

SOLUTION For a pile, let C be the event that a soft clay layer was encountered and let F be the event that the static load test was failed. We are given $\mathrm{P}[C] = 0.2$, $\mathrm{P}[F \mid C] = 0.6$, and $\mathrm{P}[F \mid C^c] = 0.1$.

1. We have the event tree in Figure 1.10 and thus $\mathrm{P}[F] = 0.2(0.6) + 0.8(0.1) = 0.2$.
2. From the above tree, we have

$$\mathrm{P}[C \mid F] = \frac{0.2 \times 0.6}{0.2} = 0.6$$

1.4.3 Problem-Solving Methodology

Solving real-life problems (i.e., "word problems") is not always easy. It is often not perfectly clear what is meant by a worded question. Two things improve one's chances of successfully solving problems which are expressed using words: (a) a systematic approach, and (b) *practice*. It is practice that allows you to identify those aspects of the question that need further clarification, if any. Below, a few basic recommendations are outlined.

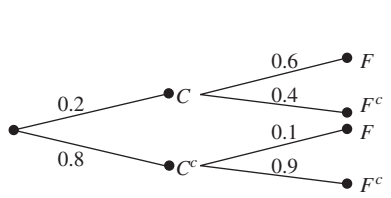


Figure 1.10 Event tree for pile encounter events.

1. Solving a word problem generally involves the computation of some quantity. Clearly identify this quantity at the beginning of the problem solution. Before starting any computations, it is good practice to write out your concluding sentence first. This forces you to concentrate on the essentials.
2. In any problem involving the probability of events, you should:

(a) *Clearly define your events.* Use the following guidelines:
   (i) Keep events as simple as possible.
   (ii) if your event definition includes the words *and*, *or*, *given*, *if*, *when*, and so on, then *it is NOT a good event definition*. Break your event into two (or more, if required) events and use the $\cap$, $\cup$, or | operators to express what you had originally intended. The complement is also a helpful operator, see (iii).
   (iii) You do not need to define separate events for, for example, "an accident occurs" and "an accident does not occur". In fact, this will often lead to confusion. Simply define A to be one of the events and use A^c when you want to refer to the other. This may also give you some hints as to how to proceed since you know that $\mathrm{P}[A^c] = 1 - \mathrm{P}[A]$.

(b) Once your events are defined, you need to go through the worded problem to extract the given numerical information. Write this information down in the form of probabilities of the events that you defined above. For example, $\mathrm{P}[A] = 0.23$, $\mathrm{P}[B \mid A] = 0.6$, and so on. Note that the conditional probabilities, are often difficult to unravel. For example, the following phrases all translate into a probability statement of the form $\mathrm{P}[A \mid B]$:

If . . . occurs, the probability of . . . doubles. . . .
In the event that . . . occurs, the probability of . . . becomes 0.6.
When . . . occurs, the probability of . . . becomes 0.43.
Given that . . . occurs, the probability of . . . is 0.3.

In this case, you will likely be using one of the conditional probability relationship ($\mathrm{P}[A \cap B] = \mathrm{P}[B \mid A]\,\mathrm{P}[A]$), the total probability theorem, or Bayes' Theorem.

(c) Now review the worded problem again and write down the probability that the question is asking for in terms of the events defined above. Although the question may be in worded form, you should be writing down something like $\mathrm{P}[A \cap B]$ or $\mathrm{P}[B \mid A]$. Make sure that you can express the desired probability in terms of the events you defined above. If you cannot, then you need to revise your original event definitions.

(d) Finally, use the rules of combining probabilities (e.g., probabilities of unions or intersections, Bayes' Theorem) to compute the desired probability.

1.5 RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Although probability theory is based on the idea of events and associated set theory, it becomes very unwieldy to treat random events like "time to failure" using explicit event definitions. One would conceivably have to define a separate event for each possible time of failure and so would soon run out of symbols for the various events. For this reason, and also because they allow the use of a wealth of mathematical tools, random variables are used to represent a suite of possible events. In addition, since most engineering problems are expressed in terms of numerical quantities, random variables are particularly appropriate.

Definition Consider a sample space S consisting of a set of outcomes $\{s_1, s_2, \ldots\}$. If X is a function that assigns a real number $X(s)$ to every outcome $s \in S$, then X is a *random variable*. Random variables will be denoted with uppercase letters.

Now what does this mean in plain English? Essentially a random variable is a means of identifying events in numerical terms. For example, if the outcome s_1 means that an apple was selected and s_2 means that an orange was selected, then $X(s_1)$ could be set equal to 1 and $X(s_2)$ could be set equal to 0. Then $X > 0$ means that an apple was selected. Now mathematics can be used on X, that is, if the fruit-picking experiment is repeated n times and $x_1 = X_1(s)$ is the outcome of the first experiment, $x_2 = X_2(s)$ the outcome of the second, and so on, then the total number of apples picked is $\sum_{i=1}^{n} x_i$. Note that mathematics could not be used on the actual outcomes themselves; for example, picking an apple is a real event which knows nothing about mathematics nor can it be used in a mathematical expression without first mapping the event to a number.

For each outcome s, there is exactly one value of $x = X(s)$, but different values of s may lead to the same x. We will see examples of this shortly.

The above discussion illustrates in a rather simple way one of the primary motivations for the use of random variables—simply so that mathematics can be used. One other thing might be noticed in the previous paragraph. After the "experiment" has taken place and the outcome is known, it is referred to using lowercase, x_i. That is x_i has a known fixed value while X is unknown. In other words

x is a realization of the random variable X. This is a rather subtle distinction, but it is important to remember that X is unknown. The most that we can say about X is to specify what its likelihoods of taking on certain values are—we cannot say exactly what the value of X is.

Example 1.21 Two piles are to be randomly selected for testing from a group of 60 piles. Five of the piles are 0.5 m in diameter, the rest are 0.3 m in diameter. If X is the number of 0.5-m-diameter piles selected for testing, then X is a random variable that assigns a number to each outcome in the sample space according to:

Sample Space	X
NN	0
NL	1
LN	1
LL	2

The sample space is made up of pairs of possible outcomes, where N represents a "normal" diameter pile (0.3 m) and L represents a large -diameter pile (0.5 m). For example, *LN* means that the first pile selected was large and the second pile selected was normal. Notice that the outcomes $\{NL\}$ and $\{LN\}$ both lead to $X = 1$.

Sample spaces corresponding to random variables may be discrete or continuous:

Discrete: A random variable is called a discrete random variable if its set of possible outcomes is countable. This usually occurs for any random variable which is a count of occurrences or of items, for example, the number of large-diameter piles selected in the previous example.

Continuous: A random variable is called a continuous random variable if it can take on values on a continuous scale. This is usually the case with measured data, such as cohesion.

Example 1.22 A few examples:

1. Let X be the number of blows in a standard penetration test—X is discrete.
2. Let Y be the number of piles driven in one day—Y is discrete.
3. Let Z be the time until consolidation settlement exceeds some threshold—Z is continuous.
4. Let W be the number of grains of sand involved in a sand cone test—W is discrete but is often approximated as continuous, particularly since W can be very large.

1.5.1 Discrete Random Variables

Discrete random variables are those that take on only discrete values $\{x_1, x_2, \ldots\}$, that is, have a countable number of outcomes. Note that countable just means that the outcomes can be numbered $1, 2, \ldots$, however there could still be an infinite number of them. For example, our experiment might be to count the number of soil tests performed before one yields a cohesion of 200 MPa. This is a discrete random variable since the outcome is one of $0, 1, \ldots$, but the number may be very large or even (in concept) infinite (implying that a soil sample with cohesion 200 MPa was never found).

Discrete Probability Distributions As mentioned previously, we can never know for certain what the value of a random variable is (if we do measure it, it becomes a *realization*—presumably the next measurement is again uncertain until it is measured, and so on). The most that we can say about a random variable is what its probability is of assuming each of its possible values. The set of probabilities assigned to each possible value of X is called a *probability distribution*. The sum of these probabilities over all possible values must be 1.0.

Definition The set of ordered pairs $(x, f_X(x))$ is the probability distribution of the discrete random variable X if, for each possible outcome x,

1. $0 \le f_X(x) \le 1$
2. $\sum_{\text{all } x} f_X(x) = 1$
3. $\mathrm{P}[X = x] = f_X(x)$

Here, $f_X(x)$ is called the *probability mass function* of X. The subscript is used to indicate what random variable is being governed by the distribution. We shall see when we consider continuous random variables why we call this a probability "mass" function.

Example 1.23 Recall Example 1.21. We can compute the probability mass function of the number of large piles selected by using the counting rules of Section 1.2.3. Specifically,

$$f_X(0) = \mathrm{P}[X = 0] = \frac{\binom{5}{0}\binom{55}{2}}{\binom{60}{2}} = 0.8390$$

$$f_X(1) = \mathrm{P}[X = 1] = \frac{\binom{5}{1}\binom{55}{1}}{\binom{60}{2}} = 0.1554$$

$$f_X(2) = \mathrm{P}[X = 2] = \frac{\binom{5}{2}\binom{55}{0}}{\binom{60}{2}} = 0.0056$$

and thus the probability mass function of the random variable X is

x	$f_X(x)$
0	0.8390
1	0.1554
2	0.0056

Discrete Cumulative Distributions An *equivalent* description of a random variable is the cumulative distribution function (cdf), which is defined as follows:

Definition The cumulative distribution function $F_X(x)$ of a discrete random variable X with probability mass function $f_X(x)$ is defined by

$$F_X(x) = \mathrm{P}[X \leq x] = \sum_{t \leq x} f_X(t) \tag{1.18}$$

We say that this is equivalent to the probability mass function because one can be obtained from the other,

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \tag{1.19}$$

Example 1.24 In the case of an experiment involving tossing a fair coin three times we can count the number of heads which appear and assign that to the random variable X. The random variable X can assume four values 0, 1, 2, and 3 with probabilities $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, and $\frac{1}{8}$ (do you know how these probabilities were computed?). Thus, $F_X(x)$ is defined as

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{8} & \text{if } 0 \leq x < 1 \\ \frac{4}{8} & \text{if } 1 \leq x < 2 \\ \frac{7}{8} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } 3 \leq x \end{cases}$$

and a graph of $F_X(x)$ appears in Figure 1.11. The values of $F_X(x)$ at $x = 0, 1, \ldots$ are shown by the closed circles.

Discrete probability mass functions are often represented using a bar plot, where the height of each bar is equal to the probability that the random variable takes that value. For example, the bar plot of the pile problem (Examples 1.21 and 1.23) would appear as in Figure 1.12.

1.5.2 Continuous Random Variables

Continuous random variables can take on an infinite number of possible outcomes—generally X takes values from the real line $\Re$. To illustrate the changes involved when we

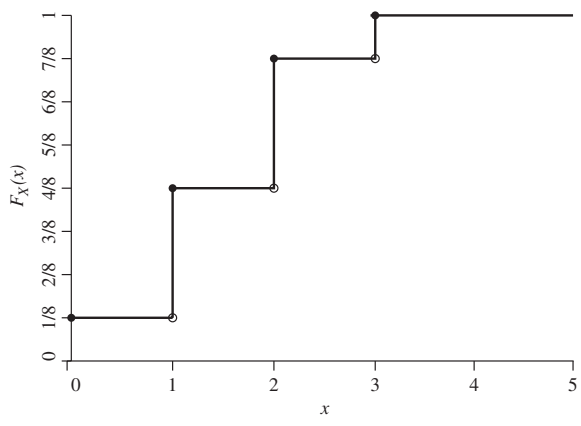


Figure 1.11 Cumulative distribution function for the three-coin toss.

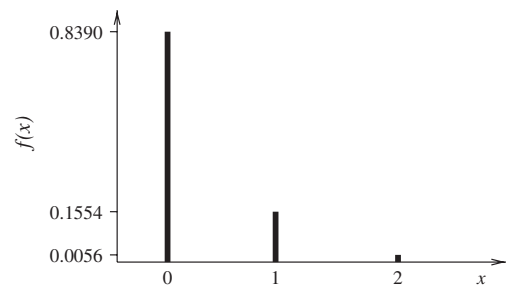


Figure 1.12 Bar plot of $f_X(x)$ for number of large piles selected, X.

go from the discrete to the continuous case, consider the probability that a grain silo experiences a bearing capacity failure at *exactly* 4.3673458212... years from when it is installed. Clearly the probability that it fails at exactly that instant in time is essentially zero. In general the probability that it fails at any one instant in time is vanishingly small. In order to characterize probabilities for continuous random variables, we cannot use probabilities directly (since they are all essentially zero); we must use *relative likelihoods*. That is, we say that the probability that X lies in the small interval between x and $x + dx$ is $f_X(x)\,dx$, or

$$\mathrm{P}[x < X \leq x + dx] = f_X(x)\,dx \tag{1.20}$$

where $f_X(x)$ is now called the *probability density function* (pdf) of the random variable X. The word *density* is used because "density" must be multiplied by a length measure in order to get a "mass." Note that the above probability is vanishingly small because dx is vanishingly small. The function $f_X(x)$ is now the relative likelihood that X lies in a very small interval near x. Roughly speaking, we can think of this as $\mathrm{P}[X = x] = f_X(x)\,dx$.

Continuous Probability Distributions

Definition The function $f_X(x)$ is a probability density function for the continuous random variable X defined over the set of real numbers if

1. $0 \leq f_X(x) < \infty$ for all $-\infty < x < +\infty$,
2. $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$ (i.e., the area under the pdf is 1.0), and
3. $\mathrm{P}[a < X < b] = \int_a^b f_X(x)\, dx$ (i.e., the area under $f_X(x)$ between a and b).

Note: it is important to recognize that, in the continuous case, $f_X(x)$ is no longer a probability. It has units of probability per unit length. In order to get probabilities, we have to find *areas* under the pdf, that is, sum values of $f_X(x)\, dx$.

Example 1.25 Suppose that the time to failure, T in years, of a clay barrier has the probability density function

$$f_T(t) = \begin{cases} 0.02e^{-0.02t} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

This is called an *exponential distribution* and distributions of this exponentially decaying form have been found to well represent many lifetime-type problems. What is the probability that T will exceed 100 years?

SOLUTION The distribution is shown in Figure 1.13. If we consider the more general case where

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

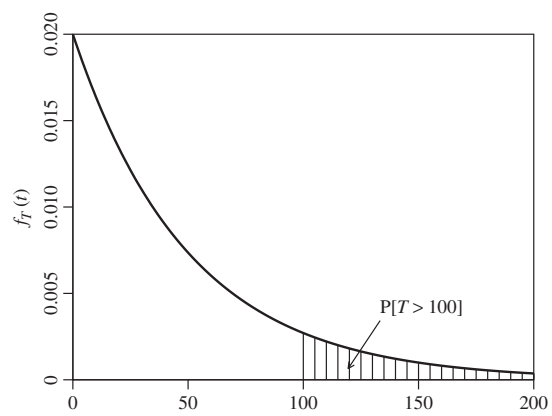


Figure 1.13 Exponential distribution illustrating $\mathrm{P}[T > 100]$.

then we get

$$\mathrm{P}[T > 100] = \mathrm{P}[100 < T < \infty] = \int_{100}^{\infty} \lambda e^{-\lambda t}\, dt$$
$$= -e^{-\lambda t}\Big|_{100}^{\infty} = -e^{-\infty\lambda} + e^{-100\lambda}$$
$$= e^{-100\lambda}$$

For $\lambda = 0.02$, as is the case in this problem,

$$\mathrm{P}[T > 100] = e^{-100\times 0.02} = e^{-2} = 0.1353$$

Continuous Cumulative Distribution The *cumulative distribution function* (cdf) for a continuous random variable is basically defined in the same way as it is for a discrete distribution (Figure 1.14).

Definition The cumulative distribution function $F_X(x)$ of a continuous random variable X having probability density function $f_X(x)$ is defined by the area under the density function to the left of x:

$$F_X(x) = \mathrm{P}[X \leq x] = \int_{-\infty}^{x} f_X(t)\, dt \qquad (1.21)$$

As in the discrete case, the cdf is equivalent to the pdf in that one can be obtained from the other. It is simply another way of expressing the probabilities associated with a random variable. Since the cdf is an integral of the pdf, the pdf can be obtained from the cdf as a derivative:

$$f_X(x) = \frac{dF_X(x)}{dx} \qquad (1.22)$$

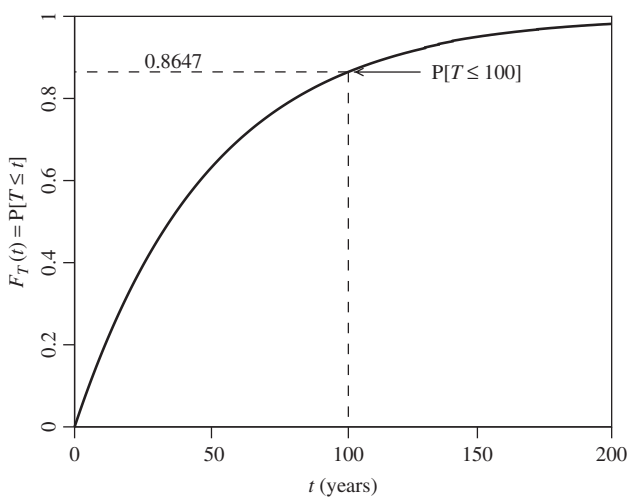


Figure 1.14 Cumulative distribution function for the exponential distribution.

Example 1.26 Note that we could also have used the cumulative distribution in Example 1.25. The cumulative distribution function of the exponential distribution is

$$F_T(t) = \mathrm{P}[T \le t] = \int_0^t \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda t}$$

and thus

$$\begin{aligned}\mathrm{P}[T > 100] &= 1 - \mathrm{P}[T \le 100] = 1 - F_T(100) \\ &= 1 - (1 - e^{-100\lambda}) = e^{-100\lambda}\end{aligned}$$

1.6 MEASURES OF CENTRAL TENDENCY, VARIABILITY, AND ASSOCIATION

A random variable is completely described, as well as can be, if its probability distribution is specified. However, we will never know the precise distribution of any natural phenomenon. Nature cares not at all about our mathematical models and the "truth" is usually far more complex than we are able to represent. So we very often have to describe a random variable using less complete but more easily estimated measures. The most important of these measures are *central tendency* and *variability*. Even if the complete probability distribution is known, these quantities remain useful because they convey information about the properties of the random variable that are of first importance in practical applications. Also, the parameters of the distribution are often derived as functions of these quantities or they may be the parameters themselves.

The most common measures of central tendency and variability are the *mean* and the *variance*, respectively. In engineering, the variability of a random quantity is often expressed using the dimensionless *coefficient of variation*, which is the ratio of the *standard deviation* over the mean. Also, when one has two random variables X and Y, it is frequently of interest to measure how strongly they are related (or associated) to one another. A typical measure of the strength of the relationship between two random variables is their *covariance*. As we shall see, covariance depends on the units of the random variables involved and their individual variabilities, and so a more intuitive measure of the strength of the relationship between two random variables is the *correlation coefficient*, which is both dimensionless and bounded. All of these characteristics will be covered in this section.

1.6.1 Mean

The *mean* is the most important characteristic of a random variable, in that it tells us about its central tendency. It is defined mathematically as follows:

Definition Let X be a random variable with probability density function $f(x)$. The mean, or *expected value*, of X, denoted μ_X, is defined by

$$\mu_X = \begin{cases} \mathrm{E}[X] = \displaystyle\sum_x xf(x) \\ \qquad \text{if } X \text{ is discrete} & (1.23a) \\ \mathrm{E}[X] = \displaystyle\int_{-\infty}^{\infty} xf(x)\, dx \\ \qquad \text{if } X \text{ is continuous} & (1.23b) \end{cases}$$

where the subscript on μ, when present, denotes what μ is the mean of.

Example 1.27 Let X be a discrete random variable which takes on the values listed in the table below with associated probabilities:

x	-2	-1	0	1	2
$f(x)$	$\frac{1}{12}$	$\frac{1}{6}$	k	$\frac{1}{3}$	$\frac{1}{4}$

1. Find the constant k such that $f_X(x)$ is a legitimate probability mass function for the random variable X.
2. Find the mean (expected value) of X.

SOLUTION

1. We know that the sum of all possible probabilities must be 1, so that $k = 1 - (\frac{1}{12} + \frac{1}{6} + \frac{1}{3} + \frac{1}{4}) = \frac{1}{6}$.
2. $\mathrm{E}[X] = (-2)(\frac{1}{12}) + (-1)(\frac{1}{6}) + 0(\frac{1}{6}) + 1(\frac{1}{3}) + 2(\frac{1}{4}) = \frac{1}{2}$.

Expectation The notation $\mathrm{E}[X]$ refers to a mathematical operation called *expectation*. The expectation of any random variable is a sum of all possible values of the random variable weighted by the probability of each value occurring. For example, if X is a random variable with probability (mass or density) function $f_X(x)$, then the expected value of the random variable $g(X)$, where g is any function of X, is

$$\mu_{g(X)} = \begin{cases} \mathrm{E}\big[g(X)\big] = \displaystyle\sum_x g(x) f_X(x) \\ \qquad \text{if } X \text{ is discrete} \\ \mathrm{E}\big[g(X)\big] = \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x)\, dx \\ \qquad \text{if } X \text{ is continuous} \end{cases} \quad (1.24)$$

Example 1.28 A researcher is looking at fibers as a means of reinforcing soil. The fibers being investigated are nominally of radius 10 μm. However, they actually have

random radius R with probability density function (in units of micrometers)

$$f_R(r) = \begin{cases} \frac{3}{4}\left[1-(10-r)^2\right] & \text{if } 9 \le r \le 11 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected area of a reinforcing fiber?

SOLUTION The area of a circle of radius R is πR^2. Thus,

$$\begin{aligned} \mathrm{E}\left[\pi R^2\right] &= \pi \mathrm{E}\left[R^2\right] = \pi \int_9^{11} r^2 \frac{3}{4}\left[1-(10-r)^2\right] dr \\ &= \frac{3}{4}\pi \int_9^{11} \left[-99r^2 + 20r^3 - r^4\right] dr \\ &= \frac{3}{4}\pi \left[-33r^3 + 5r^4 - \frac{r^5}{5}\right]_9^{11} \\ &= \frac{3}{4}\pi \left(\frac{668}{5}\right) = \frac{501}{5}\pi \\ &= 314.8 \quad \mu\mathrm{m}^2 \end{aligned}$$

If we have a sample of observations $x_1, x_2, \ldots, x_n$ of some population X, then the population mean μ_X is estimated by the *sample mean* $\bar{x}$, defined as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.25}$$

Example 1.29 Suppose $\mathbf{x} = \{x_1, x_2, \ldots, x_n\} = \{1, 3, 5, 7, 9\}$.

(a) What is $\bar{x}$?
(b) What happens to $\bar{x}$ if $\mathbf{x} = \{1, 3, 5, 7, 79\}$?

SOLUTION In both cases, the sample size is $n = 5$.

(a) $\bar{x} = \frac{1}{5}(1+3+5+7+9) = 5$

(b) $\bar{x} = \frac{1}{5}(1+3+5+7+79) = 19$

Notice that the one (possible erroneous) observation of 79 makes a big difference to the sample mean. An alternative measure of central tendency, which enthusiasts of robust statistics vastly prefer, is the median, discussed next.

1.6.2 Median

The *median* is another measure of central tendency. We shall denote the median as $\tilde{\mu}$. It is the point which divides the distribution into two equal halves. Most commonly, $\tilde{\mu}$ is found by solving

$$F_X(\tilde{\mu}) = \mathrm{P}\left[X \le \tilde{\mu}\right] = 0.5$$

for $\tilde{\mu}$. For example, if $f_X(x) = \lambda e^{-\lambda x}$, then $F_X(x) = 1 - e^{-\lambda x}$, and we get

$$1 - e^{-\lambda\tilde{\mu}} = 0.5 \qquad \Longrightarrow \qquad \tilde{\mu}_X = -\frac{\ln(0.5)}{\lambda} = \frac{0.693}{\lambda}$$

While the mean is strongly affected by extremes in the distribution, the median is largely unaffected.

In general, the mean and the median are not the same. If the distribution is positively skewed (or skewed right, which means a longer tail to the right than to the left), as are most soil properties, then the mean will be to the right of the median. Conversely, if the distribution is skewed left, then the mean will be to the left of the median. If the distribution is symmetric, then the mean and the median will coincide.

If we have a sample of observations $x_1, x_2, \ldots, x_n$ of some population X, then the population median $\tilde{\mu}_X$ is estimated by the *sample median* $\tilde{x}$. To define $\tilde{x}$, we must first order the observations from smallest to largest, $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$. When we have done so, the sample median is defined as

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n+1)/2}\right) & \text{if } n \text{ is even} \end{cases}$$

Example 1.30 Suppose $\mathbf{x} = \{x_1, x_2, \ldots, x_n\} = \{1, 3, 5, 7, 9\}$.

(a) What is $\tilde{x}$?
(b) What happens to $\tilde{x}$ if $\mathbf{x} = \{1, 3, 5, 7, 79\}$?

SOLUTION In both cases, the sample size is odd with $n = 5$. The central value is that value having the same number of smaller values as larger values. In this case,

(a) $\tilde{x} = x_3 = 5$
(b) $\tilde{x} = x_3 = 5$

so that the (possibly erroneous) extreme value does not have any effect on this measure of the central tendency.

Example 1.31 Suppose that in 100 samples of a soil at a particular site, 99 have cohesion values of 1 kPa and 1 has a cohesion value of 3901 kPa (presumably this single sample was of a boulder or an error). What are the mean and median cohesion values at the site?

SOLUTION The mean cohesion is

$$\bar{x} = \frac{1}{100}(1 + 1 + \cdots + 1 + 3901) = 40 \quad \text{kPa}$$

The median cohesion is

$$\tilde{x} = 1 \quad \text{kPa}$$

Clearly, in this case, the median is a much better representation of the site. To design using the mean would almost certainly lead to failure.

1.6.3 Variance

The mean (expected value) or median of the random variable X tells where the probability distribution is "centered." The next most important characteristic of a random variable is whether the distribution is "wide," "narrow," or somewhere in between. This distribution "variability" is commonly measured by a quantity call the variance of X.

Definition Let X be a random variable with probability (mass or density) function $f_X(x)$ and mean μ_X. The *variance* σ_X^2 of X is defined by

$$\sigma_X^2 = \mathrm{Var}[X] = \mathrm{E}\left[(X-\mu_X)^2\right] = \begin{cases} \displaystyle\sum_x (x-\mu_X)^2 f_X(x) & \text{for discrete } X \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,dx & \text{for continuous } X \end{cases} \tag{1.26}$$

The variance of the random variable X is sometimes more easily computed as

$$\sigma_X^2 = \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X] = \mathrm{E}\left[X^2\right] - \mu_X^2 \tag{1.27}$$

The variance σ_X^2 has units of X^2. The square root of the variance, σ_X, is called the *standard deviation* of X, which is illustrated in Figure 1.15. Since the standard deviation has the same units as X, it is often preferable to report the standard deviation as a measure of variability.

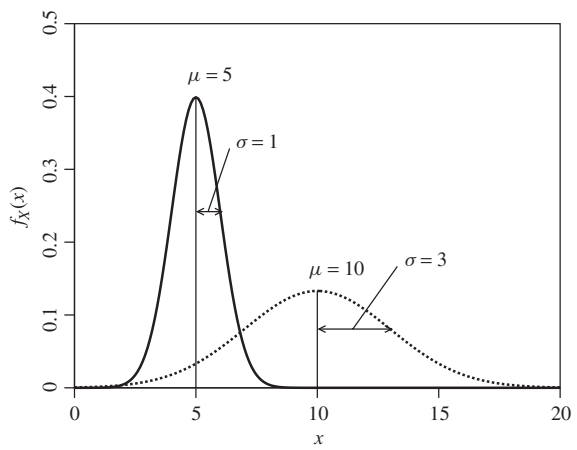


Figure 1.15 Two distributions illustrating how the position and shape change with changes in mean and variance.

Example 1.32 Recall Example 1.27. Find the variance and standard deviation of X.

SOLUTION $\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X]$
where

$$\mathrm{E}\left[X^2\right] = (-2)^2\left(\tfrac{1}{12}\right) + (-1)^2\left(\tfrac{1}{6}\right) + 0^2\left(\tfrac{1}{6}\right) + 1^2\left(\tfrac{1}{3}\right) + 2^2\left(\tfrac{1}{4}\right) = \tfrac{11}{6}$$

Thus, $\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X] = \frac{11}{6} - \left(\frac{1}{2}\right)^2 = \frac{19}{12}$ and

$$\sigma_X = \sqrt{\mathrm{Var}[X]} = \sqrt{\tfrac{19}{12}} = 1.258$$

Even though the standard deviation has the same units as the mean, it is often still not particularly informative. For example, a standard deviation of 1.0 may indicate significant variability when the mean is 1.0 but indicates virtually deterministic behavior when the mean is one million. For example, an error of 1 m on a 1-m survey would be considered unacceptable, whereas an error of 1-m on a 1000-km survey might be considered quite accurate. A measure of variability which both is nondimensional and delivers a relative sense of the magnitude of variability is the *coefficient of variation*, defined as

$$v = \frac{\sigma}{\mu} \tag{1.28}$$

Example 1.33 Recall Examples 1.27 and 1.29. What is the coefficient of variation of X?

SOLUTION

$$v_X = \frac{\sqrt{19/12}}{1/2} = 2.52$$

or about 250%, which is a highly variable process.

Note that the coefficient of variation becomes undefined if the mean of X is zero. It is, however, quite popular as a way of expressing variability in engineering, particularly for material property and load variability, which generally have nonzero means.

1.6.4 Covariance

Often one must consider more than one random variable at a time. For example, the two components of a drained soil's shear strength, $\tan(\phi')$ and c', will vary randomly from location to location in a soil. These two quantities can be modeled by two random variables, and since they may influence one another (or they may be jointly influenced by some other factor), they are characterized by a *bivariate distribution*. See Figure 1.16.

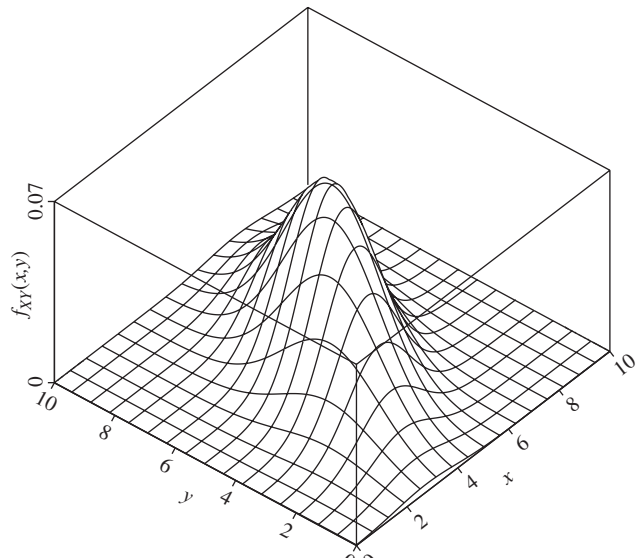


Figure 1.16 Example bivariate probability density function, $f_{XY}(x, y)$.

Properties of Bivariate Distribution

Discrete: $f_{XY}(x, y) = \mathrm{P}\left[X = x \cap Y = y\right]$

$$0 \le f_{XY}(x, y) \le 1$$

$$\sum_{\text{all } x} \sum_{\text{all } y} f_{XY}(x, y) = 1$$

Continuous: $f_{XY}(x, y)\, dx\, dy = \mathrm{P}\left[x < X \le x + dx \cap y < Y \le y + dy\right]$

$$f_{XY}(x, y) \ge 0 \text{ for all } (x, y) \in \Re^2$$

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

$$\mathrm{P}\left[x_1 < X \le x_2 \cap y_1 < Y \le y_2\right] = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x, y)\, dx\, dy$$

Definition Let X and Y be random variables with joint probability distribution $f_{XY}(x, y)$. The *covariance* between X and Y is defined by

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] \tag{1.29a}$$

$$= \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f_{XY}(x, y)$$

(discrete case)

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{XY}(x, y)\, dx\, dy$$

(continuous case) (1.29b)

The covariance between two random variables X and Y, having means μ_X and μ_Y, respectively, may also be computed as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = \mathrm{E}[XY] - \mu_X \mu_Y \tag{1.30}$$

Example 1.34 In order to determine the frequency of electrical signal transmission errors during a cone penetration test, a special cone penetrometer is constructed with redundant measuring and electrical systems. Using this penetrometer, the number of errors detected in the transmission of tip resistance during a typical cone penetration test can be measured and will be called X and the number of errors detected in the transmission of side friction will be called Y. Suppose that statistics are gathered using this penetrometer on a series of penetration tests and the following joint discrete probability mass function is estimated:

$f_{XY}(x, y)$		y (side)				
		0	1	2	3	4
x (tip)	0	0.24	0.13	0.04	0.03	0.01
	1	0.16	0.10	0.05	0.04	0.01
	2	0.08	0.05	0.01	0.00	0.00
	3	0.02	0.02	0.01	0.00	0.00

Assuming that these numbers are correct, compute

1. The expected number of errors in the transmission of the tip resistance
2. The expected number of errors in the transmission of the side friction
3. The variance of the number of errors in the transmission of the tip resistance
4. The variance of the number of errors in the transmission of the side friction
5. The covariance between the number of errors in the transmission of the tip resistance and the side friction

SOLUTION We expand the table by summing rows and columns to obtain the "marginal distributions" (i.e., unconditional distributions), $f_X(x)$ and $f_Y(y)$, of X and Y:

$f_{XY}(x, y)$		y (side)					
		0	1	2	3	4	$f_X(x)$
x (tip)	0	0.24	0.13	0.04	0.03	0.01	0.45
	1	0.16	0.10	0.05	0.04	0.01	0.36
	2	0.08	0.05	0.01	0.00	0.00	0.14
	3	0.02	0.02	0.01	0.00	0.00	0.05
$f_Y(y)$		0.50	0.30	0.11	0.07	0.02	1.00

so that

1. $\mathrm{E}[X] = \sum_x x f_X(x) = 0(0.45) + 1(0.36) + 2(0.14) + 3(0.05) = 0.79$
2. $\mathrm{E}[Y] = \sum_y y f_Y(y) = 0(0.50) + 1(0.30) + 2(0.11) + 3(0.07) + 4(0.02) = 0.81$
3. $\mathrm{E}\left[X^2\right] = \sum_x x^2 f_X(x) = 0^2(0.45) + 1^2(0.36) + 2^2(0.14) + 3^2(0.05) = 1.37$
   $\sigma_X^2 = \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X] = 1.37 - 0.79^2 = 0.75$
4. $\mathrm{E}\left[Y^2\right] = \sum_y y^2 f_Y(y) = 0^2(0.50) + 1^2(0.30) + 2^2(0.11) + 3^2(0.07) + 4^2(0.02) = 1.69$
   $\sigma_Y^2 = \mathrm{E}\left[Y^2\right] - \mathrm{E}^2[Y] = 1.69 - 0.81^2 = 1.03$
5. $\mathrm{E}[XY] = \sum_x \sum_y xy f_{XY}(x,y) = (0)(0)(0.24) + (0)(1)(0.13) + \cdots + (3)(2)(0.01) = 0.62$
   $\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$
   $= 0.62 - 0.79(0.81) = -0.02$

Although the covariance between two random variables does give information regarding the nature of the relationship, the magnitude of $\mathrm{Cov}[X,Y]$ does not indicate anything regarding the strength of the relationship. This is because $\mathrm{Cov}[X,Y]$ depends on the units and variability of X and Y. A quantity which is both normalized and nondimensional is the correlation coefficient, to be discussed next.

1.6.5 Correlation Coefficient

Definition Let X and Y be random variables with joint probability distribution $f_{XY}(x,y)$. The *correlation coefficient* between X and Y is defined to be

$$\rho_{XY} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X \sigma_Y} \tag{1.31}$$

Figure 1.17 illustrates the effect that the correlation coefficient has on the shape of a bivariate probability density function, in this case for X and Y jointly normal. If $\rho_{XY} = 0$, then the contours form ovals with axes aligned with the cartesian axes (if the variances of X and Y are equal, then the ovals are circles). When $\rho_{XY} > 0$, the ovals become stretched and the major axis has a positive slope. What this means is that when Y is large X will also tend to be large. For example, when $\rho_{XY} = 0.6$, as shown on the right plot of Figure 1.17, then when $Y = 8$, the most likely value X will take is around 7, since this is the peak of the distribution along the line $Y = 8$. Similarly, if $\rho_{XY} < 0$, then the ovals will be oriented so that the major axis has a negative slope. In this case, large values of Y will tend to give small values of X.

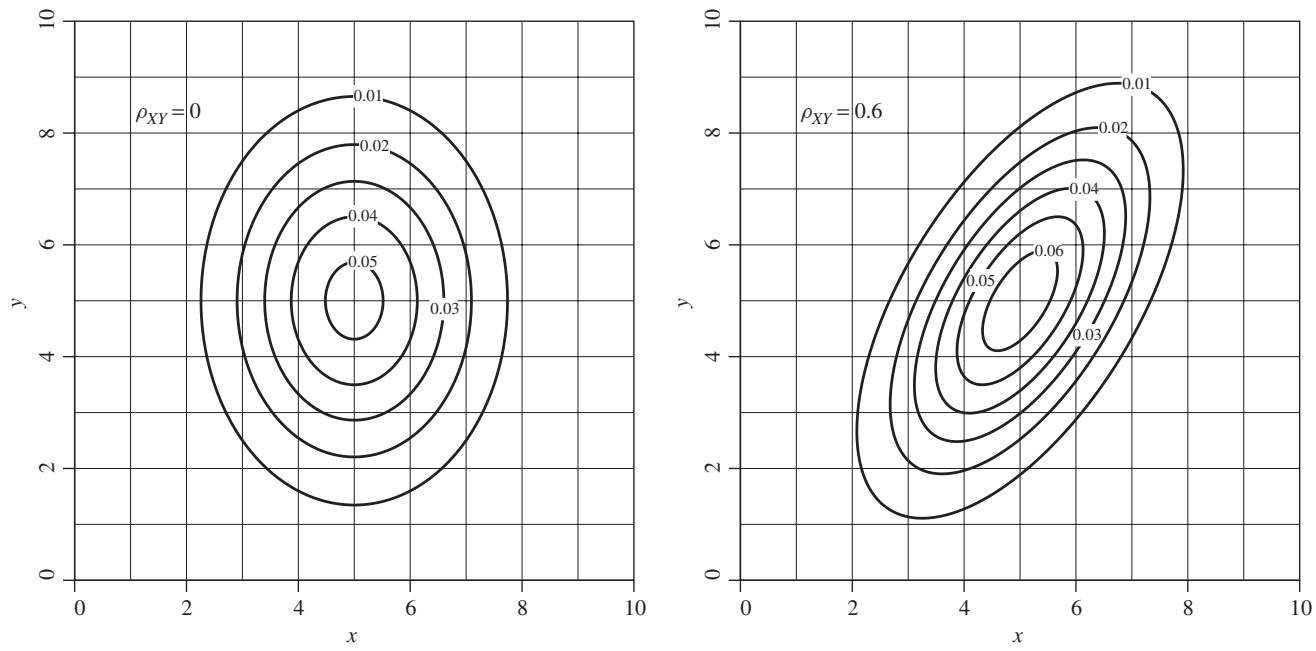


Figure 1.17 Effect of correlation coefficient ρ_{XY} on contours of a bivariate probability density function $f_{XY}(x,y)$ having $\mu_X = \mu_Y = 5$, $\sigma_X = 1.5$ and $\sigma_Y = 2.0$.

We can show that $-1 \leq \rho_{XY} \leq 1$ as follows: Consider two random variables X and Y having variances σ_X^2 and σ_Y^2, respectively, and correlation coefficient ρ_{XY}. Then

$$\begin{aligned}\text{Var}\left[\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right] &= \frac{\sigma_X^2}{\sigma_X^2} + \frac{\sigma_Y^2}{\sigma_Y^2} + 2\frac{\text{Cov}[X,Y]}{\sigma_X\sigma_Y} \\ &= 2\Big[1 + \rho_{XY}\Big] \\ &\geq 0\end{aligned}$$

which implies that $\rho_{XY} \geq -1$. Similarly,

$$\begin{aligned}\text{Var}\left[\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right] &= \frac{\sigma_X^2}{\sigma_X^2} + \frac{\sigma_Y^2}{\sigma_Y^2} - 2\frac{\text{Cov}[X,Y]}{\sigma_X\sigma_Y} \\ &= 2\Big[1 - \rho_{XY}\Big] \\ &\geq 0\end{aligned}$$

which implies that $\rho_{XY} \leq 1$. Taken together, these imply that $-1 \leq \rho_{XY} \leq 1$.

The correlation coefficient is a direct measure of the degree of *linear* dependence between X and Y. When the two variables are perfectly linearly related, ρ_{XY} will be either $+1$ or -1 ($+1$ if Y increases with X and -1 if Y decreases when X increases). When $|\rho_{XY}| < 1$, the dependence between X and Y is not completely linear; however, there could still be a strong nonlinear dependence. If two random variables X and Y are independent, then their correlation coefficient will be zero. If the correlation coefficient between two random variables X and Y is 0, *it does not mean* that they are independent, only that they are uncorrelated. Independence is a much stronger statement than is $\rho_{XY} = 0$, since the latter only implies linear independence. For example, $Y = X^2$ may be linearly independent of X (this depends on the range of X), but clearly Y and X are completely (nonlinearly) dependent.

Example 1.35 Recall Example 1.30.

1. Compute the correlation coefficient between the number of errors in the transmission of tip resistance and the number of errors in the transmission of the side friction.
2. Interpret the value you found in 1.

SOLUTION

1. $\rho_{XY} = \dfrac{-0.02}{\sqrt{0.75}\sqrt{1.03}} = -0.023$
2. With ρ_{XY} as small as -0.023, there is essentially no linear dependence between the error counts.

1.7 LINEAR COMBINATIONS OF RANDOM VARIABLES

Consider the random variables $X_1, X_2, \ldots, X_n$ and the constants $a_1, a_2, \ldots, a_n$. If

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i \qquad (1.32)$$

then Y is also a random variable, being a *linear combination* of the random variables $X_1, \ldots, X_n$. Linear combinations of random variables are common in engineering applications; any sum is a linear combination. For example, the weight of a soil mass is the sum of the weights of its constitutive particles. The bearing strength of a soil is due to the sum of the shear strengths along the potential failure surface. This section reviews the basic results associated with linear combinations.

1.7.1 Mean of Linear Combinations

The mean, or expectation, of a linear combination can be summarized by noting that *the expectation of a sum is the sum of the expectations*. Also, since constants can be brought out in front of an expectation, we have the following rules:

1. If a and b are constants, then
$$\text{E}[aX \pm b] = a\text{E}[X] \pm b \qquad (1.33)$$
2. If g and h are functions of the random variable X, then
$$\text{E}\Big[g(X) \pm h(X)\Big] = \text{E}\Big[g(X)\Big] \pm \text{E}[h(X)] \qquad (1.34)$$
3. Similarly, for any two random variables X and Y,
$$\text{E}\Big[g(X) \pm h(Y)\Big] = \text{E}\Big[g(X)\Big] \pm \text{E}[h(Y)] \qquad (1.35)$$
Note that this means, for example, $\text{E}[X \pm Y] = \text{E}[X] \pm \text{E}[Y]$.
4. If X and Y are two *uncorrelated* random variables, then
$$\text{E}[XY] = \text{E}[X]\,\text{E}[Y] \qquad (1.36)$$
by virtue of the fact that $\text{Cov}[X,Y] = \text{E}[XY] - \text{E}[X]\,\text{E}[Y] = 0$ when X and Y are uncorrelated. (This actually has nothing to do with linear combinations but often occurs in problems involving linear combinations.)

In general, if

$$Y = \sum_{i=1}^{n} a_iX_i \qquad (1.37)$$

as in Eq. 1.32, then

$$\mathrm{E}[Y] = \sum_{i=1}^{n} a_i \mathrm{E}[X_i] \tag{1.38}$$

1.7.2 Variance of Linear Combinations

The variance of a linear combination is complicated by the fact that the X_i's in the combination may or may not be correlated. If they are correlated, then the variance calculation will involve the covariances between the X_i's. In general, the following rules apply:

1. If a and b are constants, then

$$\mathrm{Var}[aX + b] = \mathrm{Var}[aX] + \mathrm{Var}[b]$$
$$= a^2 \mathrm{Var}[X] = a^2 \sigma_X^2 \tag{1.39}$$

that is, the variance of a constant is zero, and since variance is defined in terms of *squared* deviations from the mean, all quantities, including constants, are squared. Variance has units of X^2 (which is why we often prefer the standard deviation in practice).

2. If X and Y are random variables with joint probability distribution $f_{XY}(x, y)$ and a and b are constants, then

$$\mathrm{Var}[aX + bY] = a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \ \mathrm{Cov}[X, Y] \tag{1.40}$$

Note that the sign on the last term depends on the sign of a and b but that the variance terms are always positive. Note also that, if X and Y are uncorrelated, then $\mathrm{Cov}[X, Y] = 0$, so that, in this case, the above simplifies to

$$\mathrm{Var}[aX + bY] = a^2 \sigma_X^2 + b^2 \sigma_Y^2 \tag{1.41}$$

If we consider the more general case where (as in Eq. 1.37)

$$Y = \sum_{i=1}^{n} a_i X_i$$

then we have the following results:

3. If $X_1, X_2, \ldots, X_n$ are correlated, then

$$\mathrm{Var}[Y] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \ \mathrm{Cov}\left[X_i, X_j\right] \tag{1.42}$$

where we note that $\mathrm{Cov}[X_i, X_i] = \mathrm{Var}[X_i]$. If $n = 2$, the equation given in item 2 is obtained by replacing X_1 with X and X_2 with Y.

4. If $X_1, X_2, \ldots, X_n$ are uncorrelated random variables, then

$$\mathrm{Var}[a_1 X_1 + \cdots + a_n X_n]$$
$$= a_1^2 \sigma_{X_1}^2 + \cdots + a_n^2 \sigma_{X_n}^2 = \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 \tag{1.43}$$

which follows from item 3 by noting that, if X_i and X_j are uncorrelated for all $i \neq j$, then $\mathrm{Cov}\left[X_i, X_j\right] = 0$ and we are left only with the $\mathrm{Cov}[X_i, X_i] = \sigma_{X_i}^2$ terms above. This means that, if the X's are uncorrelated, then *the variance of a sum is the sum of the variances*. (However, remember that this rule *only* applies if the X's are uncorrelated.)

Example 1.36 Let X and Y be independent random variables with $\mathrm{E}[X] = 2$, $\mathrm{E}\left[X^2\right] = 29$, $\mathrm{E}[Y] = 4$, and $\mathrm{E}\left[Y^2\right] = 52$. Consider the random variables $W = X + Y$ and $Z = 2X$. The random variables W and Z are clearly dependent since they both involve X. What is their covariance? What is their correlation coefficient?

SOLUTION Given $\mathrm{E}[X] = 2$, $\mathrm{E}\left[X^2\right] = 29$, $\mathrm{E}[Y] = 4$, and $\mathrm{E}\left[Y^2\right] = 52$; X and Y independent; and $W = X + Y$ and $Z = 2X$.

Thus,

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X] = 29 - 2^2 = 25$$
$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - \mathrm{E}^2[Y] = 52 - 4^2 = 36$$
$$\mathrm{E}[W] = \mathrm{E}[X + Y] = 2 + 4 = 6$$
$$\mathrm{Var}[W] = \mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$
$$= 25 + 36 = 61$$
(due to independence)
$$\mathrm{E}[Z] = \mathrm{E}[2X] = 2(2) = 4$$
$$\mathrm{Var}[Z] = \mathrm{Var}[2X] = 4\mathrm{Var}[X] = 4(25) = 100$$
$$\mathrm{Cov}[W, Z] = \mathrm{E}[WZ] - \mathrm{E}[W]\mathrm{E}[Z]$$
$$\mathrm{E}[WZ] = \mathrm{E}[(X + Y)(2X)] = \mathrm{E}\left[2X^2 + 2XY\right]$$
$$= 2\mathrm{E}\left[X^2\right] + 2\mathrm{E}[X]\mathrm{E}[Y]$$
$$= 2(29) + 2(2)(4) = 74$$
$$\mathrm{Cov}[W, Z] = 74 - 6(4) = 50$$
$$\rho_{WZ} = \frac{50}{\sqrt{61}\sqrt{100}} = \frac{5}{\sqrt{61}} = 0.64$$

1.8 FUNCTIONS OF RANDOM VARIABLES

In general, deriving the distribution of a function of random variables [i.e., the distribution of Y where $Y = g(X_1, X_2, \ldots)$] can be quite a complex problem and exact solutions may be unknown or impractical to find.

In this section, we will cover only relatively simple cases (although even these can be difficult) and also look at some approximate approaches.

1.8.1 Functions of a Single Variable

Consider the function

$$Y = g(X) \tag{1.44}$$

and assume we know the distribution of X, that is, we know $f_X(x)$. When X takes on a specific value, that is, when $X = x$, we can compute $Y = y = g(x)$. If we assume, for now, that each value of x gives only one value of y and that each value of y arises from only one value of x (i.e., that $y = g(x)$ is a *one-to-one function*), then we must have the probability that $Y = y$ is just equal to the probability that $X = x$. That is, for discrete X,

$$\mathrm{P}\left[Y = y\right] = \mathrm{P}\left[X = x\right] = \mathrm{P}\left[X = g^{-1}(y)\right] \tag{1.45}$$

where $g^{-1}(y)$ is the inverse function, obtained by solving $y = g(x)$ for x, i.e. $x = g^{-1}(y)$. Eq. 1.45 implies that

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \tag{1.46}$$

In terms of the discrete cumulative distribution function,

$$\begin{aligned} F_Y(y) &= \mathrm{P}\left[Y \le y\right] = F_X(g^{-1}(y)) = \mathrm{P}\left[X \le g^{-1}(y)\right] \\ &= \sum_{x_i \le g^{-1}(y)} f_X(x_i) \end{aligned} \tag{1.47}$$

In the continuous case, the distribution of Y is obtained in a similar fashion. Considering Figure 1.18, the probability that X lies in a neighborhood of x_1 is the area A_1. If X lies in the shown neighborhood of x_1, Y must lie in a corresponding neighborhood of y_1 and will do so with equal probability A_1. Since the two probabilities are equal, this defines the height of the distribution of Y in the neighborhood of y_1. Considering the situation in the neighborhood of x_2, we see that the height of the distribution of Y near y_2 depends not only on A_2, which is the probability that X is in the neighborhood of x_2, but also on the slope of $y = g(x)$ at the point x_2. As the slope flattens, the height of $f(y)$ increases; that is, $f(y)$ increases as the slope decreases.

We will develop the theory by first considering the continuous analog of the discrete cumulative distribution function developed above,

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{g^{-1}(y)} f_X(x)\, dx \\ &= \int_{-\infty}^{y} f_X(g^{-1}(y)) \left[\frac{d}{dy} g^{-1}(y)\right] dy \end{aligned} \tag{1.48}$$

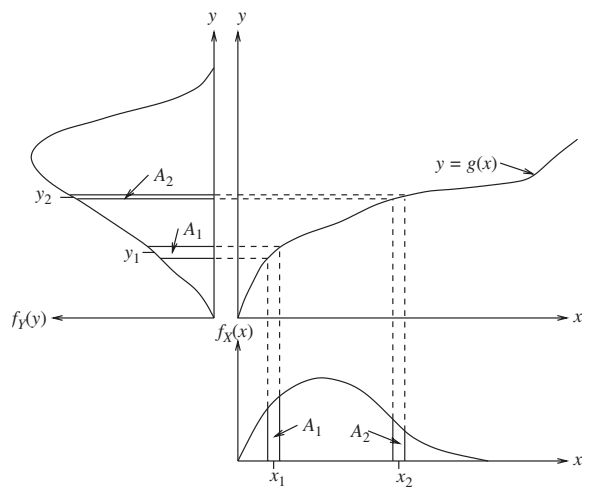


Figure 1.18 Deriving the distribution of $Y = g(X)$ from the distribution of X.

where we let $x = g^{-1}(y)$ to get the last result. To get the probability density function of Y, we can differentiate,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \left[\frac{d}{dy} g^{-1}(y)\right] \tag{1.49}$$

Note that the left-hand side here is found under the assumption that y always increases with increasing x. However, if y decreases with increasing x, then $\mathrm{P}\left[Y \le y\right]$ corresponds to $\mathrm{P}\left[X > x\right]$, leading to (see Eq. 1.47),

$$\begin{aligned} F_Y(y) &= 1 - F_X(g^{-1}(y)) \\ f_Y(y) &= f_X(g^{-1}(y)) \left[-\frac{d}{dy} g^{-1}(y)\right] \end{aligned}$$

To handle both possibilities (and since probabilities are always positive), we write

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left|\frac{d}{dy} g^{-1}(y)\right| \tag{1.50}$$

In terms of Figure 1.18 we can leave $x = g^{-1}(y)$ in the relationship and write our result as

$$f_Y(y) = f_X(x) \left|\frac{dx}{dy}\right| \tag{1.51}$$

which means that $f_Y(y)$ increases as the *inverse* of the slope, $|dx/dy|$, increases, which agrees with what is seen in Figure 1.18.

Example 1.37 Suppose that X has the following continuous distribution:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

which is the normal distribution, which we will discuss further in Section 1.10.4. If $Z = (X - \mu)/\sigma$, then what is $f_Z(z)$? (Note, we use Z intentionally here, rather than Y, because as we shall see in Section 1.10.8, Z is the so-called standard normal.)

SOLUTION In order to determine $f_Z(z)$, we need to know both $f_X(x)$ and dx/dz. We know $f_X(x)$ is the normal distribution, as shown above. To compute dx/dz we need an expression for x, which we can get by inverting the given relationship for Z (note, for the computation of the slope, we assume that both X and Z are known, and are replaced by their lowercase equivalents):

$$x = g^{-1}(z) = \mu + \sigma z$$

which gives us

$$\left|\frac{dx}{dz}\right| = \left|\frac{dg^{-1}(z)}{dz}\right| = \sigma$$

Putting these results together gives us

$$f_Z(z) = f_X(x)\left|\frac{dx}{dz}\right| = f_X(\mu + \sigma z)\,\sigma$$
$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}$$

Notice that the parameters μ and σ have now disappeared from the distribution of Z. As we shall see, Z is also normally distributed with $\mu_Z = 0$ and $\sigma_Z = 1$.

The question now arises as to what happens if the function $Y = g(X)$ is not one to one. The answer is that the probabilities of all the $X = x$ values which lead to each y are added into the probability that $Y = y$. That is, if $g(x_1), g(x_2), \ldots$ all lead to the same value of y, then

$$f_Y(y) = f_X(x_1)\left|\frac{dx_1}{dy}\right| + f_X(x_2)\left|\frac{dx_2}{dy}\right| + \cdots$$

The number of terms on the right-hand-side generally depends on y, so this computation over all y can be quite difficult. For example, the function $Y = a + bX + cX^2 + dX^3$ might have three values of x leading to the same value of y over some ranges in y but only one value of x leading to the same value of y on other ranges.

1.8.2 Functions of Two or More Random Variables

Here we consider functions of the form

$$\begin{aligned} Y_1 &= g_1(X_1, X_2, \ldots) \\ Y_2 &= g_2(X_1, X_2, \ldots) \\ &\vdots \end{aligned} \qquad (1.52)$$

In the theory which follows, we require that the number of equations above equals the number of random variables $X_1, X_2, \ldots$ and that the equations be independent so that a unique inverse can be obtained. The theory will then give us the joint distribution of $Y_1, Y_2, \ldots$ in terms of the joint distribution of $X_1, X_2, \ldots$

More commonly, we only have a single function of the form

$$Y_1 = g_1(X_1, X_2, \ldots, X_n) \qquad (1.53)$$

in which case an additional $n - 1$ *independent* equations, corresponding to $Y_2, \ldots, Y_n$, must be arbitrarily added to the problem in order to use the theory to follow. Once these equations have been added and the complete joint distribution has been found, the $n - 1$ arbitrarily added Y's can be integrated out to obtain the *marginal distribution* of Y_1. For example, if $Y_1 = X_1/X_2$ and we want the pdf of Y_1 given the joint pdf of (X_1, X_2), then we must

1. choose some function $Y_2 = g(X_1, X_2)$ which will allow us to find an inverse—for example, if we choose $Y_2 = X_2$, then we get $X_1 = Y_1 Y_2$ and $X_2 = Y_2$ as our inverse;
2. obtain the joint pdf of (Y_1, Y_2) in terms of the joint pdf of (X_1, X_2); and
3. obtain the marginal pdf of Y_1 by integrating $f_{Y_1 Y_2}$ over all possible values of Y_2.

In detail, suppose we start with the two-dimensional set of equations

$$\left.\begin{aligned} Y_1 &= g_1(X_1, X_2) \\ Y_2 &= g_2(X_1, X_2) \end{aligned}\right\} \quad \Longleftrightarrow \quad \left\{\begin{aligned} X_1 &= h_1(Y_1, Y_2) \\ X_2 &= h_2(Y_1, Y_2) \end{aligned}\right. \qquad (1.54)$$

where the right-hand equations are obtained by inverting the (given) left-hand equations. Recall that for one variable we had $f_Y(y) = f_X(x)\,|dx/dy|$. The generalization to multiple variables is

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(h_1, h_2)\,|J| \qquad (1.55)$$

where J is the *Jacobian* of the transformation,

$$J = \det \begin{bmatrix} \dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_2} \\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_2} \end{bmatrix} \qquad (1.56)$$

For more than two variables, the extension is

$$\left.\begin{aligned} Y_1 &= g_1(X_1, X_2, \ldots, X_n) \\ Y_2 &= g_2(X_1, X_2, \ldots, X_n) \\ &\;\;\vdots \\ Y_n &= g_n(X_1, X_2, \ldots, X_n) \end{aligned}\right\}$$

$$\Longleftrightarrow \quad \begin{cases} X_1 = h_1(Y_1, Y_2, \ldots, Y_n) \\ X_2 = h_2(Y_1, Y_2, \ldots, Y_n) \\ \quad \cdot \\ \quad \cdot \\ \quad \cdot \\ X_n = h_n(Y_1, Y_2, \ldots, Y_n) \end{cases} \tag{1.57}$$

$$J = \det \begin{bmatrix} \dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_2} & \cdots & \dfrac{\partial h_1}{\partial y_n} \\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_2} & \cdots & \dfrac{\partial h_2}{\partial y_n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \dfrac{\partial h_n}{\partial y_1} & \dfrac{\partial h_n}{\partial y_2} & \cdots & \dfrac{\partial h_n}{\partial y_n} \end{bmatrix} \tag{1.58}$$

and

$$f_{Y_1 Y_2 \cdots Y_n}(y_1, y_2, \ldots, y_n) = \begin{cases} f_{X_1 X_2 \cdots X_n}(h_1, h_2, \ldots, h_n)\,|J| & \\ \qquad\qquad \text{for } (y_1, y_2, \ldots, y_n) \in T & \\ 0 \qquad\qquad \text{otherwise} & \end{cases} \tag{1.59}$$

where T is the region in $\mathbf{Y}$ space corresponding to possible values of $\mathbf{x}$, specifically

$$T = \{g_1, g_2, \ldots, g_n : (x_1, x_2, \ldots, x_n) \in S\} \tag{1.60}$$

and S is the region on which $f_{X_1 X_2 \cdots X_n}$ is nonzero.

Example 1.38 Assume X_1 and X_2 are jointly distributed according to

$$f_{X_1 X_2}(x_1, x_2) = \begin{cases} 4x_1 x_2 & \text{for } 0 < x_1 < 1 \text{ and } 0 < x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

and that the following relationships exist between $\mathbf{Y}$ and $\mathbf{X}$:

$$\left.\begin{aligned} Y_1 &= \frac{X_1}{X_2} \\ Y_2 &= X_1 X_2 \end{aligned}\right\} \quad \Longleftrightarrow \quad \begin{cases} X_1 = \sqrt{Y_1 Y_2} \\ X_2 = \sqrt{\dfrac{Y_2}{Y_1}} \end{cases}$$

What is the joint pdf of (Y_1, Y_2)?

SOLUTION We first of all find the Jacobian,

$$\frac{\partial x_1}{\partial y_1} = \frac{1}{2}\sqrt{\frac{y_2}{y_1}}, \qquad \frac{\partial x_1}{\partial y_2} = \frac{1}{2}\sqrt{\frac{y_1}{y_2}}$$

$$\frac{\partial x_2}{\partial y_1} = -\frac{1}{2}\sqrt{\frac{y_2}{y_1^3}}, \qquad \frac{\partial x_2}{\partial y_2} = \frac{1}{2}\sqrt{\frac{1}{y_1 y_2}}$$

so that

$$J = \det \begin{bmatrix} \dfrac{1}{2}\sqrt{\dfrac{y_2}{y_1}} & \dfrac{1}{2}\sqrt{\dfrac{y_1}{y_2}} \\ -\dfrac{1}{2}\sqrt{\dfrac{y_2}{y_1^3}} & \dfrac{1}{2}\sqrt{\dfrac{1}{y_1 y_2}} \end{bmatrix} = \frac{1}{2y_1}$$

This gives us

$$\begin{aligned} f_{Y_1 Y_2}(y_1, y_2) &= f_{X_1 X_2}\left(\sqrt{y_1 y_2}, \sqrt{\frac{y_2}{y_1}}\right)|J| \\ &= 4\sqrt{y_1 y_2}\sqrt{\frac{y_2}{y_1}}\left(\frac{1}{2|y_1|}\right) \\ &= \frac{2y_2}{|y_1|} \end{aligned} \tag{1.61}$$

We must still determine the range of y_1 and y_2 over which this joint distribution is valid. We know that $0 < x_1 < 1$ and $0 < x_2 < 1$, so it must also be true that $0 < \sqrt{y_1 y_2} < 1$ and $0 < \sqrt{y_2/y_1} < 1$. Now, if x_1 lies between 0 and 1, then x_1^2 must also lie between 0 and 1, so we can eliminate the square root signs and write our constraints on y_1 and y_2 as

$$0 < y_1 y_2 < 1 \qquad \text{and} \qquad 0 < \frac{y_2}{y_1} < 1$$

If we consider the lines generated by replacing the inequalities above with equalities, we get the following bounding relationships:

$$y_1 y_2 = 0, \qquad y_1 y_2 = 1$$

$$\frac{y_2}{y_1} = 0, \qquad \frac{y_2}{y_1} = 1$$

If we plot these bounding relationships, the shape of the region, T, where $f_{Y_1 Y_2}$ is defined by Eq. 1.61, becomes apparent. This is illustrated in Figure 1.19.

We see from Figure 1.19 that the range, T, is defined by

$$0 < y_2 < 1 \qquad \text{and} \qquad y_2 < y_1 < \frac{1}{y_2}$$

Our joint distribution can now be completely specified as

$$f_{Y_1 Y_2}(y_1, y_2) = \begin{cases} \dfrac{2y_2}{y_1} & \text{for } 0 < y_2 < 1 \text{ and } y_2 < y_1 < \dfrac{1}{y_2} \\ 0 & \text{otherwise} \end{cases}$$

where we dropped the absolute value because y_1 is strictly positive.

Example 1.39 Consider the relationship

$$X = A\ \cos\ \Phi$$

where A and Φ are random variables with pdf $f_{A\Phi}(a, \phi)$. Assume that A and Φ are independent, that A follows a Rayleigh distribution with parameter s^2, and that Φ is uniformly distributed between 0 and 2π. What is the distribution of X?

SOLUTION First we must define a second function, Y, to give us a unique inverse relationship. Let us somewhat arbitrarily take

$$Y = A\ \sin\ \Phi$$

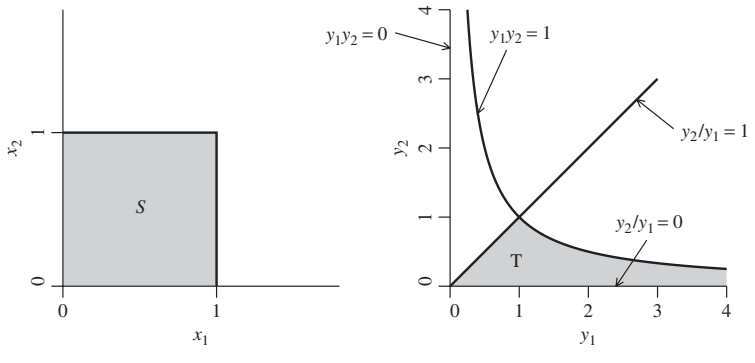


Figure 1.19 The ranges of (x_1, x_2) and (y_1, y_2) over which $f_{X_1X_2}$ and $f_{Y_1Y_2}$ are defined.

Note that there is no particular requirement for the choice of the second function so long as it leads to an inverse. This choice leads to the inverse relationships

$$\left.\begin{array}{l} X = A\ \cos\ \Phi \\ Y = A\ \sin\ \Phi \end{array}\right\} \iff \left\{\begin{array}{l} A = \sqrt{X^2 + Y^2} \\ \Phi = \tan^{-1}\left(\dfrac{Y}{X}\right) \pm 2k\pi, \qquad k = 0, 1, \ldots \end{array}\right.$$

where we have assumed that $\tan^{-1}(Y/X)$ gives a unique value between 0 and 2π—for this, we must make use of the signs of Y and X in the determination of the angle. Notice that Φ is not single valued for each X and Y.

In determining the Jacobian, we will revert to lowercase letters to emphasize that the Jacobian is deterministic (despite the fact that J itself is uppercase),

$$\frac{\partial a}{\partial x} = \frac{x}{\sqrt{x^2 + y^2}}, \qquad \frac{\partial a}{\partial y} = \frac{y}{\sqrt{x^2 + y^2}}$$
$$\frac{\partial \phi}{\partial x} = -\frac{y}{x^2 + y^2}, \qquad \frac{\partial \phi}{\partial y} = \frac{x}{x^2 + y^2}$$

so that

$$J = \det\begin{bmatrix} \dfrac{\partial a}{\partial x} & \dfrac{\partial a}{\partial y} \\ \dfrac{\partial \phi}{\partial x} & \dfrac{\partial \phi}{\partial y} \end{bmatrix} = \frac{1}{\sqrt{x^2 + y^2}}$$

Since A and Φ are independent, their joint distribution is just the product of their individual (marginal) distributions, namely $f_{A\Phi}(a, \phi) = f_A(a) f_\Phi(\phi)$. The joint distribution of X and Y is thus

$$f_{XY}(x, y) = \frac{f_A\left(\sqrt{x^2 + y^2}\right) \sum_{k=-\infty}^{\infty} f_\Phi\left(\tan^{-1}(y/x) + 2k\pi\right)}{\sqrt{x^2 + y^2}} \tag{1.62}$$

where the sum arises because Φ takes on an infinite number of possible values for each x and y—we must include the probability of each in the joint probability of X and Y.

The Rayleigh distribution, which is discussed further in Section 1.10.5, has probability density function

$$f_A(a) = \frac{a}{s^2} \exp\left\{-\frac{a^2}{2s^2}\right\}, \qquad a \geq 0$$

while the uniform distribution is

$$f_\Phi(\phi) = \frac{1}{2\pi}, \qquad 0 < \phi \leq 2\pi$$

Since Φ has zero probability of being outside the interval $(0, 2\pi]$ and exactly one value of $[\tan^{-1}(y/x) + 2k\pi]$ will lie inside that interval, then only one term in the infinite sum is nonzero and the sum simplifies to

$$\sum_{k=-\infty}^{\infty} f_\Phi\left(\tan^{-1}\left(\frac{y}{x}\right) + 2k\pi\right) = \frac{1}{2\pi}$$

In this case, Eq. 1.62 becomes

$$f_{XY}(x, y) = \frac{\sqrt{x^2 + y^2}/s^2 \exp\left\{-(x^2 + y^2)/2s^2\right\}}{2\pi\sqrt{x^2 + y^2}}$$
$$= \frac{1}{2\pi s^2} \exp\left\{-\frac{x^2 + y^2}{2s^2}\right\} \tag{1.63}$$

To find the marginal distribution of X (which was the original aim), we must integrate over all possible values of Y using the total probability theorem:

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi s^2} \exp\left\{-\frac{x^2 + y^2}{2s^2}\right\}\, dy$$
$$= \frac{e^{-x^2/(2s^2)}}{2\pi s^2} \int_{-\infty}^{\infty} e^{-y^2/(2s^2)}\, dy$$
$$= \frac{1}{\sqrt{2\pi}\, s} e^{-x^2/(2s^2)} \tag{1.64}$$

To get the final result, we needed to use the fact that

$$\int_{-\infty}^{\infty} e^{-y^2/(2s^2)}\, dy = s\sqrt{2\pi}$$

We note that Eq. 1.64 is just the normal distribution with mean zero and variance s^2. In addition, we see that Eq. 1.63 is separable and can be written as $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$, so that X and Y must be independent. A similar computation as was carried out above will show that $f_Y(y)$ is also a lognormal distribution with mean zero and variance s^2.

In summary, we see that if A is Rayleigh distributed with parameter s^2 and Φ is uniformly distributed between 0 and 2π, then

$$X = A\ \cos\ \Phi, \qquad Y = A\ \sin\ \Phi$$

will be a pair of identically normally distributed independent random variables, each with mean zero and variance s^2. As we shall see in Chapter 6, the above results suggest a very good approach to simulating normally distributed random variables.

1.8.2.1 Linear Transformations Say we have the simultaneous system of equations

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n \\ &\vdots \\ Y_n &= a_{n1}X_1 + a_{n2}X_2 + \cdots + a_{nn}X_n \end{aligned}$$

which we can write using matrix notation as

$$\mathbf{Y} = \boldsymbol{A}\mathbf{X} \tag{1.65}$$

If this relationship holds, then $\mathbf{X} = \boldsymbol{A}^{-1}\mathbf{Y}$ for nonsingular $\boldsymbol{A}$ (implies a one-to-one transformation). The joint distribution of $\mathbf{Y}$ is thus

$$f_y(\mathbf{y}) = f_x(\boldsymbol{A}^{-1}\mathbf{y})\,|J| \tag{1.66}$$

where

$$J = \det\left[\boldsymbol{A}^{-1}\right] = \frac{1}{\det[\boldsymbol{A}]} \tag{1.67}$$

Example 1.40 Say that $Y_1 = X_1 + X_2$ and that the joint pdf of $\mathbf{X}$ is

$$f_{X_1X_2}(x_1,x_2) = \begin{cases} e^{-(x_1+x_2)} & \text{for } x_1,\ x_2 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

What is the distribution of Y_1?

SOLUTION Choose $Y_2 = X_2$ as our second equation. Then

$$\left.\begin{aligned} Y_1 &= X_1 + X_2 \\ Y_2 &= X_2 \end{aligned}\right\} \quad \Longleftrightarrow \quad \left\{\begin{aligned} X_1 &= Y_1 - Y_2 \\ X_2 &= Y_2 \end{aligned}\right.$$

or

$$\begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} Y_1 \\ Y_2 \end{Bmatrix}$$

where we see from this that

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

so that $J = \det A^{-1} = 1$. This gives us

$$\begin{aligned} f_{Y_1Y_2}(y_1,y_2) &= f_{X_1X_2}(y_1 - y_2, y_2)(1) \\ &= e^{-(y_1-y_2)-y_2}, \qquad y_1 - y_2 \geq 0 \text{ and } y_2 \geq 0 \\ &= e^{-y_1}, \qquad y_1 \geq 0 \text{ and } 0 \leq y_2 \leq y_1 \end{aligned}$$

To find the distribution of Y_2, we must integrate over all possible values of Y_1 using the total probability theorem,

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f_{Y_1Y_2}(y_1,y_2)\, dy_2 = \int_0^{y_1} e^{-y_1}\, dy_2 \\ &= y_1 e^{-y_1}, \qquad y_1 \geq 0 \end{aligned}$$

In general, if $Y = X_1 + X_2$ and X_1 is independent of X_2 [so that their joint distribution can be written as the product $f_{X_1X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$], then the distribution of Y can be written as the convolution

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X_1}(y-x) f_{X_2}(x)\, dx \\ &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(y-x)\, dx \end{aligned} \tag{1.68}$$

1.8.3 Moments of Functions

In many cases the full distribution of a function of random variables is difficult to obtain. So we would like to be able to get at least the mean and variance (often the central limit theorem, discussed later, can be relied upon to suggest that the final distribution is either normal or lognormal). Obtaining just the mean and variance, at least approximately, is typically much easier than obtaining the complete distribution. In the following we will consider a function of the form $Y = g(X_1, X_2, \ldots, X_n)$ whose nth moment is defined by

$$\begin{aligned} \mathrm{E}\left[Y^n\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} & g^n(x_1,x_2,\ldots,x_n) \\ & \times f_{\mathbf{x}}(x_1,x_2,\ldots,x_n)\, dx_1\, dx_2 \cdots dx_n \end{aligned} \tag{1.69}$$

where $\mathbf{X}$ is the vector of $X's$; $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$.

1.8.3.1 Arbitrary Function of One Variable If g is an arbitrary function of one variable, $Y = g(X)$, then

$$\mathrm{E}\left[Y^n\right] = \int_{-\infty}^{\infty} g^n(x) f_X(x)\, dx \tag{1.70}$$

Various levels of approximations exist for this moment. Consider a Taylor's series expansion of $g(X)$ about μ_X,

$$Y = g(X) = g(\mu_X) + (X - \mu_X)\left.\frac{dg}{dx}\right|_{\mu_X} + \frac{1}{2}(X - \mu_X)^2 \left.\frac{d^2 g}{dx^2}\right|_{\mu_X} + \cdots \tag{1.71}$$

A *first-order* approximation to the moments uses just the first two terms of the Taylor's series expansion:

$$\mathrm{E}[Y] \simeq \mathrm{E}\left[g(\mu_X) + (X - \mu_X)\left.\frac{dg}{dx}\right|_{\mu_X}\right] = g(\mu_X) \tag{1.72a}$$

$$\mathrm{Var}[Y] \simeq \mathrm{Var}\left[g(\mu_X) + (X - \mu_X)\left.\frac{dg}{dx}\right|_{\mu_X}\right] = \mathrm{Var}[X]\left(\left.\frac{dg}{dx}\right|_{\mu_X}\right)^2 \tag{1.72b}$$

This approximation is often referred to as the *first-order second-moment* (FOSM) method. Although it is generally only accurate for small variability and small nonlinearity, it is a widely used approximation because of its simplicity (see the next section).

The *second-order* approximation uses the first three terms of the Taylor's series expansion and so is potentially more accurate:

$$\mathrm{E}[Y] \simeq g(\mu_X) + \frac{1}{2}\,\mathrm{Var}[X]\left(\left.\frac{d^2 g}{dx^2}\right|_{\mu_X}\right) \tag{1.73a}$$

$$\mathrm{Var}[Y] \simeq \mathrm{Var}[X]\left(\left.\frac{dg}{dx}\right|_{\mu_X}\right)^2 - \left(\frac{1}{2}\,\mathrm{Var}[X]\left.\frac{d^2 g}{dx^2}\right|_{\mu_X}\right)^2 + \mathrm{E}\left[(X - \mu_X)^3\right]\left(\left.\frac{dg}{dx}\,\frac{d^2 g}{dx^2}\right|_{\mu_X}\right) + \frac{1}{4}\,\mathrm{E}\left[(X - \mu_X)^4\right]\left(\left.\frac{d^2 g}{dx^2}\right|_{\mu_X}\right)^2 \tag{1.73b}$$

Notice that the second-order approximation to the variance of Y involves knowledge of the third and fourth moments of X, which are generally difficult to estimate. Often, in practice, the second-order estimate of the mean is used along with the first-order estimate of the variance, since these both require no more than second-moment estimates of X.

1.8.3.2 Arbitrary Function of Several Variables If Y is an arbitrary function of several variables, $Y = g(X_1, X_2, \ldots, X_n)$, then the corresponding Taylor's series expansion is

$$Y = g(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}) + \sum_{i=1}^{n}(X_i - \mu_{X_i})\left.\frac{\partial g}{\partial x_i}\right|_{\boldsymbol{\mu}} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \mu_{X_i})(X_j - \mu_{X_j})\left.\frac{\partial^2 g}{\partial x_i\, \partial x_j}\right|_{\boldsymbol{\mu}} + \cdots \tag{1.74}$$

where $\boldsymbol{\mu}$ is the vector of means, $\boldsymbol{\mu} = \{\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n}\}$. First-order approximations to the mean and variance of Y are then

$$\mathrm{E}[Y] \simeq g(\boldsymbol{\mu}) \tag{1.75a}$$

$$\mathrm{Var}[Y] \simeq \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}\left[X_i, X_j\right]\left[\left.\frac{\partial g}{\partial x_i}\cdot\frac{\partial g}{\partial x_j}\right|_{\boldsymbol{\mu}}\right] \tag{1.75b}$$

Second-order approximations are

$$\mathrm{E}[Y] \simeq g(\boldsymbol{\mu}) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}\left[X_i, X_j\right]\left(\left.\frac{\partial^2 g}{\partial x_i\, \partial x_j}\right|_{\boldsymbol{\mu}}\right) \tag{1.76a}$$

$$\mathrm{Var}[Y] = (\text{involves quadruple sums and fourth-order moments}) \tag{1.76b}$$

Example 1.41 The average degree of consolidation, C, under combined vertical and radial drainage is given by the relationship (e.g., Craig, 2001)

$$C = R + V - RV \tag{1.77}$$

where R is the average degree of consolidation due to horizontal (radial) drainage only and V is the average degree of consolidation due to vertical drainage only. From observations of a particular experiment which was repeated many times, suppose that we have determined the following:

$$\mu_R = \mathrm{E}[R] = 0.3, \qquad \sigma_R^2 = \mathrm{Var}[R] = 0.01$$
$$\mu_V = \mathrm{E}[V] = 0.5, \qquad \sigma_V^2 = \mathrm{Var}[V] = 0.04$$
$$\mathrm{Cov}[R, V] = 0.015, \qquad (\rho_{RV} = 0.75)$$

Estimate the mean μ_C and variance σ_C^2 of the average degree of consolidation.

SOLUTION First, we will expand Eq. 1.77 in a Taylor's series about $\boldsymbol{\mu} = (\mu_R, \mu_V)$ as follows

$$\begin{aligned} C = {} & (\mu_R + \mu_V - \mu_R\mu_V) \\ & + (R - \mu_R)\frac{\partial C}{\partial R}|_{\boldsymbol{\mu}} + \frac{1}{2}(R - \mu_R)^2\frac{\partial^2 C}{\partial R^2}|_{\boldsymbol{\mu}} \\ & + (V - \mu_V)\frac{\partial C}{\partial V}|_{\boldsymbol{\mu}} + \frac{1}{2}(V - \mu_V)^2\frac{\partial^2 C}{\partial V^2}|_{\boldsymbol{\mu}} \\ & + (R - \mu_R)(V - \mu_V)\frac{\partial^2 C}{\partial R \partial V}|_{\boldsymbol{\mu}} + \cdots \end{aligned}$$

Truncating the approximation at second-order terms and taking the expectation result in a second-order approximation to the mean:

$$\begin{aligned} \mu_C \simeq {} & (\mu_R + \mu_V - \mu_R\mu_V) \\ & + \mathrm{E}\left[R - \mu_R\right]\frac{\partial C}{\partial R}|_{\boldsymbol{\mu}} + \frac{1}{2}\mathrm{E}\left[(R - \mu_R)^2\right]\frac{\partial^2 C}{\partial R^2}|_{\boldsymbol{\mu}} \\ & + \mathrm{E}\left[V - \mu_V\right]\frac{\partial C}{\partial V}|_{\boldsymbol{\mu}} + \frac{1}{2}\mathrm{E}\left[(V - \mu_V)^2\right]\frac{\partial^2 C}{\partial V^2}|_{\boldsymbol{\mu}} \\ & + \mathrm{E}\left[(R - \mu_R)(V - \mu_V)\right]\frac{\partial^2 C}{\partial R \partial V}|_{\boldsymbol{\mu}} \\ = {} & (\mu_R + \mu_V - \mu_R\mu_V) \\ & + \frac{1}{2}\sigma_R^2\frac{\partial^2 C}{\partial R^2}|_{\boldsymbol{\mu}} \\ & + \frac{1}{2}\sigma_V^2\frac{\partial^2 C}{\partial V^2}|_{\boldsymbol{\mu}} \\ & + \mathrm{Cov}\left[R, V\right]\frac{\partial^2 C}{\partial R \partial V}|_{\boldsymbol{\mu}} \end{aligned}$$

The partial derivatives are

$$\frac{\partial^2 C}{\partial R^2} = 0, \qquad \frac{\partial^2 C}{\partial V^2} = 0, \qquad \frac{\partial^2 C}{\partial R \partial V} = -1$$

so that

$$\begin{aligned} \mu_C &= (\mu_R + \mu_V - \mu_R\mu_V) - \mathrm{Cov}\left[R, V\right] \\ &= 0.3 + 0.5 - (0.3)(0.5) - 0.015 = 0.635 \end{aligned}$$

Note that since derivatives higher than second order disappear, this result is exact and could have been obtained directly:

$$\begin{aligned} \mathrm{E}\left[C\right] &= \mathrm{E}\left[R + V - RV\right] = \mu_R + \mu_V - \mathrm{E}\left[RV\right] \\ &= \mu_R + \mu_V - \Big(\mathrm{Cov}\left[R, V\right] + \mu_R\mu_V\Big) \\ &= 0.3 + 0.5 - (0.015 + 0.3 \times 0.5) \\ &= 0.635 \end{aligned}$$

Can we also get an exact result for σ_C^2? If so, we would need to find

$$\begin{aligned} \mathrm{E}\left[C^2\right] &= \mathrm{E}\left[(R + V - RV)^2\right] \\ &= \mathrm{E}\left[R^2 + V^2 + 2RV - 2R^2V - 2RV^2 + R^2V^2\right] \end{aligned}$$

which involves third and fourth moments, which we do not know. We must therefore approximate σ_C^2. The first-order approximation involves just second-moment information, which we were given, and appears as follows:

$$\begin{aligned} \sigma_C^2 \simeq {} & \mathrm{Cov}\left[R, R\right]\left(\frac{\partial C}{\partial R}\right)^2|_{\boldsymbol{\mu}} \\ & + 2\ \mathrm{Cov}\left[R, V\right]\left(\frac{\partial C}{\partial R}\right)\left(\frac{\partial C}{\partial V}\right)|_{\boldsymbol{\mu}} \\ & + \mathrm{Cov}\left[V, V\right]\left(\frac{\partial C}{\partial V}\right)^2|_{\boldsymbol{\mu}} \end{aligned}$$

where

$$\begin{aligned} \frac{\partial C}{\partial R} &= 1 - V|_{\boldsymbol{\mu}} = 1 - \mu_V = 1 - 0.5 = 0.5 \\ \frac{\partial C}{\partial V} &= 1 - R|_{\boldsymbol{\mu}} = 1 - \mu_R = 1 - 0.3 = 0.7 \end{aligned}$$

Recalling that $\mathrm{Cov}\left[R, R\right] = \sigma_R^2$ and $\mathrm{Cov}\left[V, V\right] = \sigma_V^2$, we get

$$\begin{aligned} \sigma_C^2 \simeq {} & (0.01)(0.5)^2 + 2(0.015)(0.5)(0.7) \\ & + (0.04)(0.7)^2 = 0.0326 \end{aligned}$$

and $\sigma_C = 0.18$.

1.8.4 First-Order Second-Moment Method

The FOSM method is a relatively simple method of including the effects of variability of input variables on a resulting dependent variable. It is basically a formalized methodology based on a first-order Taylor series expansion, as discussed in the previous section. Since it is a commonly used method, it is worth describing it explicitly in this section.

The FOSM method uses a Taylor series expansion of the function to be evaluated. This expansion is truncated after the linear term (hence "first order"). The modified expansion is then used, along with the first two moments of the random variable(s), to determine the values of the first two moments of the dependent variable (hence "second moment").

Due to truncation of the Taylor series after first-order terms, the accuracy of the method deteriorates if second

and higher derivatives of the function are significant. Furthermore, the method takes no account of the form of the probability density function describing the random variables, using only their mean and standard deviation. The skewness (third moment) and higher moments are ignored.

Another limitation of the traditional FOSM method is that explicit account of spatial correlation of the random variable is not typically done. For example, the soil properties at two geotechnical sites could have identical mean and standard deviations; however, at one site the properties could vary rapidly from point to point ("low" spatial correlation length) and at another they could vary gradually ("high spatial correlation length").

Consider a function $f(X, Y)$ of two random variables X and Y. The Taylor series expansion of the function about the mean values (μ_X, μ_Y), truncated after first-order terms from Eq. 1.74, gives

$$f(X,Y) \approx f(\mu_X, \mu_Y) + (X - \mu_X)\frac{\partial f}{\partial x} + (Y - \mu_Y)\frac{\partial f}{\partial y} \tag{1.78}$$

where derivatives are evaluated at (μ_X, μ_Y).

To a first order of accuracy, the expected value of the function is given by

$$\mathrm{E}\left[f(X,Y)\right] \approx f\left(\mathrm{E}[X], \mathrm{E}[Y]\right) \tag{1.79}$$

and the variance by

$$\mathrm{Var}\left[f(X,Y)\right] \approx \mathrm{Var}\left[(X - \mu_X)\frac{\partial f}{\partial x} + (Y - \mu_Y)\frac{\partial f}{\partial y}\right] \tag{1.80}$$

Hence,

$$\mathrm{Var}\left[f(X,Y)\right] \approx \left(\frac{\partial f}{\partial x}\right)^2 \mathrm{Var}[X] + \left(\frac{\partial f}{\partial y}\right)^2 \mathrm{Var}[Y] + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\mathrm{Cov}[X,Y] \tag{1.81}$$

If X and Y are uncorrelated,

$$\mathrm{Var}\left[f(X,Y)\right] \approx \left(\frac{\partial f}{\partial x}\right)^2 \mathrm{Var}[X] + \left(\frac{\partial f}{\partial y}\right)^2 \mathrm{Var}[Y] \tag{1.82}$$

In general, for a function of n uncorrelated random variables, the FOSM method tells us that

$$\mathrm{Var}\left[f(X_1, X_2, \ldots, X_n)\right] \approx \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 \mathrm{Var}[X_i] \tag{1.83}$$

where the first derivatives are evaluated at the mean values $(\mu_{X_1}, \mu_{X_2}, \ldots, \mu_{X_n})$.

1.9 COMMON DISCRETE PROBABILITY DISTRIBUTIONS

Many engineered systems have the same statistical behavior: We generally only need a handful of probability distributions to characterize most naturally occurring phenomena. In this section, the most common discrete distribution will be reviewed (the next section looks at the most comment continuous distributions). These are the *Bernoulli family* of distributions, since they all derive from the first:

1. Bernoulli
2. Binomial
3. Geometric
4. Negative binomial
5. Poisson
6. Exponential
7. Gamma

The Poisson, exponential, and gamma are the continuous-time analogs of the binomial, geometric, and negative binomial, respectively, arising when each instant in time is viewed as an independent Bernoulli trial. In this section we consider the *discrete* members of the Bernoulli family, which are the first five members listed above, looking briefly at the main characteristics of each of these distributions and describing how they are most commonly used in practice. Included with the statistical properties of each distribution is the maximum-likelihood estimate (MLE) of their parameters. We do not formally cover the maximum-likelihood method until Section 5.2.1.2, but we present these results along with their distributions to keep everything together.

For a more complete description of these distributions, the interested reader should consult an introductory textbook on probability and statistics, such as Law and Kelton (1991) or Devore (2003).

1.9.1 Bernoulli Trials

All of the discrete distributions considered in this section (and the first two in the next section) are derived from the idea of *Bernoulli trials*. A Bernoulli trial is an experiment which has only two possible outcomes, *success* or *failure* (or [1, 0], or [true, false], or $[<5, \geq 5]$, etc). If a sequence of Bernoulli trials are mutually independent with constant (stationary) probability p of success, then the sequence is called a *Bernoulli process*. There are many examples of Bernoulli processes: One might model the failures of earth dams using a Bernoulli process. The success or failure of each of a sequence of bids made by a company might be a Bernoulli process. The failure of piles to support the load applied on them might be a Bernoulli process if it can

be assumed that the piles fail (or survive) independently and with constant probability. However, if the failure of one pile is dependent on the failure of adjacent piles, as might be the case if the soil structures are similar and load transfer takes place, the Bernoulli model may not be appropriate and a more complex, "dependent," model may be required, for example, random field modeling of the soil and finite-element analysis of the structural response within a Monte Carlo simulation. Evidently, when we depart from satisfying the assumptions underlying the simple models, such as those required for the Bernoulli model, the required models rapidly become very much more complicated. In some cases, applying the simple model to the more complex problem will yield a ballpark estimate, or at least a bound on the probability, and so it may be appropriate to proceed with a Bernoulli model taking care to treat the results as approximate. The degree of approximation depends very much on the degree of dependence between "trials" and the "stationarity" of the probability of "success," p.

If we let

$$X_j = \begin{cases} 1 & \text{if the } j\text{th trial results in a success} \\ 0 & \text{if the } j\text{th trial results in a failure} \end{cases} \tag{1.84}$$

then the Bernoulli distribution, or probability mass function, is given by

$$\begin{aligned} \mathrm{P}\left[X_j = 1\right] &= p \\ \mathrm{P}\left[X_j = 0\right] &= 1 - p = q \end{aligned} \tag{1.85}$$

for all $j = 1, 2, \ldots$. Note that we commonly denote $1 - p$ as q for simplicity.

For a single Bernoulli trial the following results hold:

$$\begin{aligned} \mathrm{E}\left[X_j\right] &= \sum_{i=0}^{1} i \cdot \mathrm{P}\left[X_j = i\right] \\ &= 0(1-p) + 1(p) = p \end{aligned} \tag{1.86a}$$

$$\mathrm{E}\left[X_j^2\right] = \sum_{i=0}^{1} i^2 \cdot \mathrm{P}\left[X_j = i\right] = 0^2(1-p) + 1^2(p) = p$$

$$\mathrm{Var}\left[X_j\right] = \mathrm{E}\left[X_j^2\right] - \mathrm{E}^2\left[X_j\right] = p - p^2 = pq \tag{1.86b}$$

For a sequence of trials, the assumption of independence between the trials means that

$$\begin{aligned} &\mathrm{P}\left[X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_n = x_n\right] \\ &\quad = \mathrm{P}\left[X_1 = x_1\right] \mathrm{P}\left[X_2 = x_2\right] \cdots \mathrm{P}\left[X_n = x_n\right] \end{aligned} \tag{1.87}$$

The MLE of p is just the average of the set of observations, $x_1, x_2, \ldots, x_n$, of X,

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1.88}$$

Notice that we use a hat to indicate that this is just an *estimate* of the true parameter p. Since the next set of observations will likely give a different value for $\hat{p}$, we see that $\hat{p}$ is actually a random variable itself, rather than the true population parameter, which is nonrandom. The mean and variance of the sequence of $\hat{p}$ can be found by considering the random $\hat{P}$,

$$\hat{P} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1.89}$$

obtained *prior* to observing the results of our Bernoulli trials. We get

$$\begin{aligned} \mathrm{E}\left[\hat{P}\right] &= \mathrm{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[X_i\right] = \frac{1}{n}(np) \\ &= p \end{aligned} \tag{1.90}$$

which means that the estimator given by Eq. 1.88 is *unbiased* (that is, the estimator is "aimed" at its desired target on average).

The estimator variance is

$$\begin{aligned} \mathrm{Var}\left[\hat{P}\right] &= \mathrm{Var}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\left[X_i\right] = \frac{1}{n^2}(npq) \\ &= \frac{pq}{n} \end{aligned} \tag{1.91}$$

where we made use of the fact that the variance of a sum is the sum of the variances *if the random variables are uncorrelated*. We are assuming that, since this is a Bernoulli process, not only are the random variables uncorrelated, but also they are completely independent (the probability of one occurring is not affected by the probability of other occurrences).

Note that the estimator variance depends on the true value of p on the right-hand-side of Eq. 1.91. Since we are estimating p, we obviously do not know the true value. The solution is to use our estimate of p to estimate its variance, so that

$$\hat{\sigma}_{\hat{P}}^2 \simeq \frac{\hat{p}\hat{q}}{n} \tag{1.92}$$

Once we have determined the estimator variance, we can compute its *standard error*, which is commonly taken to be equal to the standard deviation and which gives an

indication of how accurate our estimate is,

$$\sigma_{\hat{p}} \simeq \sqrt{\frac{\hat{p}\hat{q}}{n}} \tag{1.93}$$

For example, if $\hat{p} = 0.01$, then we would prefer $\sigma_{\hat{p}}$ to be quite a bit smaller than 0.01 and we can adjust the number of observations n to achieve this goal.

In Part 2 of this book, we will be estimating the probability of failure, p_f, of various classic geotechnical problems using a technique called *Monte Carlo simulation*. The standard error given by Eq. 1.93 will allow us to estimate the accuracy of our failure probability estimates, assuming that each "simulation" results in an independent failure/success trial.

Applications The classic Bernoulli trial is the toss of a coin, but many other experiments can lead to Bernoulli trials under the above conditions. Consider the following examples:

1. Soil anchors at a particular site have a 1% probability of pulling out. When an anchor is examined, it is classified as a success if it has not pulled out or a failure if it has. This is a Bernoulli trial with $p = 0.99$ if the anchors fail independently and if the probability of success remains constant from trial to trial.
2. Suppose that each sample of soil at a site has a 10% chance of containing significant amounts of chromium. A sample is analyzed and classified as a success if it does not contain significant amounts of chromium and a failure if it does. This is a Bernoulli trial with $p = 0.90$ if the samples are independent and if the probability of success remains constant from trial to trial.
3. A highway through a certain mountain range passes below a series of steep rock slopes. It is estimated that each rock slope has a 2% probability of failure (resulting in some amount of rock blocking the highway) over the next 10 years. If we define each rock slope as a trial which is a success if it does not fail in the next 10 years, then this can be modeled as a Bernoulli trial with $p = 0.98$ (assuming rock slopes fail independently, which might not be a good assumption if they generally fail due to, e.g., earthquakes).

1.9.2 Binomial Distribution

Let N_n be the number of successes in n Bernoulli trials, each with probability of success p. Then N_n follows a binomial distribution where

$$\mathrm{P}[N_n = k] = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, 2, \ldots, n \tag{1.94}$$

The quantity $p^k q^{n-k}$ is the probability of obtaining k successes and $n - k$ failures in n trials and $\binom{n}{k}$ is the number of possible ways of arranging the k successes over the n trials.

For example, consider eight trials which can be represented as a series of eight dashes:

___ ___ ___ ___ ___ ___ ___ ___

One possible realization of three successes in eight trials might be

F S F F S S F F

where successes are shown as S and failures as F. Another possible realization might be

S F F S F F F S

and so on. Clearly these involve three successes, which have probability p^3, and five failures, which have probability q^5. Combining these two probabilities with the fact that three successes in eight trials can be arranged in $\binom{8}{3}$ different ways leads to

$$\mathrm{P}[N_8 = 3] = \binom{8}{3} p^3 q^{8-3}$$

which generalizes to the binomial distribution for n trials and k successes given above. See Figure 1.20.

Properties In the following proofs, we make use of the binomial theorem, which states that

$$(\alpha + \beta)^n = \sum_{i=0}^{n} \binom{n}{i} \alpha^i \beta^{n-i} = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} \alpha^i \beta^{n-i} \tag{1.95}$$

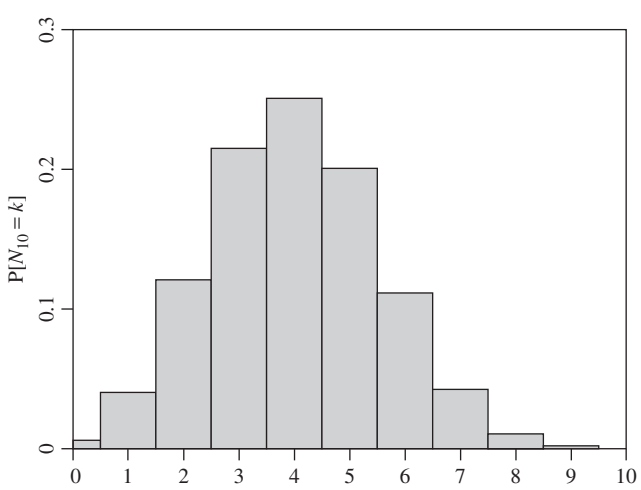


Figure 1.20 Binomial distribution for $n = 10$ and $p = 0.4$.

The expected number of successes in n trials can be found directly from the definition of the discrete-case expectation,

$$\begin{aligned}
\mathrm{E}[N_n] &= \sum_{i=0}^{n} i \binom{n}{i} p^i q^{n-i} \\
&= \sum_{i=0}^{n} i \left(\frac{n!}{i!(n-i)!} \right) p^i q^{n-i} \\
&= np \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} q^{n-i} \\
&= np \sum_{i=0}^{(n-1)} \frac{(n-1)!}{i!((n-1)-i)!} p^i q^{(n-1)-i} \\
&= np(p+q)^{n-1} \\
&= np
\end{aligned} \tag{1.96}$$

since $p + q = 1$.

Alternatively, we could write

$$\begin{aligned}
\mathrm{E}[N_n] &= \mathrm{E}[X_1 + X_2 + \cdots + X_n] \\
&= \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n] \\
&= np
\end{aligned}$$

where X_i is a Bernoulli random variable having expectation p.

To find the variance of N_n, we first need to find

$$\begin{aligned}
\mathrm{E}\left[N_n^2\right] &= \sum_{i=0}^{n} i^2 \binom{n}{i} p^i q^{n-i} = \sum_{i=1}^{n} i^2 \left(\frac{n!}{i!(n-i)!} \right) p^i q^{n-i} \\
&= np \sum_{i=1}^{n} i \left(\frac{(n-1)!}{(i-1)!(n-i)!} \right) p^{i-1} q^{n-i} \\
&= np \sum_{i=0}^{n-1} (i+1) \left(\frac{(n-1)!}{i!(n-1-i)!} \right) p^i q^{n-1-i} \\
&= np \left\{ \sum_{i=0}^{n-1} i \left(\frac{(n-1)!}{i!(n-1-i)!} \right) p^i q^{n-1-i} \right. \\
&\qquad \left. + \sum_{i=0}^{n-1} \left(\frac{(n-1)!}{i!(n-1-i)!} \right) p^i q^{n-1-i} \right\} \\
&= np\{(n-1)p + 1\} \\
&= npq + n^2 p^2
\end{aligned}$$

where for the first sum we made use of the result given by Eq. 1.96. The variance is thus

$$\mathrm{Var}[N_n] = \mathrm{E}\left[N_n^2\right] - \mathrm{E}^2[N_n] = npq + n^2p^2 - n^2p^2 = npq \tag{1.97}$$

The same result could have been obtained much more easily by considering the variance of a sum of independent random variables, since in this case the variance of a sum is the sum of the variances:

$$\mathrm{Var}[N_n] = \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}\, X_i = npq$$

The MLE of p is

$$\hat{p} = \frac{\bar{N}_n}{n} \tag{1.98}$$

if n is known, where $\bar{N}_n$ is the average of the observed values of N_n. If both n and p are unknown, see Law and Kelton (2000) for the MLE. This estimator is precisely the same as that given by Eq. 1.89 since $N_n = \sum_{i=1}^{n} X_i$, and so its mean and standard error are discussed in the previous section (with n replaced by the total number of trials making up $\bar{N}_n$).

Example 1.42 A manufacturer of geotextile sheets wishes to control the quality of its product by rejecting any lot in which the proportion of textile sheets having unacceptably low tensile strength appears to be too high. To this end, out of each large lot (1000 sheets), 25 will be selected and tested. If 5 or more of these sheets have an unacceptably low tensile strength, the entire lot will be rejected. What is the probability that a lot will be rejected if

1. 5% of the sheets in the lot have unacceptably low tensile strength?
2. 10% of the sheets in the lot have unacceptably low tensile strength?

SOLUTION

1. Let N_{25} be the number of sheets that have unacceptably low tensile strengths out of the 25 sampled.
   If the sheets fail the tension test independently with constant probability of failure, then N_{25} follows a binomial distribution with $p = 0.05$. We note that since the number of low-strength sheets in a lot is fixed, the probability of failure will change as sheets are tested. For example, if 50 out of 1000 sheets are low strength, then the probability of failure of the first sheet tested is 0.05. The probability of failure of the second sheet tested is either 49/999 or 50/999, depending on whether the first sheet tested was low strength or not. However, if the lot size (1000 in this case) is large relative to the number selected for testing (25 in this case), then the approximation that p is constant is reasonable and will lead to fairly accurate results. We will make this assumption here, so

that

$$\begin{aligned}
\mathrm{P}[N_{25} \geq 5] &= 1 - \mathrm{P}[N_{25} \leq 4] \\
&= 1 - \mathrm{P}[N_{25} = 0] - \mathrm{P}[N_{25} = 1] \\
&\quad - \mathrm{P}[N_{25} = 2] - \mathrm{P}[N_{25} = 3] \\
&\quad - \mathrm{P}[N_{25} = 4] \\
&= 1 - \binom{25}{0}(0.05)^0(0.95)^{25} \\
&\quad - \binom{25}{1}(0.05)^1(0.95)^{24} \\
&\quad - \binom{25}{2}(0.05)^2(0.95)^{23} \\
&\quad - \binom{25}{3}(0.05)^3(0.95)^{22} \\
&\quad - \binom{25}{4}(0.05)^4(0.95)^{21} \\
&= 0.00716
\end{aligned}$$

Thus, there is a very small probability of rejecting a lot where 5% of the sheets have an unacceptably low tensile strength.

2. Let N_{25} be the number of sheets that have unacceptably low tensile strengths out of the 25 sampled. Then N_{25} follows a binomial distribution with $p = 0.10$ (we will again assume sheets fail the test independently and that the probability of this happening remains constant from sheet to sheet):

$$\begin{aligned}
\mathrm{P}[N_{25} \geq 5] &= 1 - \mathrm{P}[N_{25} \leq 4] \\
&= 1 - \mathrm{P}[N_{25} = 0] - \mathrm{P}[N_{25} = 1] \\
&\quad - \mathrm{P}[N_{25} = 2] - \mathrm{P}[N_{25} = 3] \\
&\quad - \mathrm{P}[N_{25} = 4] \\
&= 1 - \binom{25}{0}(0.10)^0(0.90)^{25} \\
&\quad - \binom{25}{1}(0.10)^1(0.90)^{24} \\
&\quad - \binom{25}{2}(0.10)^2(0.90)^{23} \\
&\quad - \binom{25}{3}(0.10)^3(0.90)^{22} \\
&\quad - \binom{25}{4}(0.10)^4(0.90)^{21} \\
&= 0.098
\end{aligned}$$

There is now a reasonably high probability (about 10%) that a lot will be rejected if 10% of the sheets have an unacceptably low tensile strength.

1.9.3 Geometric Distribution

Consider a Bernoulli process in which T_1 is the number of trials required to achieve the first success. Thus, if $T_1 = 3$, then we must have had two failures followed by a success (the value of T_1 fully prescribes the sequence of trials). This has probability

$$\mathrm{P}[T_1 = 3] = \mathrm{P}[\{\text{failure, failure, success}\}] = q^2 p$$

In general

$$\mathrm{P}[T_1 = k] = q^{k-1} p, \qquad k = 1, 2, \ldots \tag{1.99}$$

Note that this is a valid probability mass function since

$$\sum_{k=1}^{\infty} q^{k-1} p = p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1$$

where we used the fact that for any $\alpha < 1$ (see, e.g., Gradshteyn and Ryzhik, 1980)

$$\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha} \tag{1.100}$$

As an example, in terms of the actual sequence of trials, the event that the first success occurs on the eighth trial appears as

F F F F F F F S

That is, the single success always occurs on the last trial. If $T_1 = 8$, then we have had seven failures, having probability q^7, and one success, having probability p. Thus

$$\mathrm{P}[T_1 = 8] = q^7 p$$

Generalizing this for $T_1 = k$ leads to the geometric distribution shown in Figure 1.21.

Because trials are assumed independent, the geometric distribution also models the number of trials *between* successes in a Bernoulli process. That is, suppose we observe the result of the Bernoulli process at trial number 1032. We will observe either a success or failure, but whichever is observed, it is now *known*. We can then ask a question such as: What is the probability that the next success occurs on trial 1040? To determine this, we start with trial 1032. Because we have observed that there is no uncertainty associated with trial 1032, it does not enter into the probability problem. However, trials 1033, 1034, ..., 1040 are unknown. We are asking for the probability that trial 1040 is the first success after 1032. In order for this *event* to occur, trials 1033–1039 must be failures. Thus, the eight trials,

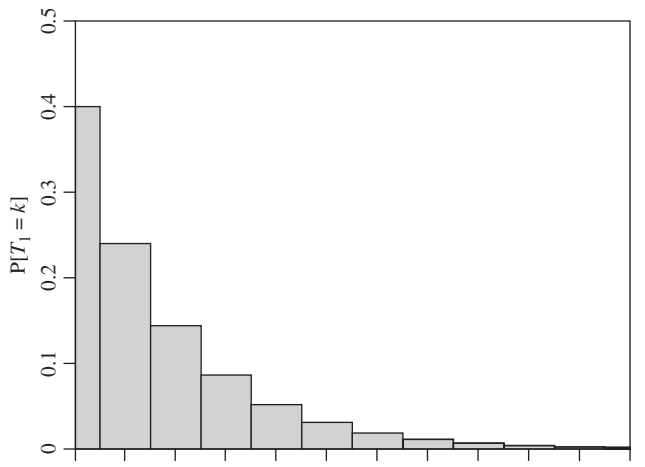


Figure 1.21 Geometric distribution for $p = 0.4$.

1033–1040, must involve seven failures (q^7) followed by one success (p). The required probability is just the product

$$\mathrm{P}[T_1 = 8] = q^7 p$$

What this means is that the geometric distribution, by virtue of the independence between trials, is *memoryless*. It does not matter when you start looking at a Bernoulli process, the number of trials to the next success is given by the geometric distribution (and is independent of the trial number).

Properties The mean of T_1, which is also sometimes referred to as the *return period* or the *mean recurrence time*, is determined as

$$\begin{aligned}\mathrm{E}[T_1] &= \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} \\ &= p\frac{d}{dq}\sum_{k=1}^{\infty} q^k = p\frac{d}{dq}\left(\frac{q}{1-q}\right) \\ &= p\left(\frac{1}{(1-q)^2}\right) = \frac{1}{p} \end{aligned} \tag{1.101}$$

where we used Eq. 1.100 to evaluate the final sum above. We will use the second to last sum in the following proof.

The variance of T_1 is obtained from $\mathrm{Var}[T_1] = \mathrm{E}\left[T_1^2\right] - \mathrm{E}^2[T_1]$ as

$$\mathrm{E}\left[T_1^2\right] = \sum_{k=1}^{\infty} k^2 pq^{k-1} = p\sum_{k=1}^{\infty} k^2 q^{k-1} = p\frac{d}{dq}\sum_{k=1}^{\infty} kq^k$$

$$\begin{aligned} &= p\frac{d}{dq}\left(\frac{q}{(1-q)^2}\right) \\ &= \frac{1}{p} + \frac{2q}{p^2} \end{aligned}$$

Thus

$$\begin{aligned}\mathrm{Var}[T_1] &= \mathrm{E}\left[T_1^2\right] - \mathrm{E}^2[T_1] \\ &= \frac{1}{p} + \frac{2q}{p^2} - \frac{1}{p^2} \\ &= \frac{q}{p^2} \end{aligned} \tag{1.102}$$

As an aside, in engineering problems, we often reverse the meaning of success and failure and use the geometric distribution to model time to failure, where time is measured in discrete steps (trials).

The MLE of p is

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{1}{\bar{t}} \tag{1.103}$$

where $t_1, t_2, \ldots, t_n$ are n independent observations of T_1.

Example 1.43 Recall the previous example where a manufacturer of geotextile sheets wishes to control the quality of its product by rejecting any lot in which the proportion of textile sheets having unacceptably low tensile strength appears to be too high. Suppose now that the sampling scheme is changed and the manufacturer decides to only sample geotextile sheets until one is encountered having an unacceptably low tensile strength. If this occurs on or before the eighth sheet tested, the entire lot will be rejected. What is the probability that a lot will be rejected if

1. 5% of the sheets in the lot have unacceptably low tensile strengths?
2. 10% of the sheets in the lot have unacceptably low tensile strengths?

If having 5% of the sheets in a lot with unacceptably low tensile strength is detrimental to the manufacturer's image and such a lot should not be sent to market, it appears that this control approach would work better than that of Example 1.39. However, if the manufacturer is more concerned with profit, this control approach is definitely not to their advantage. What might be the disadvantage of this approach from the point of view of the manufacturer? Explain with the help of a numerical example.

SOLUTION

1. Let T_1 be the trial number of the first sheet to have an unacceptably low tensile strength. Then, assuming independence between sheets and constant probability of success, T_1 follows a geometric distribution with

$p = 0.05$ and

$$\begin{aligned} \mathrm{P}[T_1 \le 8] &= \mathrm{P}[T_1 = 1] + \mathrm{P}[T_1 = 2] \\ &\quad + \cdots + \mathrm{P}[T_1 = 7] + \mathrm{P}[T_1 = 8] \\ &= 0.05 + 0.95(0.05) \\ &\quad + \cdots + 0.95^6(0.05) + 0.95^7(0.05) \\ &= 0.337 \end{aligned}$$

2. Let T_1 be the trial number of the first sheet to have an unacceptably low tensile strength. Then, under the same assumptions as in item 1, T_1 follows a geometric distribution with $p = 0.10$ and

$$\begin{aligned} \mathrm{P}[T_1 \le 8] &= \mathrm{P}[T_1 = 1] + \mathrm{P}[T_1 = 2] \\ &\quad + \cdots + \mathrm{P}[T_1 = 7] + \mathrm{P}[T_1 = 8] \\ &= 0.10 + 0.90(0.10) \\ &\quad + \cdots + 0.90^6(0.10) + 0.90^7(0.10) \\ &= 0.570 \end{aligned}$$

3. The problem with this approach, from the point of view of the manufacturer, is that a significant proportion of lots with less than 5% unacceptably low-strength sheets would be rejected (e.g., about a third). In addition, consider what happens under this quality control approach when only 2% of the sheets in the lot have unacceptably low tensile strength. (We will assume here that this is actually fairly good quality control, although, in practice, the acceptable risks can certainly vary.)
Let T_1 be the trial number of the first sheet to have an unacceptably low tensile strength. Then T_1 follows a geometric distribution, under the above assumptions, with $p = 0.02$ and

$$\begin{aligned} \mathrm{P}[T_1 \le 8] &= \mathrm{P}[T_1 = 1] + \mathrm{P}[T_1 = 2] \\ &\quad + \cdots + \mathrm{P}[T_1 = 7] + \mathrm{P}[T_1 = 8] \\ &= 0.02 + 0.98(0.02) + \cdots + 0.98^6(0.02) \\ &\quad + 0.98^7(0.02) \\ &= 0.149 \end{aligned}$$

so that there is still approximately a 15% chance that such a "good" lot would be rejected. This test does not sufficiently "resolve" the critical fraction of defectives.

1.9.4 Negative Binomial Distribution

Suppose we wish to know the number of trials (time) in a Bernoulli process until the mth success. Letting T_m be the number of trials until the mth success,

$$\mathrm{P}[T_m = k] = \binom{k-1}{m-1} p^m q^{k-m} \quad \text{for } k = m, m+1, \ldots \tag{1.104}$$

which is the *negative binomial distribution*. Whereas a binomial distributed random variable is the number of successes in a fixed number of trials, a negative binomial distributed random variable is the number of trials for a fixed number of successes. We note that the negative binomial is also often used to model the number of failures before the mth success, which results in a somewhat different distribution. We prefer the interpretation that the negative binomial distribution governs the number of trials until the mth success because it is a natural generalization of the geometric distribution and because it is then a discrete analog of the gamma distribution considered in Section 1.10.2.

The name of the negative binomial distribution arises from the *negative binomial series*

$$(1-q)^{-m} = \sum_{k=m}^{\infty} \binom{k-1}{m-1} q^{k-m} \tag{1.105}$$

which converges for $|q| < 1$. This series can be used to show that the negative binomial distribution is a valid distribution, since

$$\begin{aligned} \sum_{k=m}^{\infty} \mathrm{P}[T_m = k] &= \sum_{k=m}^{\infty} \binom{k-1}{m-1} p^m q^{k-m} \\ &= p^m \sum_{k=m}^{\infty} \binom{k-1}{m-1} q^{k-m} \\ &= p^m (1-q)^{-m} \\ &= 1 \end{aligned} \tag{1.106}$$

as expected.

We see that the geometric distribution is a special case of the negative binomial distribution with $m = 1$. The negative binomial distribution is often used to model 'time to the mth failure, where time is measured in discrete steps, or trials. Consider one possible realization which has the third success on the eighth trial:

F S S F F F F S

Another possible realization might be

F F F S F S F S

In both cases, the number of successes is 3, having probability p^3, and the number of failures is 5, having probability q^5. In terms of ordering, if $T_3 = 8$, then the third success must occur on the eighth trial (as shown above). Thus, the only other uncertainty is the ordering of the other two successes. This can occur in $\binom{7}{2}$ ways. The probability

that the third success occurs on the eighth trial is therefore given by

$$\mathrm{P}\left[T_3 = 8\right] = \binom{7}{2} p^3 q^5$$

Generalizing this for m successes and k trials leads to the negative binomial distribution shown in Eq. 1.104.

Properties The mean is determined as

$$\begin{aligned}
\mathrm{E}\left[T_m\right] &= \sum_{j=m}^{\infty} j \mathrm{P}\left[T_m = j\right] = \sum_{j=m}^{\infty} j \binom{j-1}{m-1} p^m q^{j-m} \\
&= \sum_{j=m}^{\infty} j \left(\frac{(j-1)!}{(m-1)!(j-m)!}\right) p^m q^{j-m} \\
&= m p^m \sum_{j=m}^{\infty} \left(\frac{j!}{m!(j-m)!}\right) q^{j-m} \\
&= m p^m \left[1 + (m+1)q + \frac{(m+2)(m+1)}{2!} q^2 \right. \\
&\quad \left. + \frac{(m+3)(m+2)(m+1)}{3!} q^3 + \cdots\right] \\
&= \frac{m p^m}{(1-q)^{m+1}} \\
&= \frac{m}{p} \qquad (1.107)
\end{aligned}$$

which is just m times the mean of a single geometrically distributed random variable T_1, as expected, since the number of trials between successes follows a geometric distribution. In fact, this observation leads to the following alternative representation of T_m,

$$T_m = T_{1,1} + T_{1,2} + \cdots + T_{1,m} \qquad (1.108)$$

where $T_{1,1}$ is the number of trials until the first success, $T_{1,2}$ is the number of trials after the first success until the second success, and so on. That is, the $T_{1,i}$ terms are just the times between successes. Since all trials are independent, each of the $T_{1,i}$ terms will be independent geometrically distributed random variables, all having common probability of success, p. This leads to the following much simpler computation:

$$\mathrm{E}\left[T_m\right] = \mathrm{E}\left[T_{1,1}\right] + \mathrm{E}\left[T_{1,2}\right] + \cdots + \mathrm{E}\left[T_{1,m}\right] = \frac{m}{p} \qquad (1.109)$$

since $\mathrm{E}\left[T_{1,i}\right] = 1/p$ for all $i = 1, 2, \ldots, m$. The mean in Figure 1.22 is $3/0.4 = 7.5$.

To get the variance, $\mathrm{Var}\left[T_m\right]$, we again use Eq. 1.108. Due to independence of the $T_{1,i}$ terms, the variance of the sum is the sum of the variances,

$$\begin{aligned}
\mathrm{Var}\left[T_m\right] &= \mathrm{Var}\left[T_{1,1}\right] + \mathrm{Var}\left[T_{1,2}\right] + \cdots + \mathrm{Var}\left[T_{1,m}\right] \\
&= m\ \mathrm{Var}\left[T_1\right] \\
&= \frac{mq}{p^2} \qquad (1.110)
\end{aligned}$$

which is just m times the variance of a single geometrically distributed random variable T_1, as expected.

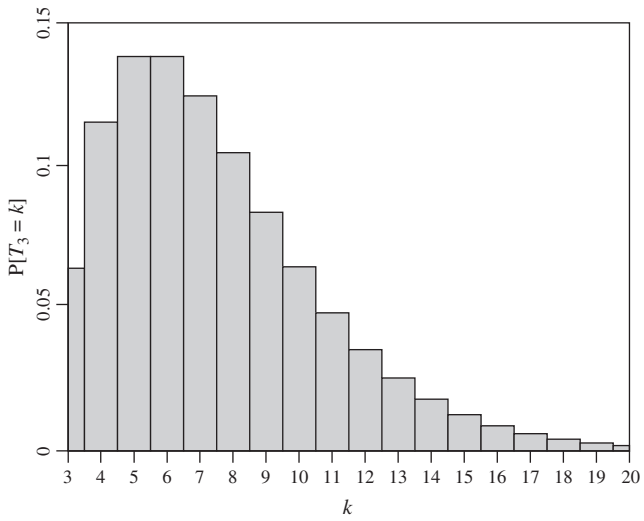


Figure 1.22 Negative binomial distribution for T_3 (i.e., $m = 3$) and $p = 0.4$.

If m is known, then the MLE of p is

$$\hat{p} = \frac{mn}{\sum_{i=1}^{n} x_i} = \frac{m}{\bar{x}} \qquad (1.111)$$

where $x_1, x_2, \ldots, x_n$ are n independent observations of T_m. If m is unknown, see Law and Kelton (2000), although beware of the fact that Law and Kelton define their negative binomial as governing the number of failures prior to the mth success, not as the number of trials until the mth success, as is done here.

Example 1.44 Consider again the problem of the tensile strength of geotextile sheets of the previous two examples. If 10% of the sheets have unacceptably low tensile strengths, what is the probability that on the next series of tests the third sheet to fail the tensile test is the eighth sheet tested?

SOLUTION Let T_3 be the number of sheets tested when the third sheet to fail the tensile test is encountered (note, this *includes* the sheet being tested). Then we are looking for

$$\mathrm{P}\left[T_3 = 8\right] = \binom{7}{2} (0.10)^3 (0.9)^{8-3} = 0.0124$$

1.9.5 Poisson Distribution

If we now allow every instant in time (or space) to be a Bernoulli trial, we get a family of three distributions: the Poisson distribution, the exponential distribution, and the gamma distribution. The latter two are continuous distributions governing the time between trial successes and are discussed in the next section. The Poisson distribution is analogous to the binomial distribution: It is derived from the binomial distribution by letting the number of trials go to infinity (one trial for each instant) and governs the number of successes in some time interval t. To see how the Poisson distribution is derived, consider the following example.

Example 1.45 Derivation from Binomial Distribution Suppose that it is known that along a certain long highway stretch an average of 1 slope subsidence occurs per year. What is the probability that exactly 10 slope subsidences will occur in the next 10-year interval?

SOLUTION If we attempt to model this using the binomial distribution, we must first divide time up into a series of intervals within each of which a slope can either subside (success) or not (failure). As a starting point, let us assume that at most one slope can subside in any half-year interval. We make this assumption because a Bernoulli trial can only have two outcomes, and if we wish to be able to count the number of subsidences, we must make these two possible outcomes either 1 (a single slope subsides) or 0 (no slopes subside). If our trials are a half-year in duration, then we have 20 trials in 10 years and the probability of a success (a slope subsides) in each trial is the rate per year divided by the number of trials per year: $p = \frac{1}{2}$. In our 10-year interval the probability we are looking for is

$$\begin{aligned}\mathrm{P}\left[10 \text{ subsidences in 10 years}\right] &\simeq \binom{20}{10}(0.5)^{10}(0.5)^{20-10} \\ &= 0.176\end{aligned}$$

Of course, we know that two or more slope subsidences could easily occur within any half-year interval. An improved solution is obtained by using a shorter trial interval. If 2-month intervals were to be used then we now have six trials per year and the probability of a slope subsidence in any interval becomes $p = \frac{1}{6}$. The number of trials in 10 years (120 months) becomes $n = \frac{120}{2} = 60$

$$\begin{aligned}\mathrm{P}\left[10 \text{ subsidences in 10 years}\right] &\simeq \binom{60}{10}\left(\frac{1}{6}\right)^{10}\left(\frac{5}{6}\right)^{50} \\ &= 0.137\end{aligned}$$

which is quite a bit more accurate.

In general, if time interval t is divided into n intervals and the mean arrival rate is λ, then

$$p = \frac{\lambda t}{n} \tag{1.112}$$

and if N_t is the number of subsidences in t years,

$$\mathrm{P}\left[N_t = k\right] = \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1 - \frac{\lambda t}{n}\right)^{n-k}$$

where λt is the mean number of subsidences ("arrivals") occurring in time interval t. If arrivals are instantaneous (so that no more than one can occur in any instant with probability 1) and can occur at any instant in time, so that each instant in time becomes a Bernoulli trial, then

$$\begin{aligned}\mathrm{P}\left[N_t = k\right] &= \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1 - \frac{\lambda t}{n}\right)^{n-k} \\ &= \lim_{n\to\infty}\left[\left\{\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n}\right\}\right. \\ &\quad \left.\times\frac{(\lambda t)^k}{k!}\left(1 - \frac{\lambda t}{n}\right)^n\left(1 - \frac{\lambda t}{n}\right)^{-k}\right]\end{aligned}$$

but since

$$\lim_{n\to\infty}\left\{\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n}\right\} = 1$$

$$\lim_{n\to\infty}\left(1 - \frac{\lambda t}{n}\right)^{-k} = 1$$

$$\lim_{n\to\infty}\left(1 - \frac{\lambda t}{n}\right)^{n} = e^{-\lambda t}$$

then our distribution simplifies to

$$\mathrm{P}\left[N_t = k\right] = \frac{(\lambda t)^k}{k!}e^{-\lambda t}$$

which is the Poisson distribution. In other words, the Poisson distribution is a limiting case of the binomial distribution, obtained when the number of trials goes to infinity, one for each instant in time, and p is replaced by the mean rate λ.

For our problem $\lambda = 1$ subsidence per year and $t = 10$ years. The probability of exactly 10 subsidences in 10 years using the Poisson distribution is

$$\mathrm{P}\left[N_{10} = 10\right] = \frac{(10)^{10}}{10!}e^{-10} = 0.125$$

and we see that the binomial model using 2-month trial intervals gives a reasonably close result (with a relative error of less than 10%).

We note that the Poisson model assumes independence between arrivals. In the subsidence problem mentioned above, there may be significant dependence between occurrences, if, for example, they are initiated by spatially

extended rainfall or freeze/thaw action. When dependence exists between trials and some common outside influence (e.g., weather), the model is complicated by the fact that the rate of occurrence becomes dependent on time. One possible solution is to apply different Poisson models for different time periods (e.g., wet season vs. dry season) or to investigate nonstationary Poisson models.

The Poisson distribution is often used to model arrival processes. We shall see in Chapter 4 that it is also useful to model "excursion" processes, for example, the number of weak pockets in a soil mass. For simplicity, we will talk about Poisson processes in time, but recognize that they can be equivalently applied over space simply by replacing t with a distance (or area, volume, etc.) measure.

For any nonzero time interval we have an infinite number of Bernoulli trials, since any time interval is made up of an infinite number of instants. Thus, the probability of success, p, in any one instant must go to zero (see Eq. 1.112); otherwise we would have an infinite number of successes in each time interval ($np \to \infty$ as $n \to \infty$). This means that we must abandon the probability of success, p, in favor of a *mean rate of success*, λ, which quantifies the mean number of successes per unit time.

The basic assumption on which the Poisson distribution rests is that each instant in time is a Bernoulli trial. Since Bernoulli trials are independent and have constant probability of success and only two possible outcomes, the Poisson process enjoys the following properties:

1. Successes (arrivals) are independently and can occur at any instant in time.
2. The mean arrival rate is constant.
3. Waiting times between arrivals are independent and exponentially distributed.
4. The time to the kth arrival is gamma distributed.

In fact, if the first two or either of the last two properties are known to hold for a sequence of arrivals, then the arrival process belongs to the Poisson family.

As in the previous example, we will define N_t to be the number of successes (arrivals or "occurrences") occurring in time t. If the above assumptions hold, then N_t is governed by the following distribution:

$$\mathrm{P}[N_t = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad k = 0, 1, 2, \ldots \tag{1.113}$$

where λ is the mean rate of occurrence (λ has units of reciprocal time). This distribution is illustrated in Figure 1.23

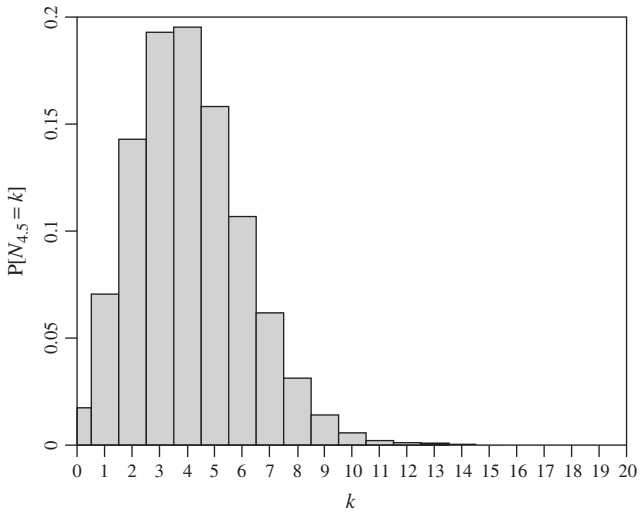


Figure 1.23 Poisson distribution for $t = 4.5$ and $\lambda = 0.9$.

Properties The mean is determined as

$$\begin{aligned}\mathrm{E}[N_t] &= \sum_{j=0}^{\infty} j \frac{(\lambda t)^j}{j!} e^{-\lambda t} = \lambda t e^{-\lambda t} \sum_{j=1}^{\infty} \frac{(\lambda t)^{j-1}}{(j-1)!} \\ &= \lambda t e^{-\lambda t} \sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} \\ &= \lambda t \end{aligned} \tag{1.114}$$

The mean of the distribution shown in Figure 1.23 is $\mathrm{E}[N_{4.5}] = 0.9(4.5) = 4.05$. To determine the variance, we first need to find

$$\begin{aligned}\mathrm{E}\left[N_t^2\right] &= \sum_{j=0}^{\infty} j^2 \frac{(\lambda t)^j}{j!} e^{-\lambda t} = \lambda t e^{-\lambda t} \sum_{j=0}^{\infty} (j+1) \frac{(\lambda t)^j}{j!} \\ &= \lambda t e^{-\lambda t} \left[\sum_{j=0}^{\infty} j \frac{(\lambda t)^j}{j!} + \sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} \right] \\ &= (\lambda t)^2 + (\lambda t) \end{aligned}$$

Thus

$$\mathrm{Var}[N_t] = \mathrm{E}\left[N_t^2\right] - \mathrm{E}^2[N_t] = \lambda t \tag{1.115}$$

That is, the mean and variance of a Poisson process are the same.

The Poisson distribution is also often written in terms of the single parameter $\nu = \lambda t$,

$$\mathrm{P}[N_t = k] = \frac{\nu^k}{k!} e^{-\nu}, \qquad k = 0, 1, 2, \ldots \tag{1.116}$$

If $x_1, x_2, \ldots, x_n$ are n independent observations of N_t, then the MLE of ν is

$$\hat{\nu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{1.117}$$

If t is known, then $\hat{\lambda} = \hat{\nu}/t$.

Example 1.46 Many research papers suggest that the arrivals of earthquakes follow a Poisson process over time. Suppose that the mean time between earthquakes is 50 years at a particular location.

1. How many earthquakes can be expected to occur during a 100-year period?
2. What is the probability that more than three earthquakes occur in a 100-year period?
3. How long must the time period be so that the probability that no earthquakes occur during that period is at most 0.1?
4. Suppose that 50 years pass without any earthquakes occurring. What is the probability that another 50 years will pass without any earthquakes occurring?

SOLUTION

1. Let N_t be the number of earthquakes occurring over t years. Then

$$\mathrm{P}[N_t = k] = \frac{(\lambda t)^k}{k!}\, e^{-\lambda t}$$

where $\lambda = \frac{1}{50} = 0.02$ per year is the mean rate of occurrence of earthquakes and $t = 100$ years. Using this, we have $\mathrm{E}[N_{100}] = 100\lambda = 100(0.02) = 2$. Thus, we can expect two earthquakes to occur during a 100-year period, which makes sense since the mean time between earthquakes is 50 years.

2. Since $\lambda t = 0.02 \times 100 = 2$, we have

$$\begin{aligned}
\mathrm{P}[N_{100} > 3] &= 1 - \mathrm{P}[N_{100} \le 3] \\
&= 1 - \Big(\mathrm{P}[N_{100} = 0] + \mathrm{P}[N_{100} = 1] \\
&\quad + \mathrm{P}[N_{100} = 2] + \mathrm{P}[N_{100} = 3]\Big) \\
&= 1 - e^{-2}\left[1 + 2 + \frac{2^2}{2} + \frac{2^3}{3!}\right] \\
&= 0.143
\end{aligned}$$

3. Let N_t be the number of occurrences over the time interval t. We want to find t such that $\mathrm{P}[N_t = 0] = e^{-\lambda t} \le 0.1$. This gives us $t \ge -\ln(0.1)/\lambda = -\ln(0.1)/0.02 = 115$ years.

4. Let N_{50} be the number of occurrences over the first 50 years and N_{100} be the number of occurrences over the first 100 years. Then, we have

$$\begin{aligned}
\mathrm{P}[N_{100} = 0 \,|\, N_{50} = 0] &= \frac{\mathrm{P}[N_{100} = 0 \,\cap\, N_{50} = 0]}{\mathrm{P}[N_{50} = 0]} \\
&= \frac{\mathrm{P}[N_{100} = 0]}{\mathrm{P}[N_{50} = 0]} = \frac{e^{-100\lambda}}{e^{-50\lambda}} \\
&= e^{-50\lambda} = e^{-1} \\
&= 0.368
\end{aligned}$$

We note that due to the memorylessness of the Poisson process (which is in turn due to the independence between trials) this result is identical to the probability of having no earthquakes in any 50-year period,

$$\mathrm{P}[N_{50} = 0] = e^{-50\lambda} = e^{-1} = 0.368$$

Now consider a Poisson process with arrival rate λ. If arrivals are retained randomly from this process with probability p and rejected with probability $q = 1 - p$, then the resulting process of retained arrivals is also Poisson with arrival rate $p\lambda$ [see Cinlar (1975) for a proof]. This is illustrated by the following example.

Example 1.47 Earthquakes in a particular region occur as a Poisson process with mean rate $\lambda = 3$ per year. In addition, it has been observed that every third earthquake, on average, has magnitude exceeding 5.

(a) What is the probability of having two or more earthquakes of magnitude in excess of 5 in the next one year?
(b) What is the probability that the next earthquake of magnitude in excess of 5 will occur within the next 2 months?

SOLUTION We are told that earthquakes occur as a Poisson process with $\lambda = 3$ per year. This means that an earthquake can occur at any instant in time but that on average there are three "successes" each year. We are also told that on average one in three of these earthquakes has a higher magnitude (i.e., exceeding 5). The "on average" part of this statement implies that each earthquake that does occur has a $\frac{1}{3}$ chance of having a higher magnitude. The mean rate of occurrence of higher magnitude earthquakes is thus $\lambda' = 1$ per year.

(a) Let N_t be the number of higher magnitude earthquakes which occur in t years. Under the above conditions,

N_t follows a Poisson distribution and the desired probability is

$$\begin{aligned} \mathrm{P}[N_1 \geq 2] &= 1 - \mathrm{P}[N_1 = 0] - \mathrm{P}[N_1 = 1] \\ &= 1 - e^{-\lambda' t}\big[1 + \lambda' t\big] \\ &= 1 - e^{-1(1)}\big[1 + 1(1)\big] \\ &= 0.2643 \end{aligned}$$

(b) The number of higher magnitude earthquakes which might occur in the next two months is $N_{1/6}$. The question is "What is the probability that one or more higher magnitude earthquakes will occur in the next two months?" which can be solved as follows:

$$\begin{aligned} \mathrm{P}\big[N_{1/6} \geq 1\big] &= 1 - \mathrm{P}\big[N_{1/6} = 0\big] = 1 - e^{-\lambda' t} \\ &= 1 - e^{-1/6} = 0.1535 \end{aligned}$$

As mentioned above, and as we will see more of shortly, the time to the next occurrence of a Poisson process is exponentially distributed (compare the above result to the exponential distribution presented in Section 1.10.1).

The previous example seems to suggest that the distribution of every third occurrence is also Poisson, which is *not* correct. This raises a rather subtle issue, but the distinction lies between whether we are selecting every third occurrence or whether we are selecting occurrences randomly with probability $\frac{1}{3}$ of success. Here are the rules and the reasoning for a process in which we are selecting every kth occurrence on average or deterministically:

1. If we are selecting every kth occurrence on average, and so randomly (i.e., the probability of selecting an occurrence is $1/k$), then the *time* until the next selection follows an exponential distribution (see Section 1.10.1) with mean rate $\lambda' = \lambda/k$, where λ is the mean occurrence rate of the original process. In this case, the likelihood of having success in the next instant is $1/k$, and the likelihood decreases exponentially thereafter. The resulting process is a Poisson process.
2. If we are selecting every kth occurrence nonrandomly (e.g., every kth customer arriving at a website is asked to fill out a survey), then the time between selections follows a gamma distribution (see Section 1.10.2). The main implication of having to have exactly $k - 1$ occurrences of the original process before a selection is that the likelihood of a selection in the next $k - 1$ instants is zero. In other words, we expect the gamma distribution to start at zero when $t = 0$. The resulting process is not Poisson.

In the above the word "likelihood" is used loosely to denote the relative probability of an occurrence in a vanishingly small time interval (i.e., an instant), dp/dt.

1.10 COMMON CONTINUOUS PROBABILITY DISTRIBUTIONS

Many naturally occurring and continuous random phenomena can be well modeled by a relatively small number of distributions. The following six continuous distributions are particularly common in engineering applications:

1. Exponential
2. Gamma
3. Uniform
4. Weibull
5. Rayleigh
6. Normal
7. Lognormal

As mentioned in the previous section, the exponential and gamma distributions are members of the *Bernoulli family*, deriving from the idea that each instant in time constitutes an independent Bernoulli trial. These are the continuous-time analogs of the geometric and negative binomial distributions.

Aside from the above, there are certainly other continuous distributions which may be considered. Distributions which involve more than two parameters are generally difficult to justify because we rarely have enough data to estimate even two parameters with much accuracy. From a practical point of view what this means is that even if a geotechnical researcher has large volumes of data at a particular site and can accurately estimate, for example, a modified six-parameter beta distribution, it is unlikely that anyone else will be able to do so at other sites. Thus, complex distributions, such as a six-parameter beta distribution, are of questionable value at any site other than the site at which it was estimated (see Chapter 4 for further discussion of this issue).

As with the common discrete distributions, this section looks briefly at the main characteristics of each of these continuous distributions and describes how they are most commonly used in practice. For a more complete description of these distributions, the interested reader should consult an introductory textbook on probability and statistics, such as Law and Kelton (1991) or Devore (2003).

1.10.1 Exponential Distribution

The exponential distribution is yet another distribution derived from the Bernoulli family: It is the continuous analog

of the geometric distribution. Recall that the geometric distribution governs the number of trials until the first success (or to the next success). If we imagine that each instant in time is now an independent trial, then the time until the first (or next) success is given by the exponential distribution (the mathematics associated with this transition from the geometric distribution involving "discrete" trials to a "continuous" sequence of trials is similar to that shown previously for the transition from the binomial to the Poisson distribution and will not be repeated here).

As with the geometric distribution, the exponential distribution is often used to describe "time-to-failure" problems. It also governs the time between arrivals of a Poisson process. If T_1 is the time to the occurrence (or failure) in question and T_1 is exponentially distributed, then its probability density function is (see Figure 1.24)

$$f_{T_1}(t) = \lambda e^{-\lambda t}, \qquad t \geq 0 \tag{1.118}$$

where λ is the *mean rate* of occurrence (or failure). Its cumulative distribution function is

$$F_{T_1}(t) = \mathrm{P}[T_1 \leq t] = 1 - e^{-\lambda t}, \qquad t \geq 0 \tag{1.119}$$

Properties

$$\mathrm{E}[T_1] = \frac{1}{\lambda} \tag{1.120a}$$

$$\mathrm{Var}[T_1] = \frac{1}{\lambda^2} \tag{1.120b}$$

That is, the mean and standard deviation of an exponentially distributed random variable are equal.

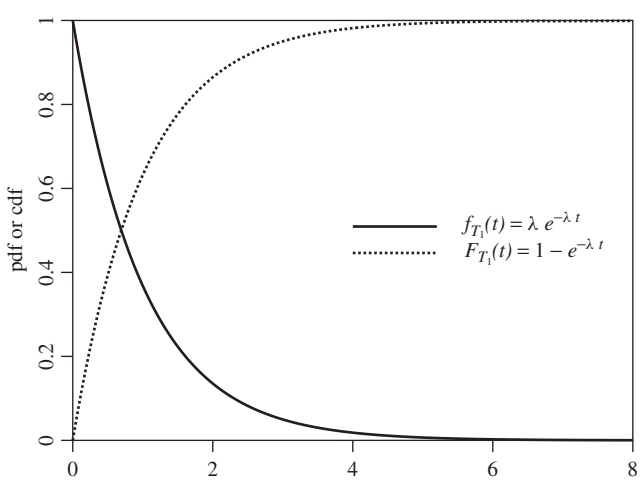


Figure 1.24 Exponential distribution for $\lambda = 1$.

Memoryless Property We will illustrate this property with an example: Let T_1 denote the time between occurrences of earthquakes in a particular region. Assume that T_1 has an exponential distribution with a mean of 4 months (i.e., on average, earthquakes in this region occur once every 4 months). Thus, T_1 has mean arrival rate of $\lambda = \frac{1}{4} = 0.25$ earthquakes per month. The probability that an earthquake occurs within the next 2 weeks (half-month) is thus

$$\begin{aligned}\mathrm{P}[T_1 < 2 \text{ weeks}] &= \mathrm{P}[T_1 < 0.5 \text{ months}] \\ &= 1 - e^{-0.5 \times 0.25} = 0.1175\end{aligned}$$

Now, suppose that we set up a ground motion accelerometer in this region and 8 months pass without an earthquake occurring. What is the probability that an earthquake will occur in the next half-month (i.e., between 8 and 8.5 months from our setup time)? Because 8 months have gone by without an earthquake occurring, you might feel that an occurrence is overdue and therefore more likely. That is, that the probability of an occurrence in the next half-month should be greater than 0.1175. However, for the exponential distribution, this is not the case, which is one of the features of the exponential distribution—the past is ignored. Each instant in time constitutes a trial which is independent of all other trials. In fact,

$$\begin{aligned}\mathrm{P}[T_1 < 8.5 \,|\, T_1 > 8] &= \frac{\mathrm{P}[8 < T_1 < 8.5]}{\mathrm{P}[T_1 > 8]} \\ &= \frac{(1 - e^{-8.5 \times 0.25}) - (1 - e^{-8 \times 0.25})}{e^{-8 \times 0.25}} \\ &= 0.1175\end{aligned}$$

Thus, after 8 months without an occurrence, the probability of an occurrence in the next half-month is the same as the probability of an occurrence in any half-month interval. We found this same property existed in the Poisson process; indeed, the times between arrivals in the Poisson process are exponentially distributed.

More generally, if T_1 is exponentially distributed with mean rate λ, then the *memoryless* property means that the probability that T_1 is greater than $t + s$, given that $T_1 > t$, is the same as the probability that T_1 is greater than s with no past history knowledge. In other words,

$$\begin{aligned}\mathrm{P}[T_1 > t + s \,|\, T_1 > t] &= \frac{\mathrm{P}[T_1 > t + s \ \cap \ T_1 > t]}{\mathrm{P}[T_1 > t]} \\ &= \frac{\mathrm{P}[T_1 > t + s]}{\mathrm{P}[T_1 > t]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= \mathrm{P}[T_1 > s]\end{aligned} \tag{1.121}$$

Link to Poisson It was mentioned above that the exponential distribution governs the time between the occurrences of a Poisson process. This can be clearly seen through the following argument: Let N_t be a Poisson distributed random variable with mean arrival rate λ. We wish to know the distribution of the time until the first arrival. Let T_1 be the time to the first arrival. Then,

$$\mathrm{P}[T_1 > t] = \mathrm{P}[N_t = 0] = \frac{(\lambda t)^0}{0!} e^{\lambda t} = e^{-\lambda t}$$

and so

$$\mathrm{P}[T_1 \leq t] = F_{T_1}(t) = 1 - e^{-\lambda t}$$

But $1 - e^{-\lambda t}$ is the cumulative distribution for the exponential probability density function $\lambda e^{-\lambda t}$. Consequently, T_1 must follow an exponential distribution with mean rate λ; that is, the time to the first occurrence in a Poisson process follows an exponential distribution with parameter λ which is equal to the Poisson rate λ. The same holds for the time *between* any occurrences of a Poisson process.

In many cases, the assumption of "independence" between trials at every instant in time makes sense (e.g., arrivals of customers at a bank, cars traveling along a highway). However, earthquakes tend to occur only once sufficient strain levels have developed between adjacent tectonic plates, and that generally takes some time. Thus, the times between measurable earthquake occurrences depend on tectonic movement rates and interplate friction, which will not generally lead to a constant probability of occurrence at each instant in time. The Poisson model is usually more reasonable for moderate to high earthquake magnitudes (in Chapter 4 we discuss the fact that higher level excursions tend to a Poisson process).

If $x_1, x_2, \ldots x_n$ are n independent observations of T_1, then the MLE of λ is

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \tag{1.122}$$

Example 1.48 Suppose the lifetime of a particular type of nuclear density meter has an exponential distribution with a mean of 28,700 h. Compute the probability of a density meter of this type failing during its 8000-h warranty?

SOLUTION Let T_1 be the lifetime of this type of density meter. Then T_1 is exponentially distributed with $\lambda = 1/28{,}700$ per hour, and

$$\begin{aligned}\mathrm{P}[T_1 < 8000] &= F_{T_1}(8000) \\ &= 1 - \exp\left\{-\frac{8000}{28{,}700}\right\} = 0.243\end{aligned}$$

Example 1.49 Let us assume that earthquakes in a certain region occur on average once every 50 years and that the number of earthquakes in any time interval follows a Poisson distribution. Under these conditions, what is the probability that less than 30 years will pass before the next earthquake occurs?

SOLUTION Let T_1 be the time to the next earthquake. Then, since the number of earthquakes follow a Poisson distribution, the time between earthquakes follows an exponential distribution. Thus, T_1 follows an exponential distribution with $\lambda = 1/50 = 0.02$ earthquakes per year (on average), and

$$\mathrm{P}\left[T_1 < 30 \text{ years}\right] = 1 - e^{-0.02 \times 30} = 0.549$$

We could also solve this using the Poisson distribution. Let N_{30} be the number of earthquakes to occur in the next 30 years. Then the event that less than 30 years will pass before the next earthquake is equivalent to the event that one or more earthquakes will occur in the next 30 years. That is,

$$\begin{aligned}\mathrm{P}\left[T_1 < 30 \text{ years}\right] &= \mathrm{P}[N_{30} \geq 1] = 1 - \mathrm{P}[N_{30} < 1] \\ &= 1 - \mathrm{P}[N_{30} = 0] = 1 - e^{-0.02 \times 30} \\ &= 0.549\end{aligned}$$

1.10.2 Gamma Distribution

We consider here a particular form of the gamma distribution which is a member of the Bernoulli family and is the continuous-time analog of the negative binomial distribution. It derives from an infinite sequence of Bernoulli trials, one at each instant in time, with mean rate of success λ, and governs the time between every kth occurrence of successes in a Poisson process. Specifically, if T_k is defined as the time to the kth success in a Poisson process, then T_k is the sum of k independent exponentially distributed random variables E_i each with parameter λ. That is, $T_k = E_1 + E_2 + \cdots + E_k$ and T_k has the probability density function

$$f_{T_k}(t) = \frac{\lambda\,(\lambda t)^{k-1}}{(k-1)!}\, e^{-\lambda t}, \qquad t \geq 0 \tag{1.123}$$

which is called the gamma distribution (Figure 1.25). This form of the gamma distribution (having integer k) is also referred to as the *k-Erlang distribution*. Note that $k = 1$ gives the exponential distribution, as expected. The above distribution can be generalized to noninteger k if $(k-1)!$ is replaced by $\Gamma(k)$, which is the gamma function; see Law and Kelton (2000) for more information on the general gamma distribution. We also give a brief discussion of noninteger k at the end of this section.

To derive the cumulative distribution function, we integrate the above probability density function (by parts) to

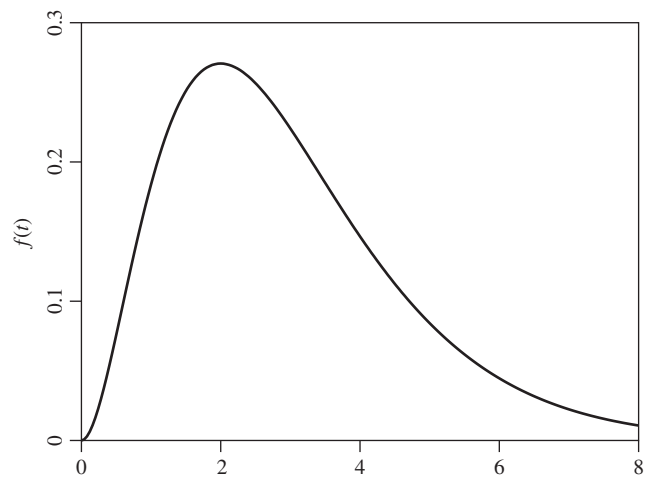


Figure 1.25 Gamma probability density function for $\lambda = 1$ and $k = 3$.

obtain, for integer k,

$$F_{T_k}(t) = \mathrm{P}[T_k \le t] = 1 - e^{-\lambda t} \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!} \tag{1.124}$$

The cumulative distribution function can also be found by recognizing that the event that the kth arrival occurs within time t (i.e., $T_k < t$) is equivalent to the event that there are k or more arrivals within time t (i.e., $N_t \ge k$). In other words,

$$\begin{aligned} F_{T_k}(t) &= \mathrm{P}[T_k \le t] = \mathrm{P}[N_t \ge k] = 1 - \mathrm{P}[N_t < k] \\ &= 1 - e^{-\lambda t} \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!} \end{aligned}$$

Properties

$$\mathrm{E}[T_k] = \frac{k}{\lambda} \qquad \left(= k\ \mathrm{E}[E_i]\right) \tag{1.125a}$$

$$\mathrm{Var}[Y] = \frac{k}{\lambda^2} \qquad \left(= k\ \mathrm{Var}[E_i]\right) \tag{1.125b}$$

If k is known and $x_1, x_2, \ldots, x_n$ are n independent observations of T_k, then the MLE of λ is

$$\hat{\lambda} = \frac{nk}{\sum_{i=1}^{n} x_i} = \frac{k}{\bar{x}} \tag{1.126}$$

Example 1.50 As in the previous example, let us assume that earthquakes in a certain region occur on average once every 50 years and that the number of earthquakes in any time interval follows a Poisson distribution. Under these conditions, what is the probability that less than 150 years will pass before two or more earthquakes occur?

SOLUTION Let T_2 be the time to the occurrence of the second earthquake. Then, since earthquakes occur according to a Poisson process, T_2 must follow a gamma distribution with $k = 2$ and $\lambda = \frac{1}{50}$ and

$$\begin{aligned} \mathrm{P}[T_2 < 150] &= F_{T_2}(150) \\ &= 1 - e^{-150/50}\left(1 + \frac{150/50}{1!}\right) = 0.801 \end{aligned}$$

Note that the same result is obtained by computing

$$\begin{aligned} \mathrm{P}[N_{150} \ge 2] &= 1 - \mathrm{P}[N_{150} < 2] \\ &= 1 - \mathrm{P}[N_{150} = 0] - \mathrm{P}[N_{150} = 1] \\ &= 1 - e^{-150/50} - \frac{150/50}{1!} e^{-150/50} \\ &= 0.801 \end{aligned}$$

The gamma distribution presented above is specialized to the sum of k independent and identically exponentially distributed random variables. It can be extended to other types of problems, so long as k is (at least approximately) a positive integer.

Example 1.51 Suppose that for clay type A the length of time in years until achieving 80% of consolidation settlement follows a gamma distribution with a mean of 4 and a variance of 8. Suppose also that for clay type B the time required to achieve the same fraction of consolidation settlement also follows a gamma distribution but with mean 4 and variance 16. Which clay type has a higher probability of reaching 80% consolidation in less than one year?

SOLUTION Let X be the time required to achieve 80% consolidation settlement for clay type A. Then X follows a gamma distribution with $\mu = k/\lambda = 4$ and $\sigma^2 = k/\lambda^2 = 8$. Solving these two equations for k and λ gives us $k = 2$ and $\lambda = \frac{1}{2}$.

Now let Y be the time required to achieve 80% consolidation settlement for clay type B. Then Y follows a gamma distribution with $\mu = k/\lambda = 4$ and $\sigma^2 = k/\lambda^2 = 16$. Solving these two equations for k and λ gives us $k = 1$ and $\lambda = \frac{1}{4}$. For clay type A we then have

$$\begin{aligned} \mathrm{P}[X < 1] &= F_{T_2}(1) \\ &= 1 - e^{-\lambda}(1 + \lambda) \\ &= 1 - e^{-1/2}(1 + \tfrac{1}{2}) \\ &= 0.0902 \end{aligned}$$

while for clay type B we have

$$\begin{aligned} \mathrm{P}[Y < 1] &= F_{T_1}(1) \\ &= 1 - e^{-\lambda} \\ &= 1 - e^{-1/4} \\ &= 0.2212 \end{aligned}$$

Thus, we are more likely to achieve 80% consolidation in under one year with clay type B.

Although the gamma distribution is not limited to integer values of k, the interpretation of the gamma PDF as the distribution of a sum of independent and identically exponentially distributed random variables is lost if k is not an integer. The more general gamma distribution has the form

$$f_X(x) = \frac{\lambda\,(\lambda x)^{k-1}}{\Gamma(k)}\, e^{-\lambda x}, \qquad x \geq 0 \tag{1.127}$$

which is valid for any $k > 0$ and $\lambda > 0$. The *gamma function* $\Gamma(k)$ for $k > 0$ is defined by the integral

$$\Gamma(k) = \int_0^\infty x^{k-1} e^{-x}\, dx \tag{1.128}$$

Tabulations of the gamma function can be found in Abramowitz and Stegun (1970), for example. When k is an integer, $\Gamma(k) = (k-1)!$.

1.10.3 Uniform Distribution

The continuous uniform distribution is the simplest of all continuous distributions since its density function is constant (over a range) (Figure 1.26). Its general definition is

$$f(x) = \frac{1}{\beta - \alpha}, \qquad \alpha \leq x \leq \beta$$

and its cumulative distribution is

$$F(x) = \mathrm{P}[X \leq x] = \frac{x - \alpha}{\beta - \alpha}, \qquad \alpha \leq x \leq \beta \tag{1.129}$$

The uniform distribution is useful in representing random variables which have known *upper* and *lower* bounds and which have equal likelihood of occurring anywhere between these bounds. Another way of looking at the uniform distribution is that it is *noninformative* or *nonpresumptive*. That is, if you know nothing else about the relative likelihood of a random variable, aside from its upper and lower bounds, then the uniform distribution is appropriate—it makes no assumptions regarding preferential likelihood of the random variable since all possible values are equally likely.

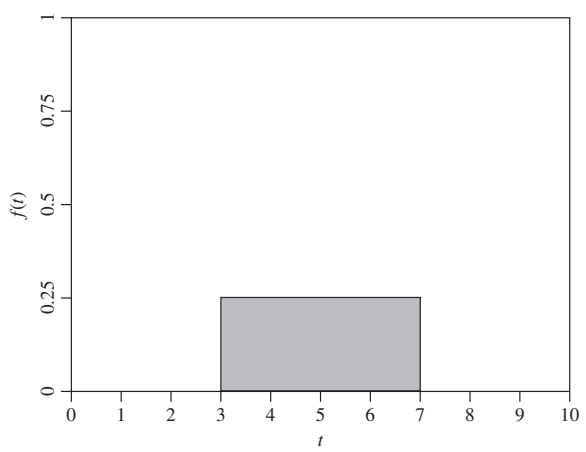


Figure 1.26 Uniform distribution for $\alpha = 3$ and $\beta = 7$.

Properties

$$\mathrm{E}[X] = \int_\alpha^\beta \frac{x\,dx}{\beta - \alpha} = \frac{\alpha + \beta}{2}$$

(this is the midpoint) (1.130a)

$$\mathrm{Var}[X] = \int_\alpha^\beta \frac{x^2}{\beta - \alpha}\, dx - \mathrm{E}^2[X] = \frac{(\beta - \alpha)^2}{12} \tag{1.130b}$$

If $x_1, x_2, \ldots, x_n$ are n independent observations of uniformly distributed X with minimum value $x_{\min}$ and maximum value $x_{\max}$, then the MLEs of α and β are

$$\hat{\alpha} = x_{\min}, \qquad \hat{\beta} = x_{\max}$$

That is, the MLEs of the lower and upper bounds of the uniform distribution are just equal to the observed minimum and maximum values.

Example 1.52 The C function `rand()` returns numbers uniformly distributed on the interval [0, `RAND_MAX`), which includes zero but excludes `RAND_MAX`. If X_i is assigned subsequent values returned by `rand()/RAND_MAX`, then each X_i is uniformly distributed on the interval $[0, 1)$. If we further define

$$Y = \alpha \left[\sum_{i=1}^{12} X_i - 6 \right]$$

then what is the mean and variance of Y?

SOLUTION

$$\begin{aligned} \mathrm{E}[Y] &= \alpha \left[\sum_{i=1}^{12} \mathrm{E}[X_i] - 6 \right] = \alpha\,[12\ \mathrm{E}[X_i] - 6] \\ &= \alpha \left[12(\tfrac{1}{2}) - 6 \right] \\ &= 0 \end{aligned}$$

$$\begin{aligned}\text{Var}[Y] &= \text{Var}\left[\alpha\left(\sum_{i=1}^{12} X_i - 6\right)\right] = \text{Var}\left[\alpha\sum_{i=1}^{12} X_i\right] \\ &= \alpha^2\ \text{Var}\left[\sum_{i=1}^{12} X_i\right] \\ &= \alpha^2 \sum_{i=1}^{12} \text{Var}[X_i] = \alpha^2(12)(\tfrac{1}{12}) \\ &= \alpha^2\end{aligned}$$

1.10.4 Weibull Distribution

Often, engineers are concerned with the strength properties of materials and the lifetimes of manufactured devices. The Weibull distribution has become very popular in describing these types of problems (Figure 1.27). One of the attractive features of the Weibull distribution is that its cumulative distribution function is quite simple.

If a continuous random variable X has a Weibull distribution, then it has probability density function

$$f(x) = \frac{\beta}{x}(\lambda x)^\beta e^{-(\lambda x)^\beta} \quad \text{for } x > 0 \tag{1.131}$$

having parameters $\lambda > 0$ and $\beta > 0$. The Weibull has a particularly simple cumulative distribution function

$$F(x) = 1 - e^{-(\lambda x)^\beta} \quad \text{if } x \geq 0 \tag{1.132}$$

Note that the exponential distribution is a special case of the Weibull distribution (simply set $\beta = 1$). While the exponential distribution has constant, memoryless failure rate, the Weibull allows for a failure rate that decreases with time ($\beta < 1$) or a failure rate that increases with time ($\beta > 1$). This gives increased flexibility for modeling lifetimes of systems that improve with time (e.g., a good red wine might have $\beta < 1$) or degrade with time (e.g., reinforced concrete bridge decks subjected to salt might have $\beta > 1$).

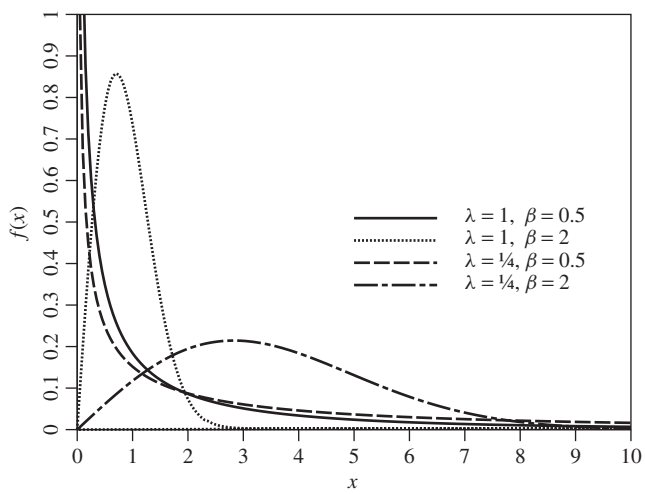


Figure 1.27 Weibull distribution.

The mean and variance of a Weibull distributed random variable are

$$\mu = \frac{1}{\lambda\beta}\,\Gamma\left(\frac{1}{\beta}\right) \tag{1.133a}$$

$$\sigma^2 = \frac{1}{\lambda^2\beta}\left\{2\Gamma\left(\frac{2}{\beta}\right) - \frac{1}{\beta}\left[\Gamma\left(\frac{1}{\beta}\right)\right]^2\right\} \tag{1.133b}$$

where Γ is the gamma function, which is commonly tabulated in math tables.

To find MLEs of λ and β, we must solve the following two equations for the estimators $\hat{\lambda}$ and $\hat{\beta}$ (Law and Kelton, 2000):

$$\frac{\sum_{i=1}^n x_i^{\hat{\beta}} \ln x_i}{\sum_{i=1}^n x_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} = \frac{1}{n}\sum_{i=1}^n \ln x_i \quad \hat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^n x_i^{\hat{\beta}}\right)^{-1/\hat{\beta}} \tag{1.134}$$

The first equation involves only $\hat{\beta}$, which can be solved for numerically. Once $\hat{\beta}$ has been obtained, the second equation can be solved directly for $\hat{\lambda}$. Thomas et al. (1969) provide an efficient general recursive formula using Newton's root-finding method,

$$\hat{\beta}_{k+1} = \hat{\beta}_k + \frac{A + (1/\hat{\beta}_k) - C_k/B_k}{(1/\hat{\beta}_k^2) + (B_k H_k - C_k^2)/B_k^2} \tag{1.135}$$

where

$$A = \frac{1}{n}\sum_{i=1}^n \ln x_i$$

$$B_k = \sum_{i=1}^n x_i^{\hat{\beta}_k}$$

$$C_k = \sum_{i=1}^n x_i^{\hat{\beta}_k} \ln x_i$$

$$H_k = \sum_{i=1}^n x_i^{\hat{\beta}_k} (\ln x_i)^2$$

An appropriate initial starting point is given by Menon (1963) and Thoman et al. (1969) to be

$$\hat{\beta}_0 = \left\{\frac{6}{(n-1)\pi^2}\left[\sum_{i=1}^n (\ln x_i)^2 - \frac{1}{n}\left(\sum_{i=1}^n \ln x_i\right)^2\right]\right\}^{-1/2} \tag{1.136}$$

See also Thoman et al. (1969) for confidence intervals on the true λ and β.

Example 1.53 The time to 90% consolidation of a sample of a certain clay has a Weibull distribution with $\beta = \frac{1}{2}$. A significant number of tests have shown that 81% of clay samples reach 90% consolidation in under 5516 h. What is the median time to attain 90% consolidation?

SOLUTION Let X be the time until a clay sample reaches 90% consolidation. Then we are told that X follows a Weibull distribution with $\beta = 0.5$. We first need to compute the other Weibull parameter, λ. To do this we make use of the fact that we know $\mathrm{P}[X < 5516] = 0.81$, and since $\mathrm{P}[X < 5516] = F(5516)$, we have

$$F(5516) = 1 - \exp\left\{-(5516\lambda)^{0.5}\right\} = 0.81$$

$$\exp\left\{-(5516\lambda)^{0.5}\right\} = 0.19$$

$$\lambda = \tfrac{1}{2000}$$

We are now looking for the median, $\tilde{x}$, which is the point which divides the distribution into half. That is, we want to find $\tilde{x}$ such that $F(\tilde{x}) = 0.5$,

$$1 - \exp\left\{-\left(\frac{\tilde{x}}{2000}\right)^{0.5}\right\} = 0.5$$

$$\exp\left\{-\left(\frac{\tilde{x}}{2000}\right)^{0.5}\right\} = 0.5$$

$$\tilde{x} = 960.9 \text{ h}$$

1.10.5 Rayleigh Distribution

The Rayleigh distribution (Figure 1.28) is a nonnegative distribution which finds application in the simulation of normally distributed random processes (see Section 3.3 and Chapter 6). In particular, consider the two orthogonal components τ_1 and τ_2 of the vector $\boldsymbol{\tau}$ in two-dimensional space. If the two components are independent and identically normally distributed random variables with zero means and common variance s^2, then the vector length $|\tau| = \sqrt{\tau_1^2 + \tau_2^2}$ will be Rayleigh distributed with probability density function

$$f(x) = \frac{x}{s^2}\exp\left\{-\frac{x^2}{2s^2}\right\}, \qquad x \geq 0 \tag{1.137}$$

and cumulative distribution function

$$F(x) = 1 - e^{-\frac{1}{2}(x/s)^2} \quad \text{if } x \geq 0 \tag{1.138}$$

which is actually a special case of the Weibull distribution ($\beta = 2$ and $\lambda = 1/(s\sqrt{2})$).

The mean and variance of a Rayleigh distributed random variable are

$$\mu = s\sqrt{\tfrac{1}{2}\pi} \qquad \sigma^2 = (2 - \tfrac{1}{2}\pi)s^2$$

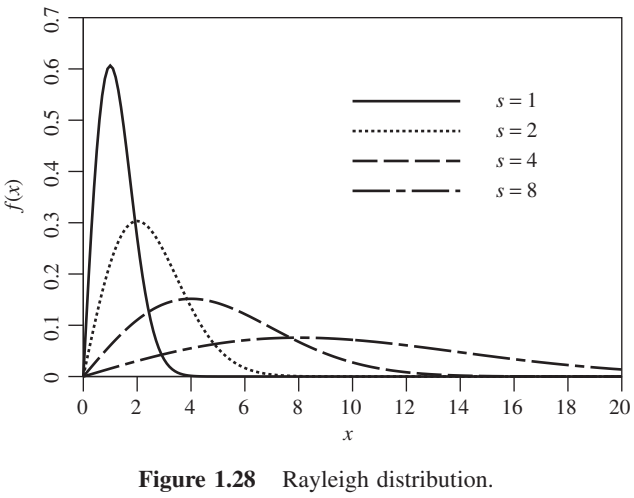


Figure 1.28 Rayleigh distribution.

1.10.6 Student t-Distribution

If Z is a standard normal random variable, V is a chi-square random variable with ν degrees of freedom, and Z and V are independent, then the random variable T defined by

$$T = \frac{Z}{\sqrt{V/\nu}} \tag{1.139}$$

follows the Student t-distribution with probability function

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}\left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < t < \infty \tag{1.140}$$

This distribution was discovered in 1908 by William Gossett, who was working for the Guinness Brewing Company in Dublin, Ireland. The company considered the discovery to be proprietary information and told Gossett he could not publish it. Gossett published it anyway under the pseudonym "Student."

Table A.2 shows values of $t_{\alpha,\nu}$ such that $\mathrm{P}\left[T > t_{\alpha,\nu}\right] = \alpha$ for commonly used values of α. We shall see more of this distribution in Chapters 2 and 3. Figure 1.29 shows some of the family of t-distributions. Notice that the t-distribution becomes wider in the tails as the number of degrees of freedom ν decreases. Conversely, as ν increases, the distribution narrows, becoming the standard normal distribution as $\nu \to \infty$. Thus, the last line of Table A.2 corresponds to the standard normal distribution, which is useful when finding z for given cumulative probability. (Note that Table A.2 is in terms of *areas to the right*.)

The mean and variance of a Student t-distributed random variable are

$$\mu = 0, \qquad \sigma^2 = \frac{\nu}{\nu - 2} \quad \text{for } \nu > 2$$

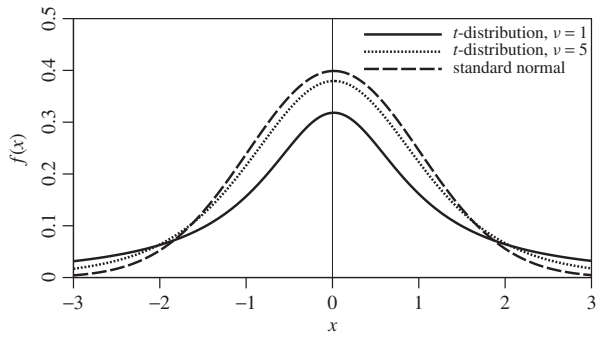


Figure 1.29 Student t-distribution.

1.10.7 Chi-Square Distribution

If $Z_1, Z_2, \ldots, Z_\nu$ are independent *standard* normal random variables [i.e., each $N(0,1)$], then the sum

$$\chi_k^2 = Z_1^2 + Z_2^2 + \cdots + Z_\nu^2 \tag{1.141}$$

has the probability density function

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})}\, x^{\nu/2-1}\, e^{-x/2} \quad \text{for } x > 0 \tag{1.142}$$

which is called a chi-square distribution with ν degrees of freedom (Figure 1.30). This is actually a special case of the gamma distribution with $k = \nu/2$ and $\lambda = \frac{1}{2}$. To get probabilities, we write

$$\mathrm{P}\left[\chi_k^2 \geq \chi_{\alpha,k}^2\right] = \alpha \;=\; \int_{\chi_{\alpha,k}^2}^{\infty} f(u)\, du \tag{1.143}$$

and use standard chi-square tables. See Table A.3. For example, $\mathrm{P}\left[\chi_{10}^2 \geq 15.99\right] = 0.10$, which is found by entering the table with $\nu = 10$ degrees of freedom, looking across for 15.99, and then reading up at the top of the table for the associated probability. Note that both Tables A.2 and A.3 are in terms of area to the right and are used with *inverse* problems where we want values on the horizontal axis having area to the right specified by a given α.

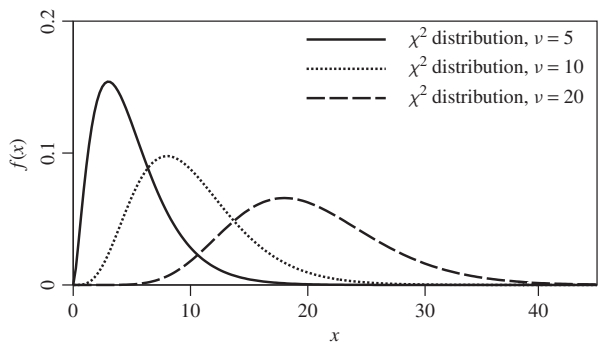


Figure 1.30 Chi-square distribution.

The mean and variance of a chi-square distributed random variable are

$$\mu = \nu, \qquad \sigma^2 = 2\nu$$

1.10.8 Normal Distribution

The *normal distribution* is probably the single most important distribution in use today (Figure 1.31). This is largely because sums of random variables tend to a normal distribution, as was proven by the *central limit theorem*—a theorem to be discussed shortly. Many natural "additive" type phenomena, or phenomena involving many accumulating factors, therefore tend to have a normal distribution. For example, the cohesive strength of a soil is due to the sum of a very large number of electrochemical interactions taking place at the molecular level; thus, the normal distribution has been widely used to represent the distribution of cohesion (its main competitor as a representative distribution is the lognormal distribution, discussed next).

A random variable X follows a normal (or *Gaussian*) distribution if its probability density function has the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad \text{for } -\infty < x < \infty \tag{1.144}$$

The notation $X \sim N(\mu, \sigma^2)$ will be used to mean that X follows a normal distribution with mean μ and variance σ^2.

Properties

1. The distribution is symmetric about the mean μ (which means that μ is also equal to the median).
2. The maximum point, or *mode*, of the distribution occurs at μ.
3. The inflection points of $f(x)$ occur at $x = \mu \pm \sigma$.

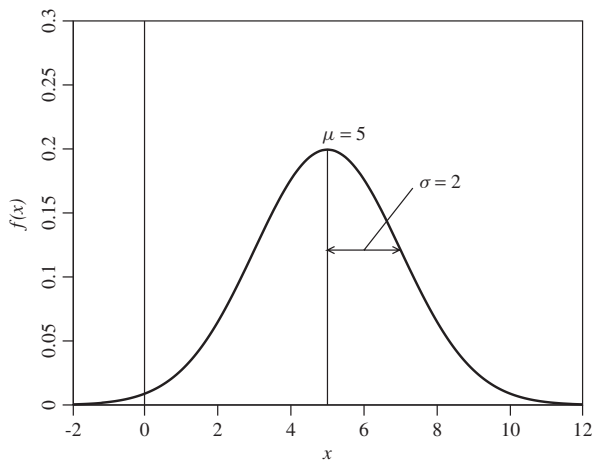


Figure 1.31 Normal distribution with $\mu = 5$ and $\sigma = 2$.

The mean and variance are given as

$$\mathrm{E}[X] = \mu, \qquad \mathrm{Var}[X] = \sigma^2$$

If $x_1, x_2, \ldots, x_n$ are n independent observations of normally distributed X, then the MLEs of μ and σ^2 are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{1.145a}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \tag{1.145b}$$

The more common estimator for σ^2 is slightly different, having the form

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \tag{1.146}$$

The latter is an *unbiased* estimator (see Section 5.2.1), which is generally more popular, especially for smaller n.

Standard Normal Unfortunately, no closed-form solution exists for the integral of the normal probability density function. Probabilities associated with the normal distribution must be obtained by numerical integration. Traditionally, this has meant that normal probabilities have had to be obtained by consulting tables presented in manuals and textbooks. Of course, no book is big enough to contain the complete set of tables necessary for all possible values of μ and σ, so some way of encapsulating the tables is necessary. As it turns out, if the random variable X is transformed by subtracting its mean and dividing by its standard deviation,

$$Z = \frac{X - \mu}{\sigma} \tag{1.147}$$

then the resulting random variable Z has mean zero and unit variance (Figure 1.32). If a probability table is developed for Z, which is called the *standard normal* variate, then probabilities for all other normally distributed random variables can be obtained by performing the above normalizing transformation. That is, probabilities for any normally distributed random variable can be obtained by performing the above transformation and then consulting the single standard normal probability table.

The distribution of the standard normal Z is given the special symbol $\phi(z)$, rather than $f(z)$, because of its importance in probability modeling and is defined by

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \quad \text{for } -\infty < z < \infty \tag{1.148}$$

The cumulative distribution function of the standard normal also has a special symbol, $\Phi(z)$, rather than $F(z)$, again because of its importance. Tables of $\Phi(z)$ are commonly included in textbooks, and one appears in Appendix

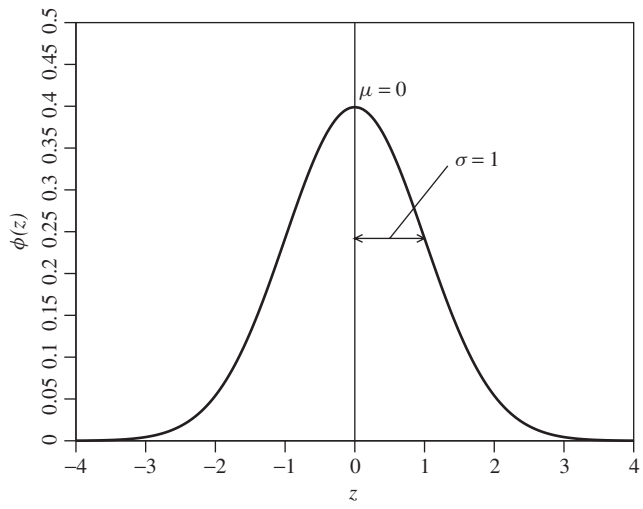


Figure 1.32 Standard normal distribution.

A. Computing probabilities for any normally distributed random variables proceeds by *standardization*, that is, by subtracting the mean and dividing by the standard deviation on both sides of the inequality in the following:

$$\begin{aligned}\mathrm{P}[X < x] &= \mathrm{P}\left[\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right] \\ &= \mathrm{P}\left[Z < \frac{x-\mu}{\sigma}\right] \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \\ &= \Phi(z)\end{aligned} \tag{1.149}$$

at which point, Table A.1 can be consulted, with $z = (x - \mu)/\sigma$, to obtain the desired probability.

Example 1.54 Suppose X is normally distributed with mean 5 and standard deviation 2. Then, what is $\mathrm{P}[X < 2.0]$?

SOLUTION In order to use Table A.1, we *standardize* on *both* sides of the inequality by subtracting the mean and dividing by the standard deviation:

$$\begin{aligned}\mathrm{P}[X < 2.0] &= \mathrm{P}\left[\frac{X-\mu}{\sigma} < \frac{2-\mu}{\sigma}\right] \\ &= \mathrm{P}\left[Z < \frac{2-5}{2}\right] = \mathrm{P}[Z < -1.5] \\ &= \Phi(-1.5)\end{aligned}$$

Table A.1 does not include negative values, so we make use of the symmetry of the standard normal. That is, the area under the distribution to the left of $z = -1.5$ (see the

figures below) is equal to the area under the distribution to the right of $z = 1.5$. The table only gives areas to the left; it is a cumulative distribution. This means that the area to the right of a point must be obtained by subtracting the area to the left from 1.0. This leaves us with

$$\mathrm{P}[X < 2.0] = 1 - \Phi(1.5) = 1 - 0.93319 = 0.06681$$

Note, for increased precision, interpolation can be used between table values, for example, if you are trying to determine $\Phi(\frac{2}{3})$. However, given the typical uncertainty in the estimates of the mean and standard deviation, there is probably little point in trying to obtain the final probability too precisely.

The probability areas involved in this question are shown below. The plot on the left illustrates the original $\mathrm{P}[X < 2]$ while the plot on the right illustrates the transformed standardized problem, $\mathrm{P}[Z < -1.5]$. The shaded areas are of equal size.

Example 1.55 The reliability of soil anchor cables against tensile failure is to be assessed. Suppose that a particular brand of cable has normally distributed tensile strength with mean 35 kN and a standard deviation of 2 kN.

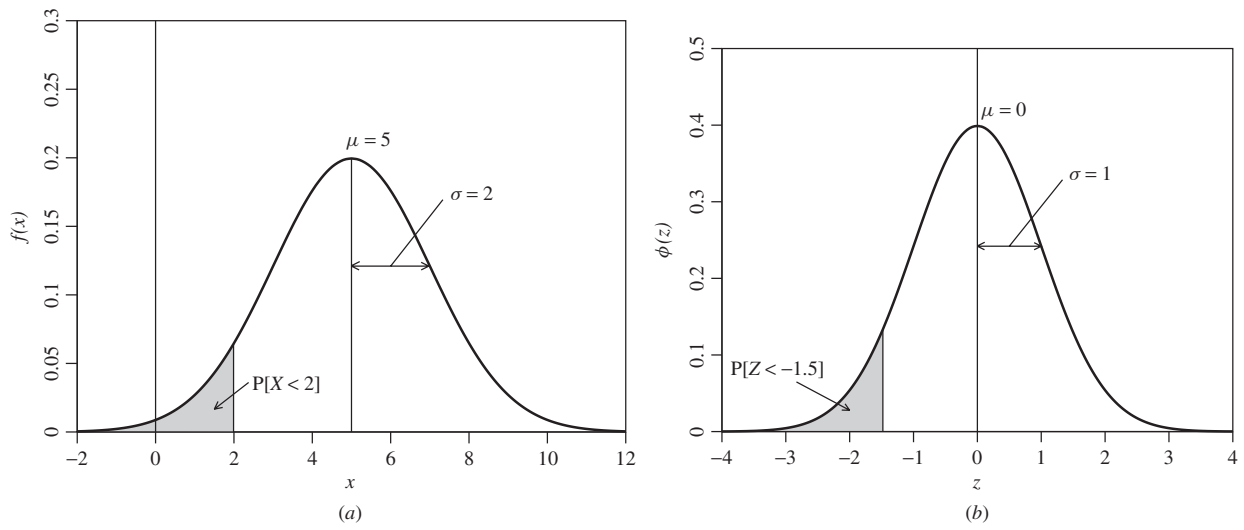
1. What is the probability that the tensile strength of a randomly selected cable is less than 40 kN?
2. Approximately 10% of all sampled cables will have a tensile strength stronger than which value?
3. Can you see any problems with modeling tensile strength using a normal distribution?

SOLUTION Let X be the tensile strength of the cable. Then X is normally distributed with mean $\mu = 35$ kN and standard deviation $\sigma = 2$ kN.

1. $\mathrm{P}[X < 40] = \mathrm{P}\left[\dfrac{X-\mu}{\sigma} < \dfrac{40-35}{2}\right] = \mathrm{P}[Z < 2.5]$ $= 0.9938$.
2. $\mathrm{P}[X > x] = 0.10 \;\rightarrow\; \mathrm{P}\left[\dfrac{X-\mu}{\sigma} > \dfrac{x-35}{2}\right]$ $= 0.10$.
   Since $\mathrm{P}[Z > 1.28] = 0.10$, we have
   $$\tfrac{1}{2}(x - 35) = 1.28 \qquad \Longrightarrow \qquad x = 37.56$$
   so that 10% of all samples are stronger than 37.56 kN.
   Note that in this solution we had to search through Table A.1 for the probability as close as possible to $1 - 0.10 = 0.9$ and then read "outwards" to see what value of z it corresponded to. A much simpler solution is to look at the last line of Table A.2 under the heading $\alpha = 0.10$. As we saw previously, Table A.2 is the inverse t-distribution, and the t-distribution collapsed to the standard normal when $\nu \to \infty$.
3. The normal distribution allows negative tensile strengths, which are not physically meaningful. This is a strong motivation for the *lognormal* distribution covered in Section 1.10.9.

1.10.8.1 Central Limit Theorem If $X_1, X_2, \ldots, X_n$ are independent random variables having arbitrary distributions, then the random variable

$$Y = X_1 + X_2 + \cdots + X_n \tag{1.150}$$



Probability plots for Example 1.54

has a normal distribution as $n \to \infty$ if all the X's have about the same mean and variance (i.e., none is dominant). See Papoulis (1991) for a proof of this theorem. In addition, if the X's are all normally distributed then Y is normally distributed for any n.

Specifically we will find the following result useful. If

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

where $X_1, X_2, \ldots, X_n$ are independent samples taken from population X having mean μ and variance σ^2 (any distribution), then

$$\lim_{n\to\infty} \mathrm{P}\left[\frac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \le x\right] = \Phi(x) \tag{1.151}$$

Implications

1. The sum of normal variates is normal (for any n) as mentioned above.
2. If the distributions of the X's are well behaved (almost normal), Then $n \ge 4$ gives a good approximation to the normal distribution.
3. If the distributions of the X's are uniform (or almost so), then $n \ge 6$ yields a reasonably good approximation to the normal distribution (out to at least about three standard deviations from the mean).
4. For poorly behaved distributions, you may need $n > 100$ before the distribution begins to look reasonably normal. This happens, for example, with distributions whose tails fall off very slowly.

Thus for n sufficiently large and $X_1, X_2, \ldots, X_n$ independent and identically distributed (iid)

$$Y = X_1 + X_2 + \cdots + X_n$$

is approximately normally distributed with

$$\mu_Y = \mathrm{E}[Y] = n\ \mathrm{E}[X_i] \tag{1.152a}$$

$$\sigma_Y^2 = \mathrm{Var}[Y] = n\ \mathrm{Var}[X_i] \tag{1.152b}$$

If the X's are *not* identically distributed but are still independent, then

$$\mu_Y = \sum_{i=1}^{n} \mathrm{E}[X_i] \tag{1.153a}$$

$$\sigma_Y^2 = \sum^{n} \mathrm{Var}[X_i] \tag{1.153b}$$

1.10.8.2 Normal Approximation to Binomial By virtue of the central limit theorem, the binomial distribution, which as you will recall arises from the sum of a sequence of Bernoulli random variables, can be approximated by the normal distribution (Figure 1.33). Specifically, if N_n is the number of successes in n trials, then

$$N_n = \sum_{i=1}^{n} X_i \tag{1.154}$$

where X_i is the outcome of a Bernoulli trial ($X_i = 1$ with probability p, $X_i = 0$ with probability $q = 1 - p$). Since N_n is the sum of identically distributed random variables, which are assumed independent, if n is large enough, the central limit theorem says that N_n can be approximated by a normal distribution. We generally consider this approximation to be reasonably accurate when both $np \ge 5$ and $nq \ge 5$. In this case, the normal distribution approximation has mean and standard deviation

$$\mu = np \tag{1.155a}$$

$$\sigma = \sqrt{npq} \tag{1.155b}$$

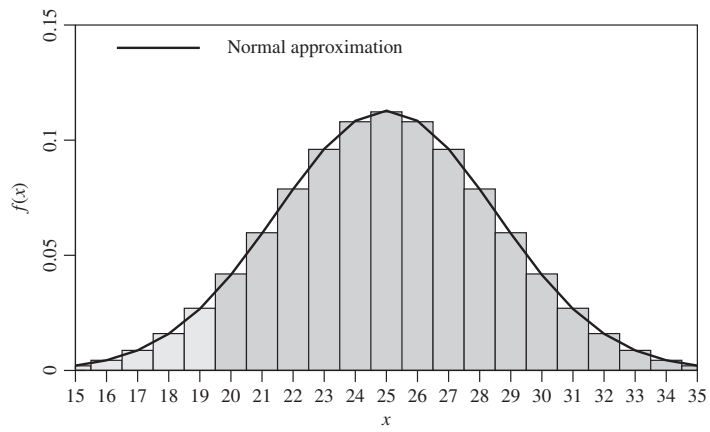


Figure 1.33 Normal approximation to binomial distribution.

Of course, we know that N_n is discrete while the normal distribution governs a continuous random variable. When we want to find the approximate probability that N_n is greater than or equal to, say, k, using the normal distribution, we should include all of the binomial *mass* at k. This means that we should look at the normal probability that $(N_n > k - \frac{1}{2})$. For example, in Figure 1.33, the probability that $N_n \geq 20$ is better captured by the area under the normal distribution above 19.5.

In general, the following corrections apply. Similar corrections apply for two-sided probability calculations.

$$\mathrm{P}[N_n \geq k] \simeq 1 - \Phi\left(\frac{k - 0.5 - \mu}{\sigma}\right) \tag{1.156a}$$

$$\mathrm{P}[N_n > k] \simeq 1 - \Phi\left(\frac{k + 0.5 - \mu}{\sigma}\right) \tag{1.156b}$$

$$\mathrm{P}[N_n \leq k] \simeq \Phi\left(\frac{k + 0.5 - \mu}{\sigma}\right) \tag{1.156c}$$

$$\mathrm{P}[N_n < k] \simeq \Phi\left(\frac{k - 0.5 - \mu}{\sigma}\right) \tag{1.156d}$$

Example 1.56 Suppose that in a certain region it is equally likely for a soil sample to pass a particular soil test as it is to fail it. If this is true, what is the probability that more than 495 samples pass the test over the next 900 tests?

SOLUTION If we assume that soil tests pass or fail independently with constant probability of passing the test, then the number of tests passing, out of n tests, is N_n, which follows a binomial distribution. The exact probability is then given by

$$\begin{aligned}\mathrm{P}[N_{900} > 495] &= \mathrm{P}[N_{900} = 496] + \mathrm{P}[N_{900} = 497] \\ &\quad + \cdots + \mathrm{P}[N_{900} = 900] \\ &= \binom{900}{496} p^{496} q^{404} + \binom{900}{497} p^{497} q^{403} \\ &\quad + \cdots + \binom{900}{900} p^{900} q^{0}\end{aligned}$$

It is not practical to compute this with a simple hand calculator, and even with a computer the calculations are prone to numerical roundoff and overflow errors. The normal approximation will give a very accurate result with a fraction of the effort. We start by computing the mean and variance of N_{900}:

$$\begin{aligned}\mathrm{E}[N_{900}] &= \mu = np = (900)(0.5) = 450 \\ \mathrm{Var}[N_{900}] &= \sigma^2 = npq = (900)(0.5)(0.5) = 225\end{aligned}$$

so that $\sigma = \sqrt{225} = 15$. We now make the following approximation:

$$\begin{aligned}\mathrm{P}[N_{900} > 495] &= \mathrm{P}\left[\frac{N_{900} - \mu}{\sigma} > \frac{495 - \mu}{\sigma}\right] \\ &\simeq \mathrm{P}\left[Z > \frac{495 + 0.5 - 450}{15}\right] \\ &= 1 - \Phi(3.03) \\ &= 0.00122\end{aligned}$$

where, in the second line of the equation, we say that $(N_{900} - \mu)/\sigma$ is approximately a standard normal, and, at the same time, apply the half-interval correction for increased accuracy. (Note that without the half-interval correction we would get $\mathrm{P}[N_{900} > 495] \simeq 0.00135$, a small absolute difference but a 10% relative difference.)

1.10.8.3 Multivariate Normal Distribution The normal distribution is also popular as a distribution governing multiple random variables because it is simply defined knowing only the mean and variance of each random variable and the covariances acting between them. Consider two random variables, X and Y; these follow a bivariate normal distribution if their joint distribution has the form

$$\begin{aligned}f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right.\right. \\ \left.\left. -2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}\end{aligned} \tag{1.157}$$

for $-\infty < x, y < \infty$, where ρ is the correlation coefficient between X and Y and μ_X, μ_Y and σ_X, σ_Y are the means and standard deviations of X and Y, respectively. Figures 1.16 and 1.17 illustrate the bivariate normal distribution.

If X and Y follow a bivariate normal distribution, then their marginal probability density functions, defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy \tag{1.158a}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx \tag{1.158b}$$

are also normal distributions. For example, the marginal distribution of X is a normal distribution with mean μ_X and standard deviation σ_X, and similarly for the marginal distribution of Y. That is,

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\} \tag{1.159a}$$

$$f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\} \qquad (1.159b)$$

Recall that the conditional probability of A given B is

$$\mathrm{P}[A \mid B] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]}$$

From this, we get the following result for *conditional distributions*:

$$f_{X \mid Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} \qquad (1.160)$$

In particular, if X and Y follow a bivariate normal distribution, then it can be shown that

$$f_{X \mid Y}(x \mid y) = \frac{1}{\sigma_X\sqrt{1-\rho^2}\sqrt{2\pi}} \times \exp\left\{-\frac{1}{2}\left[\frac{x-\mu_X-\rho(y-\mu_Y)\sigma_X/\sigma_Y}{\sigma_X\sqrt{1-\rho^2}}\right]^2\right\} \qquad (1.161)$$

It can be seen from this that the conditional distribution of X for a given $Y = y$ also follows a normal distribution with mean and standard deviation

$$\mu_{X \mid Y} = \mu_X + \frac{\rho(y-\mu_Y)\sigma_X}{\sigma_Y} \qquad (1.162a)$$

$$\sigma_{X \mid Y} = \sigma_X\sqrt{(1-\rho^2)} \qquad (1.162b)$$

Example 1.57 Suppose that the load capacities of two neighboring piles, X and Y, are jointly normally distributed with correlation coefficient $\rho = 0.7$. Based on similar pile capacities in the area, the following statistics have been determined:

$$\mu_X = \mu_Y = 2000,\ \sigma_X = \sigma_Y = 500$$

What is the probability that the load capacity of pile X is less than 1700 if nothing is known about the load capacity of pile Y? Alternatively, if the load capacity of pile Y has been measured to be 1800, what is the probability that X is less than 1700 in light of this information?

SOLUTION If nothing is known about the load capacity of Y, then the probability that X is less than 1700 depends only on the marginal distribution of X. That is,

$$\begin{aligned}\mathrm{P}[X < 1700] &= \mathrm{P}\left[Z < \frac{1700-\mu_X}{\sigma_X}\right] \\ &= \mathrm{P}\left[Z < \frac{1700-2000}{500}\right] \\ &= \Phi(-0.6) \\ &= 0.274\end{aligned}$$

If, however, we know that $Y = 1800$, then we are looking for the probability that pile $X < 1700$ conditioned on the fact that pile $Y = 1800$. The conditional mean of X given $Y = 1800$ is

$$\begin{aligned}\mu_{X \mid Y} &= \mu_X + \frac{\rho(y-\mu_Y)\sigma_X}{\sigma_Y} \\ &= 2000 + \frac{0.7(1800-2000)(500)}{500} \\ &= 1860\end{aligned}$$

This is saying, as expected, that the conditional mean of pile X is substantially reduced as a result of the fact that the neighboring pile had a relatively low load capacity. The conditional standard deviation of X given $Y = 1800$ is

$$\begin{aligned}\sigma_{X \mid Y} &= \sigma_X\sqrt{1-\rho^2} \\ &= 500\sqrt{1-0.7^2} \\ &= 357.07\end{aligned}$$

This is reduced from the unconditional standard deviation of 500 because the relatively high correlation with the neighboring pile constrains the possible values of pile X. For example, if the correlation between pile capacities were 1.0, then we would know that $X = Y$. In this case, once we know Y, we would know X with certainty. That is, when $\rho = 1$, the variance of $X \mid Y$ falls to zero. When $\rho = 0$, X and Y will be uncorrelated, and thus independent, since they are normally distributed, and the observation of Y will then make no difference to the variability (and distribution) of X.

For our question, the desired conditional probability is now

$$\begin{aligned}\mathrm{P}[X < 1700 \mid Y = 1800] &= \Phi\left(\frac{1700-\mu_{X \mid Y}}{\sigma_{X \mid Y}}\right) \\ &= \Phi\left(\frac{1700-1860}{357.07}\right) \\ &= \Phi(-0.45) \\ &= 0.326\end{aligned}$$

As expected, the observation of a low load capacity at a neighboring pile has increased the probability of a low load capacity at the pile of interest.

To extend the multivariate normal distribution to more than two random variables, it is useful to use vector–matrix notation. Define

$$\boldsymbol{\mu} = \begin{Bmatrix}\mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_n\end{Bmatrix} \qquad (1.163)$$

to be the vector of means of the sequence of n random variables $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ and

$$\boldsymbol{C} = \begin{bmatrix} C_{11} & C_{12} & \cdot & \cdot & \cdot & C_{1n} \\ C_{21} & C_{22} & \cdot & \cdot & \cdot & C_{2n} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ C_{n1} & C_{n2} & \cdot & \cdot & \cdot & C_{nn} \end{bmatrix} \tag{1.164}$$

to be the matrix of covariances between X_i and X_j, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$. Each element of the covariance matrix is defined as

$$\begin{aligned} C_{ij} &= \mathrm{Cov}\left[X_i, X_j\right] = \rho_{ij}\sigma_i\sigma_j \qquad \text{if } i \neq j \\ &= \mathrm{Var}\left[X_i\right] = \sigma_i^2 \qquad \text{if } i = j \end{aligned}$$

Note that if the X_i's are uncorrelated, then the covariance matrix is diagonal:

$$\boldsymbol{C} = \begin{bmatrix} \sigma_1^2 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \sigma_2^2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \sigma_n^2 \end{bmatrix}$$

Using these definitions, the joint normal distribution of $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) = &\frac{1}{(2\pi)^{n/2}\sqrt{|\boldsymbol{C}|}} \\ &\times \exp\left\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \end{aligned} \tag{1.165}$$

where $|\boldsymbol{C}|$ is the determinant of $\boldsymbol{C}$ and superscript T means the transpose.

As in the bivariate case, all marginal distributions are also normally distributed:

$$f_{X_i}(x_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right\} \tag{1.166}$$

The conditional distributions may be obtained by partitioning the vector $\mathbf{X}$ into two parts (Vanmarcke, 1984): $\mathbf{X}_a$ and $\mathbf{X}_b$ of size n_a and n_b, where $n_a + n_b = n$, that is,

$$\mathbf{X} = \begin{Bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_{n_a} \\ X_{n_a+1} \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{Bmatrix} = \begin{Bmatrix} \mathbf{X}_a \\ \mathbf{X}_b \end{Bmatrix} \tag{1.167}$$

having mean vectors

$$\boldsymbol{\mu}_a = \begin{Bmatrix} \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ \mu_{n_a} \end{Bmatrix}, \qquad \boldsymbol{\mu}_b = \begin{Bmatrix} \mu_{n_a+1} \\ \cdot \\ \cdot \\ \cdot \\ \mu_n \end{Bmatrix} \tag{1.168}$$

Using this partition, the covariance matrix can be split into four submatrices:

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{aa} & \boldsymbol{C}_{ab} \\ \boldsymbol{C}_{ba} & \boldsymbol{C}_{bb} \end{pmatrix} \tag{1.169}$$

where $\boldsymbol{C}_{ba} = \boldsymbol{C}_{ab}^{\mathrm{T}}$. Using these partitions, the conditional mean of the vector $\mathbf{X}_a$ given the vector $\mathbf{X}_b$ can be obtained from

$$\boldsymbol{\mu}_{a\,|\,b} = \boldsymbol{\mu}_a + \boldsymbol{C}_{ab}\boldsymbol{C}_{bb}^{-1}(\mathbf{X}_b - \boldsymbol{\mu}_b) \tag{1.170}$$

Similarly, the conditional covariance matrix is

$$\boldsymbol{C}_{a\,|\,b} = \boldsymbol{C}_{aa} - \boldsymbol{C}_{ab}\boldsymbol{C}_{bb}^{-1}\boldsymbol{C}_{ab}^{\mathrm{T}} \tag{1.171}$$

With these results, the conditional distribution of $\mathbf{X}_a$ given $\mathbf{X}_b$ is

$$\begin{aligned} f_{\mathbf{X}_a\,|\,\mathbf{X}_b}(\mathbf{x}_a\,|\,\mathbf{x}_b) = &\frac{1}{(2\pi)^{n_a/2}\sqrt{|\boldsymbol{C}_{a\,|\,b}|}} \\ &\times \exp\left\{-\tfrac{1}{2}(\mathbf{x}_a - \boldsymbol{\mu}_{a\,|\,b})^T\boldsymbol{C}_{a\,|\,b}^{-1}(\mathbf{x}_a - \boldsymbol{\mu}_{a\,|\,b})\right\} \end{aligned} \tag{1.172}$$

1.10.9 Lognormal Distribution

From the point of view of modeling material properties and loads in engineering, which are generally nonnegative, the normal distribution suffers from the disadvantage of allowing negative values. For example, if a soil's elastic modulus were to be modeled using a normal distribution, then there would be a nonzero probability of obtaining a negative elastic modulus. Since a negative elastic modulus does not occur in practice, the normal cannot be its true distribution.

As an approximation, the normal is nevertheless often used to represent material properties. The error incurred may be slight when the coefficient of variation v is small. For example, if $v \leq 0.3$, then $\mathrm{P}[X < 0] \leq 0.0004$, which may be fine unless it is at these extremes that failure is initiated. A simple way to avoid such problems is to fit a nonnegative distribution to the population in question, and one such candidate is the *lognormal distribution* (Figure 1.34). The lognormal distribution arises from the normal distribution through a simple, albeit nonlinear, transformation. In particular, if G is a normally distributed random variable, having range $-\infty < g < +\infty$, then $X = \exp\{G\}$ will have range $0 \leq x < \infty$. We say that the resulting random variable X is *lognormally* distributed—note that its natural logarithm is normally distributed.

The random variable X is lognormally distributed if $\ln(X)$ is normally distributed. If this is true, then X has probability density function

$$f(x) = \frac{1}{x\sigma_{\ln X}\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu_{\ln X}}{\sigma_{\ln X}}\right)^2\right\}, \quad 0 \leq x < \infty \tag{1.173}$$

Note that this distribution is strictly nonnegative and so is popular as a distribution of nonnegative engineering properties, such as cohesion, elastic modulus, the tangent of the friction angle, and so on. The two parameters of the distribution,

$$\mu_{\ln X} = \mathrm{E}[\ln\ X], \qquad \sigma^2_{\ln X} = \mathrm{Var}[\ln\ X]$$

are the mean and variance of the underlying normally distributed random variable, $\ln X$.

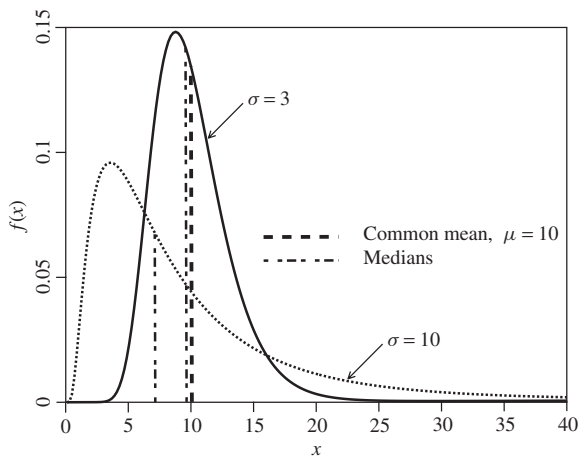


Figure 1.34 Two lognormal distributions illustrating effect of changing variance.

Computing Probabilities In order to compute probabilities from the lognormal distribution, we must make use of the fact that $\ln(X)$ is normally distributed so that we can use the standard normal table. That is, in a probability expression, we take logarithms on both sides of the inequality, then standardize by subtracting the mean and dividing by the standard deviation of $\ln X$,

$$\begin{aligned}\mathrm{P}[X \leq a] &= \mathrm{P}[\ln(X) < \ln(a)] \\ &= \mathrm{P}\left[\frac{\ln(X) - \mu_{\ln X}}{\sigma_{\ln X}} < \frac{\ln(a) - \mu_{\ln X}}{\sigma_{\ln X}}\right] \\ &= \mathrm{P}\left[Z < \frac{\ln(a) - \mu_{\ln X}}{\sigma_{\ln X}}\right] \\ &= \Phi\left(\frac{\ln(a) - \mu_{\ln X}}{\sigma_{\ln X}}\right)\end{aligned} \tag{1.174}$$

where, as before, Z is the standard normal random variate.

Mean and Variance The mean and variance of X are obtained by transforming the two parameters of the lognormal distribution,

$$\mu_X = \mathrm{E}[X] = e^{\mu_{\ln X} + \frac{1}{2}\sigma^2_{\ln X}} \tag{1.175a}$$

$$\sigma^2_X = \mathrm{Var}[X] = \mu^2_X\left(e^{\sigma^2_{\ln X}} - 1\right) \tag{1.175b}$$

Alternatively, if you are given μ_X and σ^2_X, you can obtain the parameters $\mu_{\ln X}$ and $\sigma^2_{\ln X}$ as follows:

$$\sigma^2_{\ln X} = \ln\left(1 + \frac{\sigma^2_X}{\mu^2_X}\right) \tag{1.176a}$$

$$\mu_{\ln X} = \ln(\mu_X) - \tfrac{1}{2}\sigma^2_{\ln X} \tag{1.176b}$$

Characteristics and Moments

$$\mathrm{Mode} = e^{\mu_{\ln X} - \sigma^2_{\ln X}} \tag{1.177a}$$

$$\mathrm{Median} = e^{\mu_{\ln X}} \tag{1.177b}$$

$$\mathrm{Mean} = e^{\mu_{\ln X} + \frac{1}{2}\sigma^2_{\ln X}} \tag{1.177c}$$

$$\mathrm{E}\left[X^k\right] = e^{k\mu_{\ln X} + \frac{1}{2}k^2\sigma^2_{\ln X}} \tag{1.177d}$$

Note that the mode < median < mean, and thus the lognormal distribution has *positive skew*. A distribution is skewed if one of its tails is longer than the other, and, by tradition, the sign of the skew indicates the direction of the longer tail.

Figure 1.35 illustrates the relative locations of the mode, median, and mean for the nonsymmetric lognormal distribution. Because of the positive-skewed, or "skewed-right," shape of the distribution, with the long distribution tail to the right, realizations from the lognormal distribution will

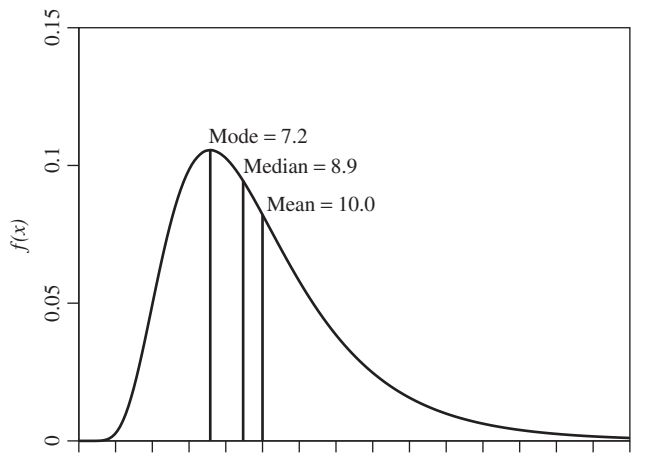


Figure 1.35 Location of mode, median, and mean in lognormal distribution for $\mu_X = 10$ and $\sigma_X = 5$.

have very large values every now and then. This results in the mean being drawn to the right (e.g., the arithmetic average is affected by very large values in the sum). Often, for the lognormal distribution, the median is actually viewed as the primary characteristic of the distribution, since it divides the distribution into equal halves.

It is worth digressing slightly at this point and consider the median of a lognormal distribution in a bit more detail, especially with respect to its estimation. Suppose that we have taken several observations $x_1, x_2, \ldots, x_n$ of a lognormally distributed random variable X. An estimate of the mean of $\ln(X)$ is just the average of $\ln(x_1), \ln(x_2), \ldots, \ln(x_n)$,

$$\hat{\mu}_{\ln X} = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \tag{1.178}$$

where the hat denotes that this is an estimate of $\mu_{\ln X}$. From this, an estimate of the median, $\tilde{x}$, is

$$\tilde{x} = \exp\{\hat{\mu}_{\ln X}\} = \exp\left\{\frac{1}{n}\sum_{i=1}^{n} \ln(x_i)\right\} \tag{1.179}$$

Alternatively, the geometric average x_g of a sequence of nonnegative numbers is defined as the nth root of the product of the n observations,

$$\begin{aligned} x_g &= (x_1 x_2 \cdots x_n)^{1/n} \\ &= \exp\left\{\ln\left((x_1 x_2 \cdots x_n)^{1/n}\right)\right\} \\ &= \exp\left\{\frac{1}{n}\sum_{i=1}^{n} \ln(x_i)\right\} \end{aligned} \tag{1.180}$$

which is identical to the equation for $\tilde{x}$, so we see that the geometric average is an estimate of the median of a lognormally distributed random variable. As we shall see in Section 4.4, this also means that the median of a lognormal distribution is preserved under geometric averaging.

Multiplicative Property If $X = Y_1 Y_2 \cdots Y_n$ and each Y_i are (positive) independent random variables of any distribution having about the same "weight," then

$$\ln X = \ln Y_1 + \ln Y_2 + \cdots + \ln Y_n \tag{1.181}$$

and by the central limit theorem $\ln X$ tends to a normal distribution with

$$\mu_{\ln X} = \mu_{\ln Y_1} + \mu_{\ln Y_2} + \cdots + \mu_{\ln Y_n} \tag{1.182a}$$

$$\sigma^2_{\ln X} = \sigma^2_{\ln Y_1} + \sigma^2_{\ln Y_2} + \cdots + \sigma^2_{\ln Y_n} \tag{1.182b}$$

Thus X tends to a lognormal distribution with parameters $\mu_{\ln X}$ and $\sigma^2_{\ln X}$. This is a useful property since it can be used to approximate the distribution of many multiplicative functions.

In particular, if X is any multiplicative function, say

$$X = \frac{AB}{C} \qquad \Longrightarrow \qquad \ln X = \ln A + \ln B - \ln C \tag{1.183}$$

and A, B, and C are independent and lognormally distributed, then X is also lognormally distributed with

$$\begin{aligned} \mu_{\ln X} &= \mu_{\ln A} + \mu_{\ln B} - \mu_{\ln C} \\ \sigma^2_{\ln X} &= \sigma^2_{\ln A} + \sigma^2_{\ln B} + \sigma^2_{\ln C} \end{aligned}$$

Recall that for variances the coefficient of -1 appearing before the last term in Eq. 1.183 is squared, so that, in the case of independence, the variance of a sum is literally the sum of the variances. (If A, B, and C were correlated, then the covariance terms which would have to be added in to find $\sigma^2_{\ln X}$ would have sign dependent on the signs appearing in the original sum.)

Consider again the geometric average, this time for random observations (i.e., before we have observed them),

$$X_g = (X_1 X_2 \cdots X_n)^{1/n} = X_1^{1/n} \times X_2^{1/n} \times \cdots \times X_n^{1/n}$$

which is a product of n random variables. By the central limit theorem, X_g will tend to a lognormal distribution so that

$$\ln(X_g) = \ln\left((X_1 X_2 \cdots X_n)^{1/n}\right) = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$$

is normally distributed. As mentioned above, X_g is an estimate of the median of X if X is lognormally distributed. However, even if X is not lognormally distributed, X_g will tend to have a lognormal distribution, by the central limit theorem, if the X_i's are nonnegative. We shall see more of this in Chapter 4 where we suggest that in a variety of cases

the lognormal distribution is a natural distribution for soil properties according to the central limit theorem.

The MLEs for $\mu_{\ln X}$ and $\sigma^2_{\ln X}$ are the same as for the normal distribution except that $\ln(X)$ is used in the estimate. If $x_1, x_2, \ldots, x_n$ are n independent observations of a lognormally distributed random variable, then the MLEs are

$$\hat{\mu}_{\ln X} = \frac{1}{n}\sum_{i=1}^{n} \ln\ x_i \tag{1.184a}$$

$$\hat{\sigma}^2_{\ln X} = \frac{1}{n}\sum_{i=1}^{n} (\ln\ x_i - \hat{\mu}_{\ln X})^2 \tag{1.184b}$$

The more common estimator for $\sigma^2_{\ln X}$ is slightly different, having the form

$$\hat{\sigma}^2_{\ln X} = \frac{1}{n-1}\sum_{i=1}^{n} (\ln\ x_i - \hat{\mu}_{\ln X})^2 \tag{1.185}$$

which is an *unbiased* estimator (see Section 5.2.1).

Example 1.58 The settlement δ of a shallow foundation, in meters, can be computed as

$$\delta = c\,\frac{L}{E}$$

where L is the footing load, E is the soil's effective elastic modulus, and c is a constant which accounts for geometry (footing area and aspect ratio, depth to bedrock, etc.) and Poisson's ratio. Assume that c is nonrandom and equal to 0.15 m^{-1} and that the load and elastic modulus are both lognormally distributed with

$$\mu_E = 20,000.0 \text{ kN/m}^2, \qquad \sigma_E = 4000.0 \text{ kN/m}^2$$
$$\mu_L = 1200.0 \text{ kN}, \qquad \sigma_L = 300.0 \text{ kN}$$

What is the probability that the footing settlement exceeds 0.025 m?

SOLUTION First write $\ln(\delta) = \ln(c) + \ln(L) - \ln(E)$, so that

$$\mu_{\ln\delta} = \ln(c) + \mu_{\ln L} - \mu_{\ln E}, \qquad \sigma^2_{\ln\delta} = \sigma^2_{\ln L} + \sigma^2_{\ln E}$$

where we assumed independence between $\ln(L)$ and $\ln(E)$ when computing the variance of $\ln(\delta)$ (so that the covariance terms can be dropped). To compute the above, we must first find the means and variances of $\ln(L)$ and $\ln(E)$:

$$\sigma^2_{\ln L} = \ln\left(1 + \frac{\sigma_L^2}{\mu_L^2}\right) = \ln\left(1 + \frac{300^2}{1200^2}\right) = 0.060625$$
$$\mu_{\ln L} = \ln(\mu_L) - \tfrac{1}{2}\sigma^2_{\ln L} = \ln(1200) - \tfrac{1}{2}(0.060625) = 7.059765$$

$$\sigma^2_{\ln E} = \ln\left(1 + \frac{\sigma_E^2}{\mu_E^2}\right) = \ln\left(1 + \frac{4000^2}{20,000^2}\right) = 0.039221$$
$$\mu_{\ln E} = \ln(\mu_E) - \tfrac{1}{2}\sigma^2_{\ln E} = \ln(20,000) - \tfrac{1}{2}(0.039221) = 9.883877$$

Thus,

$$\mu_{\ln\delta} = \ln(0.15) + 7.059765 - 9.883877 = -4.721232$$
$$\sigma^2_{\ln\delta} = 0.060625 + 0.039221 = 0.099846$$
$$\sigma_{\ln\delta} = \sqrt{0.099846} = 0.315984$$

and

$$\begin{aligned} \text{P}[\delta > 0.025] &= 1 - \text{P}[\delta \le 0.025] \\ &= 1 - \text{P}\left[Z \le \frac{\ln(0.025) - \mu_{\ln\delta}}{\sigma_{\ln\delta}}\right] \\ &= 1 - \text{P}[Z \le 3.27] \\ &= 1 - \Phi(3.27) = 1 - 0.9994622 \\ &= 0.00054 \end{aligned}$$

Most foundations are designed to have probability of failure ranging from 0.001 to 0.0001 against ultimate limit states (e.g., bearing capacity failure). This foundation would be considered very safe with respect to settlement failure, especially since excessive settlement is generally considered to be only a serviceability limit state issue.

1.10.9.1 Bivariate Lognormal Distribution Generally, the multivariate lognormal distribution is handled by directly considering the underlying multivariate normal distribution. That is, rather than considering the joint distribution between the lognormally distributed variates $X_1, X_2, \ldots$, we consider the joint distribution between $\ln X_1$, $\ln X_2, \ldots$ since these are all normally distributed and the results presented in the previous section can be used. However, we sometimes need to consider the lognormally distributed variates directly. Here we will present some results for two lognormally distributed random variables X_1 and X_2.

If X_1 and X_2 are jointly lognormally distributed, then their bivariate distribution is

$$\begin{aligned} f_{X_1X_2}(x,y) = {} & \frac{1}{2\pi\,\sigma_{\ln X_1}\sigma_{\ln X_2} r x y} \\ & \times \exp\left\{-\frac{1}{2r^2}\left[\Psi_1^2 - 2\rho_{\ln 12}\Psi_1\Psi_2 + \Psi_2^2\right]\right\}, \\ & x \ge 0, y \ge 0 \end{aligned} \tag{1.186}$$

where $\Psi_1 = (\ln x - \mu_{\ln X_1})/\sigma_{\ln X_1}$, $\Psi_2 = (\ln y - \mu_{\ln X_2})/\sigma_{\ln X_2}$, $r^2 = 1 - \rho^2_{\ln 12}$, and $\rho_{\ln 12}$ is the correlation coefficient between $\ln X_1$ and $\ln X_2$.

In general, the parameters $\mu_{\ln X_1}$, $\sigma_{\ln X_1}$ can be obtained using the transformation equations given in the previous section from the parameters μ_{X_1}, σ_{X_1}, and so on. If we happen to have an estimate for the correlation coefficient ρ_{12} acting between X_1 and X_2, we can get $\rho_{\ln 12}$ from

$$\rho_{\ln 12} = \frac{\ln(1 + \rho_{12} v_{X_1} v_{X_2})}{\sqrt{\ln(1 + v_{X_1}^2) \ln(1 + v_{X_2}^2)}} \tag{1.187}$$

where $v_{X_i} = \sigma_{X_i}/\mu_{X_i}$ is the coefficient of variation of X_i. We can also invert this relationship to obtain an expression for ρ_{12},

$$\rho_{12} = \frac{\exp\{\rho_{\ln 12}\sigma_{\ln X_1}\sigma_{\ln X_2}\} - 1}{\sqrt{\left(\exp\{\sigma_{\ln X_1}^2\} - 1\right)\left(\exp\{\sigma_{\ln X_2}^2\} - 1\right)}} \tag{1.188}$$

1.10.10 Bounded tanh Distribution

The second half of this book is devoted to a variety of traditional geotechnical problems which are approached in a nontraditional way. In particular, the soil or rock is treated as a spatially variable random field. We shall see in Chapter 3 that a random field with a multivariate normal distribution has the advantage of being fully specified by only its mean and covariance structure. In addition, the simulation of random fields is relatively straightforward when the random field is normally distributed and more complicated when it is not.

Unfortunately, the normal distribution is not appropriate for many soil and rock properties. In particular, most material properties are strictly nonnegative (e.g., elastic modulus). Since the normal distribution has range $(-\infty, +\infty)$, it will always admit some possibility of negative values. When one is simulating possibly millions of realizations of a soil or rock property using the normal distribution, some realizations will inevitably involve negative soil/rock properties, unless the coefficient of variation is quite small and chance is on your side. The occurrence of negative properties often leads to fundamental modeling difficulties (e.g., what happens when Poisson's ratio or the elastic modulus becomes negative?).

In cases where the normal distribution is not appropriate, there are usually two options: (1) choose a distribution on the interval $(0, +\infty)$ (e.g., the lognormal distribution) or (2) choose a distribution which is bounded both above and below on some interval (a, b). The latter would be appropriate for properties such as friction angle, Poisson's ratio, and void ratio.

As we saw above, the lognormal transformation $X = e^G$, where G is normally distributed, leads to a random variable X which takes values on the interval $(0, +\infty)$. Thus, the lognormal distribution derives from a simple transformation of a normally distributed random variable or field. In the case of a bounded distribution, using the transformation

$$X = a + \tfrac{1}{2}(b - a)\left[1 + \tanh\left(\frac{m + sG}{2\pi}\right)\right] \tag{1.189}$$

leads to the random variable X being bounded on the interval (a, b) if G is a standard normally distributed random variable (or at least bounded distribution—we shall assume that G is a standard normal here). The parameter m is a location parameter. If $m = 0$, then the distribution of X is symmetric about the midpoint of the interval, $\frac{1}{2}(a + b)$. The parameter s is a scale parameter—the larger s is, the more variable X is. The function tanh is defined as

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{1.190}$$

In essence, Eq. 1.189 can be used to produce a random variable with a distribution bounded on the interval (a, b), which is a simple transformation of a normally distributed random variable. Thus, a bounded property is easily simulated by first simulating the normally distributed random variable G and then applying Eq. 1.189. Such a simulation would require that the mean and covariance structure of the simulated normally distributed random process be known. To this end, Eq. 1.189 can be inverted to yield

$$m + sG = \pi \ln\left(\frac{X - a}{b - X}\right) \tag{1.191}$$

Since G is a standard normal (having mean zero and unit variance), the parameters m and s are now seen as the mean and standard deviation of the normally distributed random process $(m + sG)$. These two parameters can be estimated by observing a sequence of realizations of X, that is, $x_1, x_2, \ldots, x_n$, transforming each according to

$$y_i = \pi \ln\left(\frac{x_i - a}{b - x_i}\right) \tag{1.192}$$

and then estimating the mean m and standard deviation s using the traditional estimators,

$$m = \frac{1}{n}\sum y_i \tag{1.193a}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - m)^2} \tag{1.193b}$$

In order to estimate the correlation structure, the spatial location, $\mathbf{x}$, of each observation must also be known, so that our observations become $x(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$, and y_i also becomes a function of $\mathbf{x}_i$. The methods of estimating the correlation function discussed in Sections 5.3.6 and 5.4.1.1 can then be applied to the transformed observations, $y(\mathbf{x}_i)$.

The probability density function of X is

$$f_X(x) = \frac{\sqrt{\pi}(b-a)}{\sqrt{2}s(x-a)(b-x)} \times \exp\left\{-\frac{1}{2s^2}\left[\pi \ln\left(\frac{x-a}{b-x}\right) - m\right]^2\right\} \quad (1.194)$$

If $m = 0$, then the mean of X is at the midpoint, $\mu_X = \frac{1}{2}(a+b)$. Since most bounded distributions are symmetric about their midpoints, the remainder of this discussion will be for $m = 0$.

Figure 1.36 illustrates how the distribution of X changes as s changes for $m = 0$, $a = 0$, and $b = 1$. The distribution shapes are identical for different choices in a and b, the only change being that the horizontal axis scales with $b - a$ and the vertical axis scales with $1/(b-a)$. For example, if $a = 10$ and $b = 30$, the $s = 2$ curve looks identical to that shown in Figure 1.36 except that the horizontal axis runs from 10 to 30 while the vertical axis runs from 0 to 0.3. When $s > 5$, the distribution becomes U shaped, which is not a realistic material property shape. Practically speaking, values ranging from $s = 0$, which is nonrandom and equal to the mean, to $s = 5$, which is almost uniformly distributed between a and b, are reasonable.

The relationship between the parameter s and the standard deviation σ_X of X is also of interest. In the limit as $s \to \infty$, the transformation given by Eq. 1.189 becomes a Bernoulli distribution with $p = 0.5$ and X taking possible values a or b. The standard deviation of X for $s \to \infty$ must therefore be $0.5(b-a)$. At the other extreme, as $s \to 0$, we end up with $X = \frac{1}{2}(a+b)$, which is nonrandom. Thus, when $s \to 0$ the standard deviation of X is zero and when $s \to \infty$ the standard deviation of X is $0.5(b-a)$. We suggest, therefore, that σ_X increases from zero when $s = 0$ to $0.5(b-a)$ when $s \to \infty$.

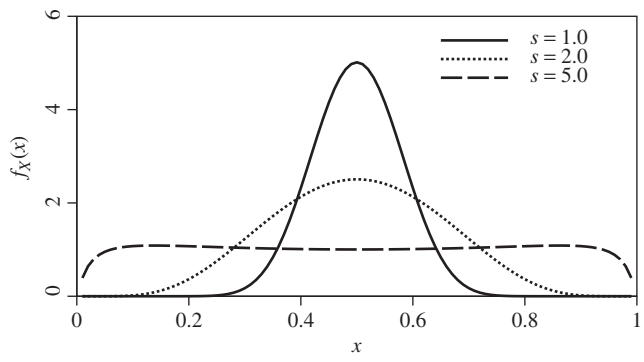


Figure 1.36 Probability density function of X determined as bounded transformation (Eq. 1.189) of normally distributed random variable $(m + sG)$ for $m = 0$ and various values of s.

The following relationship between s and the variance of X derives from a third-order Taylor series approximation to tanh and a first-order approximation to the expectation:

$$\sigma_X^2 = (0.5)^2(b-a)^2 \, \mathrm{E}\left[\tanh^2\left(\frac{sG}{2\pi}\right)\right] \simeq (0.5)^2(b-a)^2 \, \mathrm{E}\left[\frac{\left[sG/(2\pi)\right]^2}{1+\left[sG/(2\pi)\right]^2}\right] \simeq (0.5)^2(b-a)^2\frac{s^2}{4\pi^2+s^2} \quad (1.195)$$

where $\mathrm{E}\left[G^2\right] = 1$ since G is a standard normal random variable. Equation 1.195 slightly overestimates the true standard deviation of X by 0% when $s = 0$ to 11% when $s = 5$. A much closer approximation over the entire range $0 \le s \le 5$ is obtained by slightly decreasing the 0.5 factor to 0.46 (this is an empirical adjustment),

$$\sigma_X \simeq \frac{0.46(b-a)s}{\sqrt{4\pi^2+s^2}} \quad (1.196)$$

The close agreement between Eq. 1.196 and a simulation-based estimate is illustrated in Figure 1.37.

Equation 1.195 can be generalized to yield an approximation to the covariance between two random variables X_i and X_j, each derived as tanh transformations of two standard normal variables G_i and G_j according to Eq. 1.189. If G_i and G_j are correlated, with correlation coefficient ρ_{ij}, then

$$\mathrm{Cov}\left[X_i, X_j\right] = (0.5)^2(b-a)^2 \times \mathrm{E}\left[\tanh\left(\frac{sG_i}{2\pi}\right)\ \tanh\left(\frac{sG_j}{2\pi}\right)\right]$$

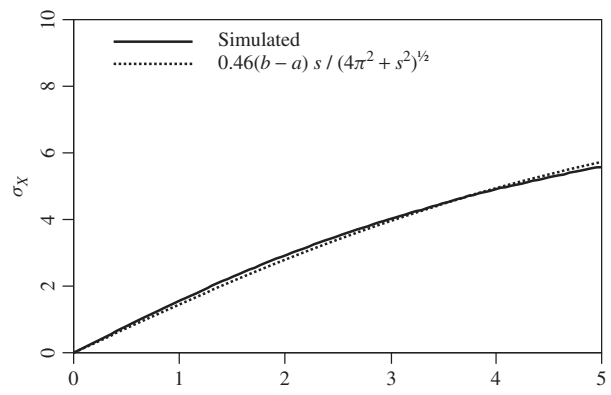


Figure 1.37 Relationship between σ_X and s derived from simulation (100,000 realizations for each s) and Taylor's series derived approximation given by Eq. 1.196. The vertical scale corresponds to $b - a = 20^\circ$.

$$\simeq (0.5)^2(b-a)^2 \times \mathrm{E}\left[\frac{\left[sG_i/(2\pi)\right]\left[sG_j/(2\pi)\right]}{1+(1/2)\left\{\left[sG_i/(2\pi)\right]^2+\left[sG_j/(2\pi)\right]^2\right\}}\right]$$

$$\simeq (0.46)^2(b-a)^2\frac{s^2\rho_{ij}}{4\pi^2+s^2}$$

where the empirical correction given in Eq. 1.196 was introduced in the last step.

1.11 EXTREME-VALUE DISTRIBUTIONS

Most engineering systems fail only when extreme loads occur and failure tends to initiate at the weakest point. Thus, it is of considerable interest to investigate the distribution of extreme values. Consider a sequence of n random variables $X_1, X_2, \ldots, X_n$. This could, for example, be the sequence of tensile strengths of individual links in a chain, or the sequence of daily average soil moisture levels, or earthquake intensities, and so on. Now define the extremes of this set of random variables as

$$Y_n = \max(X_1, X_2, \ldots, X_n) \tag{1.197a}$$

$$Y_1 = \min(X_1, X_2, \ldots, X_n) \tag{1.197b}$$

so that if X_i is the daily average soil moisture level, then Y_n is the maximum daily average soil moisture level over n days. Similarly, if X_i is the tensile strength of the ith link in a chain, then Y_1 is the tensile strength of a chain composed of n links.

1.11.1 Exact Extreme-Value Distributions

Let us first examine the behavior of the maximum, Y_n. We know that if the maximum is less than some number y, then each X_i must also be less than y. That is, the event $(Y_n \le y)$ must be equivalent to the event $(X_1 \le y \cap X_2 \le y \cap \cdots \cap X_n \le y)$. In other words the *exact* distribution of Y_n is

$$\mathrm{P}\left[Y_n \le y\right] = \mathrm{P}\left[X_1 \le y \cap X_2 \le y \cap \cdots \cap X_n \le y\right] \tag{1.198}$$

If it can be further assumed that the X's are *independent and identically distributed* (iid) (if this is not the case, the problem becomes very complex and usually only solved via simulation), then

$$\begin{aligned} F_{Y_n}(y) &= \mathrm{P}\left[Y_n \le y\right] \\ &= \mathrm{P}\left[X_1 \le y\right]\mathrm{P}\left[X_2 \le y\right]\cdots\mathrm{P}\left[X_n \le y\right] \\ &= \left[F_X(y)\right]^n \end{aligned} \tag{1.199}$$

where F_X is the cumulative distribution function of X. Taking the derivative gives us the probability density function

$$\begin{aligned} f_{Y_n}(y) &= \frac{dF_{Y_n}(y)}{dy} = n\left[F_X(y)\right]^{n-1}\frac{dF_X(y)}{dy} \\ &= n\left[F_X(y)\right]^{n-1} f_X(y) \end{aligned} \tag{1.200}$$

Example 1.59 Suppose that fissure lengths X in a rock mass have an exponential distribution with $f_X(x) = e^{-x}$. What, then, does the distribution of the maximum fissure length Y_n look like for $n = 1, 5, 50$ fissures?

SOLUTION If $n = 1$, then Y_n is the maximum of one observed fissure, which of course is just the distribution of the single fissure length. Thus, when $n = 1$, the distribution of Y_n is just the exponential distribution

$$f_{Y_1}(y) = f_X(y) = e^{-y}$$

When $n = 5$, we have

$$\begin{aligned} F_{Y_5}(y) &= \mathrm{P}\left[Y_5 \le y\right] \\ &= \mathrm{P}\left[X_1 \le y\right]\mathrm{P}\left[X_2 \le y\right]\cdots\mathrm{P}\left[X_5 \le y\right] \\ &= \left[F_X(y)\right]^5 \\ &= \left[1-e^{-y}\right]^5 \end{aligned}$$

where we used the fact that $F_X(x) = 1 - e^{-x}$. To find the probability density function (which is usually more informative graphically), we must differentiate:

$$f_{Y_5}(y) = \frac{dF_{Y_5}(y)}{dy} = 5e^{-y}\left[1-e^{-y}\right]^4$$

Similarly, when $n = 50$, we have

$$\begin{aligned} F_{Y_{50}}(y) &= \mathrm{P}\left[Y_{50} \le y\right] \\ &= \mathrm{P}\left[X_1 \le y\right]\mathrm{P}\left[X_2 \le y\right]\cdots\mathrm{P}\left[X_{50} \le y\right] \\ &= \left[F_X(y)\right]^{50} \\ &= \left[1-e^{-y}\right]^{50} \end{aligned}$$

and

$$f_{Y_{50}}(y) = \frac{dF_{Y_{50}}(y)}{dy} = 50e^{-y}\left[1-e^{-y}\right]^{49}$$

Plots of these three distributions appear as in Figure 1.38.

Example 1.60 Suppose that X follows an exponential distribution with

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \ge 0$$

Then what is the probability that the largest from a sample of five observations of X will exceed 3 times the mean?

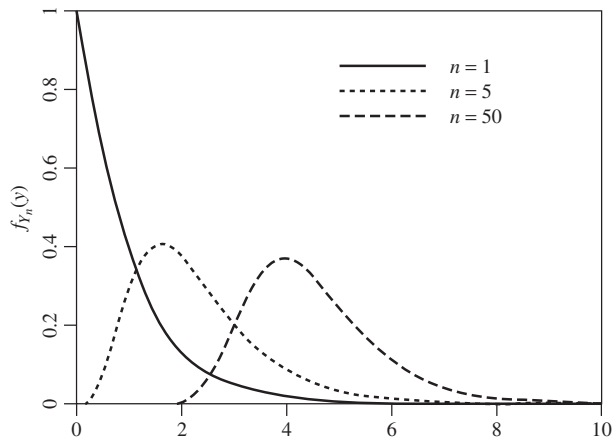


Figure 1.38 Distributions of maximum value of n observations of exponentially distributed random variable.

SOLUTION For $n = 5$, we have

$$\begin{aligned} F_{Y_5}(y) &= \mathrm{P}\left[Y_5 \le y\right] \\ &= \mathrm{P}\left[X_1 \le y\right]\mathrm{P}\left[X_2 \le y\right]\cdots\mathrm{P}\left[X_5 \le y\right] \\ &= \left[F_X(y)\right]^5 \\ &= \left[1 - e^{-\lambda y}\right]^5 \end{aligned}$$

so that

$$\mathrm{P}\left[Y_5 > y\right] = 1 - \left[1 - e^{-\lambda y}\right]^5$$

The mean of X is $1/\lambda$ (see Eq. 1.120), so the probability that Y_5 exceeds 3 times the mean is

$$\begin{aligned} \mathrm{P}\left[Y_5 > \frac{3}{\lambda}\right] &= 1 - \left[1 - e^{-\lambda(3/\lambda)}\right]^5 \\ &= 1 - \left[1 - e^{-3}\right]^5 \\ &= 0.14205 \end{aligned}$$

Now consider the distribution of the minimum out of n samples, Y_1. If we proceed as we did for Y_n, then we would look at the event $Y_1 \le y$. This event just means that $X_1 \le y$ or $X_2 \le y$ or $\ldots$, that is,

$$\mathrm{P}\left[Y_1 \le y\right] = \mathrm{P}\left[X_1 \le y \cup X_2 \le y \cup \cdots \cup X_n \le y\right] \tag{1.201}$$

The union on the right expands into $\binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n}$ terms—in other words potentially a *lot* of terms. A better way to work out this distribution is to look at the complement:

$$\mathrm{P}\left[Y_1 > y\right] = \mathrm{P}\left[X_1 > y \cap X_2 > y \cap \cdots \cap X_n > y\right]$$

$$\begin{aligned} &= \mathrm{P}\left[X_1 > y\right]\mathrm{P}\left[X_2 > y\right]\cdots\mathrm{P}\left[X_n > y\right] \\ &= \left[1 - F_X(y)\right]^n \end{aligned} \tag{1.202}$$

and since $\mathrm{P}\left[Y_1 > y\right] = 1 - F_{Y_1}(y)$ we get

$$F_{Y_1}(y) = 1 - \left[1 - F_X(y)\right]^n \tag{1.203}$$

and, taking the derivative,

$$f_{Y_1}(y) = n\left[1 - F_X(y)\right]^{n-1} f_X(y) \tag{1.204}$$

Example 1.61 A series of five soil samples are taken at a site and their shear strengths determined. Suppose that a subsequent design is going to be based on the minimum shear strength observed out of the five samples. If the shear strengths of the individual samples are exponentially distributed with parameter $\lambda = 0.025$ m^2/kN, then what is the distribution of the design shear strength?

SOLUTION If we let Y_1 be the design shear strength, where Y_1 is the minimum shear strength observed from the $n = 5$ samples, then

$$F_{Y_1}(y) = 1 - \left[1 - F_X(y)\right]^5$$

where, for the exponential distribution, $F_X(x) = 1 - e^{-\lambda x}$. Thus,

$$\begin{aligned} F_{Y_1}(y) &= 1 - \left[1 - \left(1 - e^{-\lambda y}\right)\right]^5 \\ &= 1 - e^{-5\lambda y} \end{aligned}$$

From this we see that the extreme-value distribution of the minimum of samples from an exponential distribution is also exponentially distributed with new parameter $\lambda' = n\lambda = 5(0.025) = 0.125$. Notice that while the individual samples have mean shear strength equal to $1/\lambda = 1/0.025 = 40$ kN/m^2, the mean design shear strength is one-fifth this value, $1/\lambda' = 1/0.125 = 8$ kN/m^2.

1.11.2 Asymptotic Extreme-Value Distributions

In cases where the cumulative distribution function $F_X(x)$ is not known explicitly (e.g., the normal or lognormal), the exact distributions given above are of questionable value. It turns out that if n is large enough and the sample is random (i.e., composed of independent observations), then the distribution of an extreme value tends toward one of three "asymptotic" forms, which are explained as follows. Thus, even if you do not know the precise form of the distribution of X, the distribution of the extreme value of $X_1, X_2, \ldots, X_n$ can often be deduced, since there are only three possibilities. The results presented below were developed by Gumbel (1958).

1.11.2.1 Type I Asymptotic Form If X has a distribution with an *unlimited exponentially* decaying tail in the direction of the extreme under consideration, then the distribution of the extreme will tend to the *type I asymptotic form*. Examples of such distributions are the normal (in either direction) and the exponential (in the positive direction).

In the case of the maximum, the type I extreme-value distribution has the form

$$F_{Y_n}(y) = \exp\left\{-e^{-\alpha_n(y-u_n)}\right\} \tag{1.205a}$$

$$f_{Y_n}(y) = \alpha_n\, e^{-\alpha_n(y-u_n)}\ \exp\left\{-e^{-\alpha_n(y-u_n)}\right\} \tag{1.205b}$$

where

$$\begin{aligned} u_n &= \text{characteristic largest value of } X \\ &= F_X^{-1}\left(1-\frac{1}{n}\right) \\ &= \text{mode of } Y_n \end{aligned} \tag{1.206a}$$

$$\begin{aligned} \alpha_n &= \text{inverse measure of variance of } Y_n \\ &= nf_X(u_n) \end{aligned} \tag{1.206b}$$

In particular, u_n is defined as the value that X exceeds with probability $1/n$. It is found by solving $\mathrm{P}[X > u_n] = 1/n$ for u_n, giving the result shown above. If $F_X^{-1}(p)$ is not readily available, you will either have to consult the literature or determine this extreme-value distribution via simulation.

The mean and variance of the type I maximum asymptotic distribution are as follows:

$$\mathrm{E}[Y_n] = u_n + \frac{\gamma}{\alpha_n} \tag{1.207a}$$

$$\mathrm{Var}[Y_n] = \frac{\pi^2}{6\alpha_n^2} \tag{1.207b}$$

where $\gamma = 0.577216\ldots$ is Euler's number.

Example 1.62 Suppose that a structure is supported by $n = 20$ piles and that long-term pile settlements are distributed according to $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ being the settlement, where $\lambda = 0.2\ \mathrm{mm}^{-1}$. If we make the assumption that the piles settle independently (probably a questionable assumption, so that the following results should only be considered approximate), then find the asymptotic parameters of the largest pile settlement, Y_n, out of the n piles, assuming that n is large enough that the asymptotic extreme-value distribution holds.

SOLUTION To find u_n, we solve $\mathrm{P}[X > u_n] = 1/n$ for u_n. For the exponential distribution,

$$\mathrm{P}[X > u_n] = e^{-\lambda u_n} = \frac{1}{n}$$

$$\begin{aligned} -\lambda u_n &= -\ln(n) \\ u_n &= \frac{\ln(n)}{\lambda} = \frac{\ln(20)}{0.2} \\ &= 14.98\ \text{mm} \end{aligned}$$

and

$$\alpha_n = nf_X(u_n) = n\lambda e^{-\lambda \ln(n)/\lambda} = \lambda$$

The parameter $u_n = 14.98$ is the most probable largest settlement out of the 20 piles (e.g., the mode of the distribution).

The asymptotic extreme-value distribution is then

$$F_{Y_n}(y) = \exp\left\{-e^{-\lambda y - \ln(n)}\right\} = \exp\left\{\frac{-e^{-\lambda y}}{n}\right\}$$

The distribution of the minimum value, where the distribution of X is exponentially decaying and unlimited in the direction of the minimum, has the form

$$F_{Y_1}(y) = 1 - \exp\left\{-e^{-\alpha_1(y-u_1)}\right\} \tag{1.208a}$$

$$f_{Y_1}(y) = \alpha_1\, e^{-\alpha_1(y-u_1)} \exp\left\{-e^{-\alpha_1(y-u_1)}\right\} \tag{1.208b}$$

where

$$\begin{aligned} u_1 &= \text{characteristic smallest value of } X \\ &= F_X^{-1}\left(\frac{1}{n}\right) \\ &= \text{mode of } Y_1 \end{aligned} \tag{1.209a}$$

$$\begin{aligned} \alpha_1 &= \text{inverse measure of variance of } Y_1 \\ &= nf_X(u_1) \end{aligned} \tag{1.209b}$$

In particular, u_1 is defined as the value that X has probability $1/n$ of being below. It is found by solving $\mathrm{P}[X \leq u_1] = 1/n$ for u_1. The mean and variance of Y_1 are as follows:

$$\mathrm{E}[Y_1] = u_1 - \frac{\gamma}{\alpha_1} \tag{1.210a}$$

$$\mathrm{Var}[Y_1] = \frac{\pi^2}{6\alpha_1^2} \tag{1.210b}$$

Because of the mirror symmetry of the minimum and maximum type I extreme-value distributions, the skewness coefficient of Y_n is 1.1414 whereas the skewness coefficient of Y_1 is -1.1414. That is, the two distributions are mirror images of one another.

1.11.2.2 Type II Asymptotic Form If X has a distribution with an unlimited polynomial tail, in the direction of the extreme, then its extreme value will have a type II distribution. Examples of distributions with polynomial tails are the lognormal (in the positive direction) and the Pareto

(in the positive direction) distributions, the latter of which has the form

$$F_X(x) = 1 - \left(\frac{b}{x}\right)^{\alpha} \qquad \text{for } x \geq b$$

If the coefficient b is replaced by $u_n/n^{1/\alpha}$, then we get

$$F_X(x) = 1 - \frac{1}{n}\left(\frac{u_n}{x}\right)^{\alpha} \qquad \text{for } x \geq u_n/n^{1/\alpha}$$

The corresponding extreme-value distribution for the maximum, in the limit as $n \to \infty$, is

$$F_{Y_n}(y) = \exp\left\{-\left(\frac{u_n}{y}\right)^{\alpha}\right\} \qquad \text{for } y \geq 0 \qquad (1.211a)$$

$$f_{Y_n}(y) = \left(\frac{\alpha}{u_n}\right)\left(\frac{u_n}{y}\right)^{\alpha+1} \exp\left\{-\left(\frac{u_n}{y}\right)^{\alpha}\right\} \qquad (1.211b)$$

$$\begin{aligned} u_n &= \text{characteristic largest value of } X \\ &= F_X^{-1}\left(1 - \frac{1}{n}\right) \\ &= \text{mode of } Y_n \end{aligned} \qquad (1.212a)$$

$$\begin{aligned} \alpha &= \text{shape parameter} \\ &= \text{order of polynomial decay of } F_X(x) \\ &\quad \text{in direction of extreme} \end{aligned} \qquad (1.212b)$$

Note that although the lognormal distribution seems to have an exponentially decaying tail in the direction of the maximum, the distribution is actually a function of the form $a\exp\{-b(\ln x)^2\}$, which has a polynomial decay. Thus, the extreme-value distribution of n lognormally distributed random variables follows a type II distribution with

$$\alpha = \frac{\sqrt{2\ \ln n}}{\sigma_{\ln X}}$$

$$u_n = \exp\{u_n'\}$$

$$u_n' = \sigma_{\ln X}\sqrt{2\ln n} - \frac{\sigma_{\ln X}\left[\ln(\ln n) + \ln(4\pi)\right]}{2\sqrt{2\ln n}} + \mu_{\ln X}$$

The mean and variance of the type II maximum asymptotic distribution are as follows:

$$\mathrm{E}[Y_n] = u_n\Gamma\left(1 - \frac{1}{\alpha}\right) \quad \text{if } \alpha > 1 \qquad (1.213a)$$

$$\mathrm{Var}[Y_n] = u_n^2\Gamma\left(1 - \frac{2}{\alpha}\right) - \mathrm{E}^2[Y_n] \quad \text{if } \alpha > 2 \qquad (1.213b)$$

where Γ is the gamma function (see Eq. 1.128).

The distribution of the minimum for an unbounded polynomial decaying tail can be found as the negative "reflection" of the maximum, namely as

$$F_{Y_1}(y) = 1 - \exp\left\{-\left(\frac{u_1}{y}\right)^{\alpha}\right\}, \quad y \leq 0, \quad u_1 < 0 \qquad (1.214a)$$

$$f_{Y_1}(y) = -\left(\frac{\alpha}{u_1}\right)\left(\frac{u_1}{y}\right)^{\alpha+1} \exp\left\{-\left(\frac{u_1}{y}\right)^{\alpha}\right\} \qquad (1.214b)$$

where

$$\begin{aligned} u_1 &= \text{characteristic smallest value of } X \\ &= F_X^{-1}\left(\frac{1}{n}\right) \\ &= \text{mode of } Y_1 \end{aligned} \qquad (1.215a)$$

$$\begin{aligned} \alpha &= \text{shape parameter} \\ &= \text{order of polynomial decay of } F_X(x) \\ &\quad \text{in direction of extreme} \end{aligned} \qquad (1.215b)$$

The mean and variance of the type II minimum asymptotic distribution are as follows:

$$\mathrm{E}[Y_1] = u_1\Gamma\left(1 - \frac{1}{\alpha}\right) \quad \text{if } \alpha > 1 \qquad (1.216a)$$

$$\mathrm{Var}[Y_1] = u_1^2\Gamma\left(1 - \frac{2}{\alpha}\right) - \mathrm{E}^2[Y_1] \quad \text{if } \alpha > 2 \qquad (1.216b)$$

Example 1.63 Suppose that the pile settlements, X, discussed in the last example actually have the distribution

$$f_X(x) = \frac{1}{x^2} \quad \text{for } x \geq 1 \text{ mm}$$

Determine the exact distribution of the maximum of a random sample of size n and the asymptotic distribution of the maximum.

SOLUTION We first need to find the cumulative distribution function of X,

$$F_X(x) = \int_1^x \frac{1}{t^2}dt = 1 - \frac{1}{x}, \qquad x \geq 1$$

The exact cumulative distribution function of the maximum pile settlement, Y_n, is thus

$$F_{Y_n}(y) = \left[F_X(y)\right]^n = \left[1 - \frac{1}{y}\right]^n \quad \text{for } y \geq 1$$

and the exact probability density function of Y_n is the derivative of $F_{Y_n}(y)$,

$$f_{Y_n}(y) = \frac{n}{y^2}\left[1 - \frac{1}{y}\right]^{n-1} \quad \text{for } y \geq 1$$

For the asymptotic distribution, we need to find u_n such that $F_X(u_n) = 1 - 1/n$,

$$F_X(u_n) = 1 - \frac{1}{u_n} = 1 - \frac{1}{n}$$

so that $u_n = n$. The order of polynomial decay of $F_X(x)$ in the direction of the extreme (positive direction) is $\alpha = 1$, so that the asymptotic extreme-value distribution of the maximum, Y_n, is

$$F_{Y_n}(y) = \exp\left\{-\frac{n}{y}\right\} \qquad \text{for } y \geq 0$$

$$f_{Y_n}(y) = \frac{n}{y^2}\exp\left\{-\frac{n}{y}\right\} \quad \text{for } y \geq 0$$

We see immediately that one result of the approximation is that the lower bound of the asymptotic approximations is $y \geq 0$, rather than $y \geq 1$ found in the exact distributions. However, for $n = 10$, Figure 1.39 compares the exact and asymptotic distributions, and they are seen to be very similar.

1.11.2.3 Type III Asymptotic Form If the distribution of X is bounded by a value, u, in the direction of the extreme, then the asymptotic extreme-value distribution (as $n \to \infty$) is the type III form. Examples are the lognormal and exponential distributions toward the left and the beta and uniform distributions in either direction. For the maximum, the type III asymptotic form is

$$F_{Y_n}(y) = \exp\left\{-\left(\frac{u-y}{u-u_n}\right)^{\alpha}\right\} \quad \text{for } y \leq u \tag{1.217a}$$

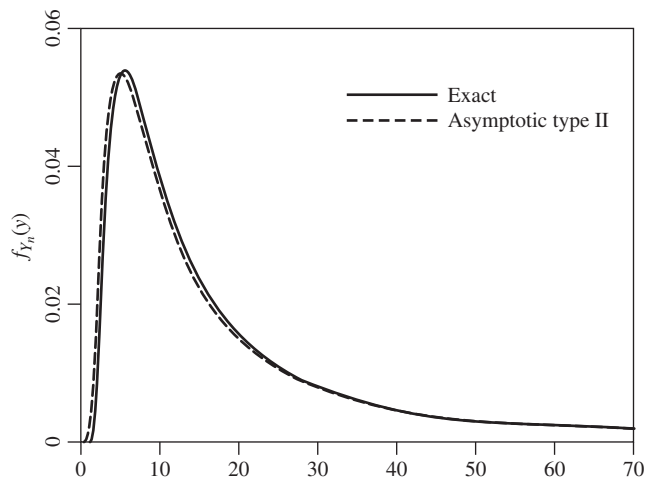


Figure 1.39 Comparison of exact and asymptotic (type II) extreme-value distributions for $n = 10$.

$$f_{Y_n}(y) = \frac{\alpha(u-y)^{\alpha-1}}{(u-u_n)^{\alpha}}\exp\left\{-\left(\frac{u-y}{u-u_n}\right)^{\alpha}\right\} \quad \text{for } y \leq u \tag{1.217b}$$

where

$$\begin{aligned} u_n &= \text{characteristic largest value of } X \\ &= F_X^{-1}\left(1 - \frac{1}{n}\right) \\ &= \text{mode of } Y_n \end{aligned} \tag{1.218a}$$

$$\begin{aligned} \alpha &= \text{shape parameter} \\ &= \text{order of polynomial decay of } F_X(x) \\ &\quad \text{in direction of extreme} \end{aligned} \tag{1.218b}$$

The mean and variance of the type III maximum asymptotic distribution are as follows:

$$\mathrm{E}[Y_n] = u - (u - u_n)\Gamma\left(1 + \frac{1}{\alpha}\right) \tag{1.219a}$$

$$\mathrm{Var}[Y_n] = (u-u_n)^2 \times \left[\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right] \tag{1.219b}$$

In the case of the minimum, the asymptotic extreme-value distribution is

$$F_{Y_1}(y) = 1 - \exp\left\{-\left(\frac{y-u}{u_1-u}\right)^{\alpha}\right\} \quad \text{for } y \geq u \tag{1.220a}$$

$$f_{Y_1}(y) = \frac{\alpha(y-u)^{\alpha-1}}{(u_1-u)^{\alpha}}\exp\left\{-\left(\frac{y-u}{u_1-u}\right)^{\alpha}\right\} \tag{1.220b}$$

where

$$\begin{aligned} u_1 &= \text{characteristic smallest value of } X \\ &= F_X^{-1}\left(\frac{1}{n}\right) \\ &= \text{mode of } Y_1 \end{aligned} \tag{1.221a}$$

$$\begin{aligned} \alpha &= \text{shape parameter} \\ &= \text{order of polynomial decay of } F_X(x) \\ &\quad \text{in direction of extreme} \end{aligned} \tag{1.221b}$$

and u is the minimum bound on X. This distribution is also a form of the Weibull distribution. The shape parameter α is, as mentioned, the order of the polynomial $F_X(x)$ in the direction of the extreme. For example, if X is exponentially distributed and we are looking at the distribution of the minimum, then $F_X(x)$ has Taylor's series expansion for small x of

$$F_X(x) = 1 - e^{-\lambda x} \simeq 1 - (1 - \lambda x) = \lambda x \tag{1.222}$$

which has order 1 as $x \to 0$. Thus, for the minimum of an exponential distribution, $\alpha = 1$.

The mean and variance of the type III maximum asymptotic distribution are as follows:

$$\mathrm{E}[Y_n] = u + (u_1 - u)\Gamma\left(1 + \frac{1}{\alpha}\right) \tag{1.223a}$$

$$\mathrm{Var}[Y_n] = (u_1 - u)^2 \times \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right] \tag{1.223b}$$

Example 1.64 A series of 50 soil samples are taken at a site and their shear strengths determined. Suppose that a subsequent design is going to be based on the minimum shear strength observed out of the 50 samples. If the shear strengths of the individual samples are exponentially distributed with parameter $\lambda = 0.025$ m^2/kN, then what is the asymptotic distribution of the design shear strength (i.e., their minimum)? Assume that n is large enough that the asymptotic extreme-value distributions hold.

SOLUTION If we let Y_1 be the design shear strength, then Y_1 is the minimum shear strength observed among the $n = 50$ samples. Since the shear strengths are exponentially distributed, they are bounded by $u = 0$ in the direction of the minimum (to the left). This means that the asymptotic extreme-value distribution of Y_1 is type III. For this distribution, we first need to find u_1 such that $F_X(u_1) = 1/n$,

$$F_X(u_1) = 1 - e^{-\lambda u_1} = 1/n$$
$$\Longrightarrow \qquad u_1 = -(1/\lambda)\ \ln(1 - 1/n)$$

so that $u_1 = -\ln(0.98)/0.025 = 0.8081$.

The order of polynomial decay of $F_X(x)$ in the direction of the extreme (toward $X = 0$) is $\alpha = 1$, as determined by Eq. 1.222, so that the asymptotic extreme-value distribution of the minimum, Y_1, is

$$F_{Y_1}(y) = 1 - \exp\left\{-\left(\frac{y}{0.8081}\right)\right\}, \quad \text{for } y \geq 0$$

$$f_{Y_1}(y) = \frac{1}{0.8081}\exp\left\{-\left(\frac{y}{0.8081}\right)\right\} \quad \text{for } y \geq 0$$

which is just an exponential distribution with parameter $\lambda' = 1/0.8081 = 1.237$. Note that the exact distribution of the minimum is exponential with parameter $\lambda' = n\lambda = 50(0.025) = 1.25$, so the asymptotic approximation is reasonably close to the exact. Figure 1.40 illustrates the close agreement between the two distributions.

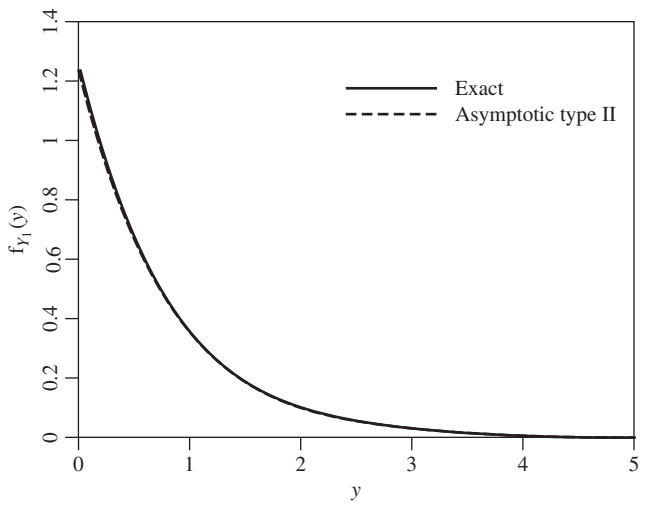


Figure 1.40 Comparison of exact and asymptotic (type III) extreme-value distributions for $n = 50$.

1.12 SUMMARY

De Morgan	$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c$
Probability	$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B] - \mathrm{P}[A \cap B]$
	$\mathrm{P}[A \cap B] = \mathrm{P}[A\|B] \cdot \mathrm{P}[B] = \mathrm{P}[B\|A] \cdot \mathrm{P}[A]$
Bayes' theorem	$\mathrm{P}[A_j \mid E] = \dfrac{\mathrm{P}[E\|A_j] \cdot \mathrm{P}[A_j]}{\mathrm{P}[E]} = \dfrac{\mathrm{P}[E\|A_j] \cdot \mathrm{P}[A_j]}{\sum_{i=1}^{n} \mathrm{P}[E\|A_i] \cdot \mathrm{P}[A_i]}$
PDFs and CDFs	$F(x) = \int_{-\infty}^{x} f(\xi)\, d\xi \Longleftrightarrow f(x) = \dfrac{d}{dx} F(x)$

Expectations

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X \, dx, \qquad \mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X \, dx$$

$$\mathrm{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f_X \, dx, \qquad \mathrm{E}[a + bX] = a + b\mathrm{E}[X]$$

$$\mathrm{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y) \, dx \, dy$$

Variance

$$\mathrm{Var}[X] = \mathrm{E}\left[(X - \mu)^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X] = \sigma^2$$
$$\mathrm{Var}[a + bX] = b^2 \, \mathrm{Var}[X]$$

Covariance

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y], \qquad \rho_{XY} = \frac{\mathrm{Cov}[X, Y]}{\sigma_X \, \sigma_Y}$$

Taylor's series

$$Y = g(X) = g(\mu_X) + (X - \mu_X) \left.\frac{dg}{dx}\right|_{\mu_X} + \frac{1}{2!}(X - \mu_X)^2 \left.\frac{d^2 g}{dx^2}\right|_{\mu_X} + \cdots$$

Linear functions

$$\text{If } Y = \sum_{i=1}^{n} a_i X_i \text{ and } Z = \sum_{i=1}^{n} b_i X_i, \text{ then } \mathrm{E}[Y] = \sum_{i=1}^{n} a_i \mathrm{E}[X_i]$$

$$\mathrm{Var}[Y] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{Cov}\left[X_i, X_j\right], \qquad \mathrm{Cov}[Y, Z] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j \, \mathrm{Cov}\left[X_i, X_j\right]$$

Functions

$$\text{If } Y = g(X) \text{ is one to one, then } f_Y(y) = f_X(x) \left|\frac{dx}{dy}\right|$$

Miscellaneous

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right\}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \qquad \Gamma(r) = (r-1)! \quad \text{(r integer)}$$

Binomial

$$\mathrm{P}[N_n = k] = \binom{n}{k} p^k q^{n-k} \qquad \text{for } 0 \le k \le n$$

$$\mathrm{E}[N_n] = np \qquad \mathrm{Var}[N_n] = npq$$

Geometric

$$\mathrm{P}[T_1 = k] = pq^{k-1} \qquad \text{for } k \ge 1$$

$$\mathrm{E}[T_1] = \frac{1}{p} \qquad \mathrm{Var}[T_1] = \frac{q}{p^2}$$

Negative binomial

$$\mathrm{P}[T_k = m] = \binom{m-1}{k-1} p^k \, q^{m-k} \qquad \text{for } m \ge k$$

$$\mathrm{E}[T_k] = \frac{k}{p} \qquad \mathrm{Var}[T_k] = \frac{kq}{p^2}$$

Poisson

$$\mathrm{P}[N_t = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad \text{for } k \ge 0$$

$$\mathrm{E}[N_t] = \lambda t \qquad \mathrm{Var}[N_t] = \lambda t$$

Uniform

$$f(x) = \frac{1}{\beta - \alpha} \qquad F(x) = \frac{x - \alpha}{\beta - \alpha} \qquad \text{for } \alpha \le x \le \beta$$

$$\mathrm{E}[X] = \tfrac{1}{2}(\alpha + \beta) \qquad \mathrm{Var}[X] = \tfrac{1}{12}(\beta - \alpha)^2$$

Exponential

$$f(t) = \lambda e^{-\lambda t} \qquad F(t) = 1 - e^{-\lambda t} \qquad \text{for } t \ge 0$$

$$\mathrm{E}[T] = \frac{1}{\lambda} \qquad \mathrm{Var}[T] = \frac{1}{\lambda^2}$$

Gamma	$f(x) = \frac{\lambda}{(k-1)!}(\lambda x)^{k-1}e^{-\lambda x}$	$F(x) = 1 - e^{-\lambda x}\sum_{j=0}^{k-1}\frac{(\lambda x)^j}{j!}$	k integer
	$\mathrm{E}[X] = \frac{k}{\lambda}$	$\mathrm{Var}[X] = \frac{k}{\lambda^2}$	
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$	$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$	for $-\infty < x < \infty$
	$\mathrm{E}[X] = \mu$	$\mathrm{Var}[X] = \sigma^2$	
	$\mathrm{P}[X \le x] = \mathrm{P}\left[Z \le \frac{x-\mu}{\sigma}\right] = \Phi\left(\frac{x-\mu}{\sigma}\right)$		
Lognormal	$f(x) = \frac{1}{x\sigma_{\ln X}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu_{\ln X}}{\sigma_{\ln X}}\right)^2\right]$	$F(x) = \Phi\left(\frac{\ln\ x - \mu_{\ln X}}{\sigma_{\ln X}}\right)$	for $0 \le x < \infty$
	$\mathrm{E}[X] = \mu_X = e^{\mu_{\ln X} + \frac{1}{2}\sigma^2_{\ln X}}$	$\mathrm{Var}[X] = \sigma_X^2 = \mu_X^2\left(e^{\sigma^2_{\ln X}} - 1\right)$	
	$\sigma^2_{\ln X} = \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)$	$\mu_{\ln X} = \ln(\mu_X) - \frac{1}{2}\sigma^2_{\ln X}$	
Weibull	$f(x) = \frac{\beta}{x}(\lambda x)^\beta e^{-(\lambda x)^\beta}$	$F(x) = 1 - e^{-(\lambda x)^\beta}$	for $x \ge 0$
	$\mathrm{E}[X] = \frac{1}{\lambda\beta}\Gamma\left(\frac{1}{\beta}\right)$	$\mathrm{Var}[X] = \frac{1}{\lambda^2\beta}\left\{2\Gamma\left(\frac{2}{\beta}\right) - \frac{1}{\beta}\left[\Gamma\left(\frac{1}{\beta}\right)\right]^2\right\}$	
Extreme Value Distributions:			
Type I	$F_{Y_n}(y) = \exp\{-e^{-\alpha_n(y-u_n)}\}$	$u_n = F_X^{-1}\left(1 - \frac{1}{n}\right)$	$\alpha_n = nf_X(u_n)$
	$F_{Y_1}(y) = 1 - \exp\{-e^{-\alpha_1(y-u_1)}\}$	$u_1 = F_X^{-1}\left(\frac{1}{n}\right)$	$\alpha_1 = nf_X(u_1)$
Type II	$F_{Y_n}(y) = \exp\left\{-\left(\frac{u_n}{y}\right)^\alpha\right\}$	$u_n = F_X^{-1}\left(1 - \frac{1}{n}\right)$	α = polynomial order
	$F_{Y_1}(y) = 1 - \exp\left\{-\left(\frac{u_1}{y}\right)^\alpha\right\}$	$u_1 = F_X^{-1}\left(\frac{1}{n}\right)$	
Type III	$F_{Y_n}(y) = \exp\left\{-\left(\frac{u-y}{u-u_n}\right)^\alpha\right\}$	$u_n = F_X^{-1}\left(1 - \frac{1}{n}\right)$	α = polynomial order
	$F_{Y_1}(y) = 1 - \exp\left\{-\left(\frac{y-u}{u_1-u}\right)^\alpha\right\}$	$u_1 = F_X^{-1}\left(\frac{1}{n}\right)$	u = bound value