

# Chapter 2

---

## *Probability Concepts*

---

---

### 2.1 Introduction

A review of probability is covered here at the outset because it provides the foundation for what is to follow: computational statistics. Readers who understand probability concepts may safely skip over this chapter.

Probability is the mechanism by which we can manage the uncertainty that underlies all real world data and phenomena. It enables us to gauge our degree of belief and to quantify the lack of certitude that is inherent in the process that generates the data we are analyzing. For example:

- To understand and use statistical hypothesis testing, one needs knowledge of the sampling distribution of the test statistic.
- To evaluate the performance (e.g., standard error, bias, etc.) of an estimate, we must know its sampling distribution.
- To adequately simulate a real system, one needs to understand the probability distributions that correctly model the underlying processes.
- To build classifiers to predict what group an object belongs to based on a set of features, one can estimate the probability density function that describes the individual classes.

In this chapter, we provide a brief overview of probability concepts and distributions as they pertain to computational statistics. In Section 2.2, we define probability and discuss some of its properties. In Section 2.3, we cover conditional probability, independence and Bayes' Theorem. Expectations are defined in Section 2.4, and common distributions and their uses in modeling physical phenomena are discussed in Section 2.5. In Section 2.6, we summarize some MATLAB functions that implement the ideas from [Chapter 2](#). Finally, in Section 2.7 we provide additional resources for the reader who requires a more theoretical treatment of probability.

---

## 2.2 Probability

### Background

A *random experiment* is defined as a process or action whose outcome cannot be predicted with certainty and would likely change when the experiment is repeated. The variability in the outcomes might arise from many sources: slight errors in measurements, choosing different objects for testing, etc. The ability to model and analyze the outcomes from experiments is at the heart of statistics. Some examples of random experiments that arise in different disciplines are given below.

- Engineering: Data are collected on the number of failures of piston rings in the legs of steam-driven compressors. Engineers would be interested in determining the probability of piston failure in each leg and whether the failure varies among the compressors [Hand, et al., 1994].
- Medicine: The oral glucose tolerance test is a diagnostic tool for early diabetes mellitus. The results of the test are subject to variation because of different rates at which people absorb the glucose, and the variation is particularly noticeable in pregnant women. Scientists would be interested in analyzing and modeling the variation of glucose before and after pregnancy [Andrews and Herzberg, 1985].
- Manufacturing: Manufacturers of cement are interested in the tensile strength of their product. The strength depends on many factors, one of which is the length of time the cement is dried. An experiment is conducted where different batches of cement are tested for tensile strength after different drying times. Engineers would like to determine the relationship between drying time and tensile strength of the cement [Hand, et al., 1994].
- Software Engineering: Engineers measure the failure times in CPU seconds of a command and control software system. These data are used to obtain models to predict the reliability of the software system [Hand, et al., 1994].

The *sample space* is the set of all outcomes from an experiment. It is possible sometimes to list all outcomes in the sample space. This is especially true in the case of some discrete random variables. Examples of these sample spaces are:

- When observing piston ring failures, the sample space is  $\{1, 0\}$ , where 1 represents a failure and 0 represents a non-failure.
- If we roll a six-sided die and count the number of dots on the face, then the sample space is  $\{1, 2, 3, 4, 5, 6\}$ .

The outcomes from random experiments are often represented by an uppercase variable such as  $X$ . This is called a *random variable*, and its value is subject to the uncertainty intrinsic to the experiment. Formally, a random variable is a real-valued function defined on the sample space. As we see in the remainder of the text, a random variable can take on different values according to a probability distribution. Using our examples of experiments from above, a random variable  $X$  might represent the failure time of a software system or the glucose level of a patient. The observed value of a random variable  $X$  is denoted by a lowercase  $x$ . For instance, a random variable  $X$  might represent the number of failures of piston rings in a compressor, and  $x = 5$  would indicate that we observed 5 piston ring failures.

Random variables can be discrete or continuous. A *discrete random variable* can take on values from a finite or countably infinite set of numbers. Examples of discrete random variables are the number of defective parts or the number of typographical errors on a page. A *continuous random variable* is one that can take on values from an interval of real numbers. Examples of continuous random variables are the inter-arrival times of planes at a runway, the average weight of tablets in a pharmaceutical production line or the average voltage of a power plant at different times.

We cannot list all outcomes from an experiment when we observe a continuous random variable, because there are an infinite number of possibilities. However, we could specify the interval of values that  $X$  can take on. For example, if the random variable  $X$  represents the tensile strength of cement, then the sample space might be  $(0, \infty)$  kg/cm<sup>2</sup>.

An *event* is a subset of outcomes in the sample space. An event might be that a piston ring is defective or that the tensile strength of cement is in the range 40 to 50 kg/cm<sup>2</sup>. The probability of an event is usually expressed using the random variable notation illustrated below.

- Discrete Random Variables: Letting 1 represent a defective piston ring and letting 0 represent a good piston ring, then the probability of the event that a piston ring is defective would be written as

$$P(X = 1).$$

- Continuous Random Variables: Let  $X$  denote the tensile strength of cement. The probability that an observed tensile strength is in the range 40 to 50 kg/cm<sup>2</sup> is expressed as

$$P(40 \text{ kg/cm}^2 \leq X \leq 50 \text{ kg/cm}^2).$$

Some events have a special property when they are considered together. Two events that cannot occur simultaneously or jointly are called *mutually exclusive events*. This means that the intersection of the two events is the empty set and the probability of the events occurring together is zero. For example, a piston ring cannot be both defective and good at the same time. So, the event of getting a defective part and the event of getting a good part are mutually exclusive events. The definition of mutually exclusive events can be extended to any number of events by considering all pairs of events. Every pair of events must be mutually exclusive for all of them to be mutually exclusive.

## Probability

**Probability** is a measure of the likelihood that some event will occur. It is also a way to quantify or to gauge the likelihood that an observed measurement or random variable will take on values within some set or range of values. Probabilities always range between 0 and 1. A *probability distribution* of a random variable describes the probabilities associated with each possible value for the random variable.

We first briefly describe two somewhat classical methods for assigning probabilities: the *equal likelihood model* and the *relative frequency method*. When we have an experiment where each of  $n$  outcomes is equally likely, then we assign a probability mass of  $1/n$  to each outcome. This is the equal likelihood model. Some experiments where this model can be used are flipping a fair coin, tossing an unloaded die or randomly selecting a card from a deck of cards.

When the equal likelihood assumption is not valid, then the relative frequency method can be used. With this technique, we conduct the experiment  $n$  times and record the outcome. The probability of event  $E$  is assigned by  $P(E) = f/n$ , where  $f$  denotes the number of experimental outcomes that satisfy event  $E$ .

Another way to find the desired probability that an event occurs is to use a *probability density function* when we have continuous random variables or a *probability mass function* in the case of discrete random variables. Section 2.5 contains several examples of probability density (mass) functions. In this text,  $f(x)$  is used to represent the probability mass or density function for either discrete or continuous random variables, respectively. We now discuss how to find probabilities using these functions, first for the continuous case and then for discrete random variables.

To find the probability that a continuous random variable falls in a particular interval of real numbers, we have to calculate the appropriate area under the curve of  $f(x)$ . Thus, we have to evaluate the integral of  $f(x)$  over the interval of random variables corresponding to the event of interest. This is represented by

$$P(a \leq X \leq b) = \int_a^b f(x)dx. \quad (2.1)$$

The area under the curve of  $f(x)$  between  $a$  and  $b$  represents the probability that an observed value of the random variable  $X$  will assume a value between  $a$  and  $b$ . This concept is illustrated in Figure 2.1 where the shaded area represents the desired probability.

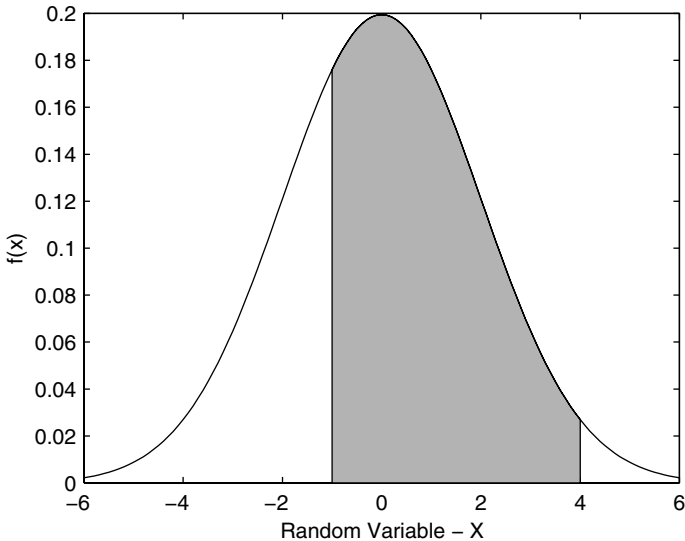


FIGURE 2.1

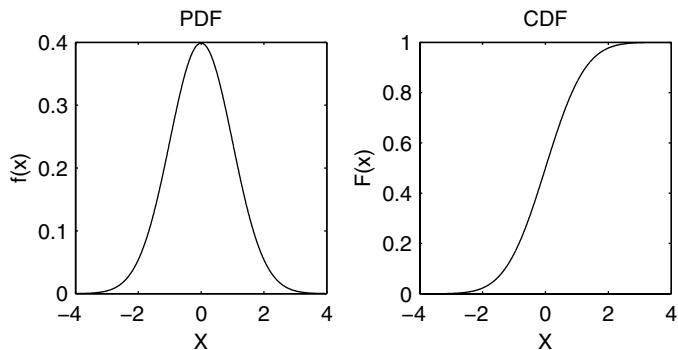
The area under the curve of  $f(x)$  between -1 and 4 is the same as the probability that an observed value of the random variable will assume a value in the same interval.

It should be noted that a valid probability density function should be non-negative, and the total area under the curve must equal 1. If this is not the case, then the probabilities will not be properly restricted to the interval  $[0, 1]$ . This will be an important consideration in Chapter 8 where we discuss probability density estimation techniques.

The *cumulative distribution function*  $F(x)$  is defined as the probability that the random variable  $X$  assumes a value less than or equal to a given  $x$ . This is calculated from the probability density function, as follows

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt. \quad (2.2)$$

It is obvious that the cumulative distribution function takes on values between 0 and 1, so  $0 \leq F(x) \leq 1$ . A probability density function, along with its associated cumulative distribution function are illustrated in [Figure 2.2](#).



**FIGURE 2.2**

This shows the probability density function on the left with the associated cumulative distribution function on the right. Notice that the cumulative distribution function takes on values between 0 and 1.

For a discrete random variable  $X$ , that can take on values  $x_1, x_2, \dots$ , the probability mass function is given by

$$f(x_i) = P(X = x_i); \quad i = 1, 2, \dots, \quad (2.3)$$

and the cumulative distribution function is

$$F(a) = \sum_{x_i \leq a} f(x_i); \quad i = 1, 2, \dots. \quad (2.4)$$

## Axioms of Probability

Probabilities follow certain axioms that can be useful in computational statistics. We let  $S$  represent the sample space of an experiment and  $E$  represent some event that is a subset of  $S$ .

### AXIOM 1

*The probability of event  $E$  must be between 0 and 1:*

$$0 \leq P(E) \leq 1 .$$

### AXIOM 2

$$P(S) = 1 .$$

### AXIOM 3

*For mutually exclusive events,  $E_1, E_2, \dots, E_k$ ,*

$$P(E_1 \cup E_2 \cup \dots \cup E_k) = \sum_{i=1}^k P(E_i) .$$

Axiom 1 has been discussed before and simply states that a probability must be between 0 and 1. Axiom 2 says that an outcome from our experiment must occur, and the probability that the outcome is in the sample space is 1. Axiom 3 enables us to calculate the probability that at least one of the mutually exclusive events  $E_1, E_2, \dots, E_k$  occurs by summing the individual probabilities.

---

## 2.3 Conditional Probability and Independence

### Conditional Probability

Conditional probability is an important concept. It is used to define independent events and enables us to revise our degree of belief given that another event has occurred. Conditional probability arises in situations where we need to calculate a probability based on some partial information concerning the experiment.

The *conditional probability* of event  $E$  given event  $F$  is defined as follows:

## CONDITIONAL PROBABILITY

$$P(E|F) = \frac{P(E \cap F)}{P(F)}; \quad P(F) > 0. \quad (2.5)$$

Here  $P(E \cap F)$  represents the **joint probability** that both  $E$  and  $F$  occur together and  $P(F)$  is the probability that event  $F$  occurs. We can rearrange Equation 2.5 to get the following rule:

### MULTIPLICATION RULE

$$P(E \cap F) = P(F)P(E|F). \quad (2.6)$$

## Independence

Often we can assume that the occurrence of one event does not affect whether or not some other event happens. For example, say a couple would like to have two children, and their first child is a boy. The gender of their second child does not depend on the gender of the first child. Thus, the fact that we know they have a boy already does not change the probability that the second child is a boy. Similarly, we can sometimes assume that the value we observe for a random variable is not affected by the observed value of other random variables.

These types of events and random variables are called **independent**. If *events* are independent, then knowing that one event has occurred does not change our degree of belief or the likelihood that the other event occurs. If *random variables* are independent, then the observed value of one random variable does not affect the observed value of another.

In general, the conditional probability  $P(E|F)$  is not equal to  $P(E)$ . In these cases, the events are called **dependent**. Sometimes we can assume independence based on the situation or the experiment, which was the case with our example above. However, to show independence mathematically, we must use the following definition.

### INDEPENDENT EVENTS

*Two events  $E$  and  $F$  are said to be independent if and only if any of the following is true:*

$$\begin{aligned} P(E \cap F) &= P(E)P(F), \\ P(E) &= P(E|F). \end{aligned} \quad (2.7)$$



Note that if events  $E$  and  $F$  are independent, then the Multiplication Rule in Equation 2.6 becomes

$$P(E \cap F) = P(F)P(E),$$

which means that we simply multiply the individual probabilities for each event together. This can be extended to  $k$  events to give

$$P(E_1 \cap E_2 \cap \dots \cap E_k) = \prod_{i=1}^k P(E_i), \quad (2.8)$$

where events  $E_i$  and  $E_j$  (for all  $i$  and  $j$ ,  $i \neq j$ ) are independent.

## Bayes Theorem

Sometimes we start an analysis with an initial degree of belief that an event will occur. Later on, we might obtain some additional information about the event that would change our belief about the probability that the event will occur. The initial probability is called a *prior probability*. Using the new information, we can update the prior probability using Bayes' Theorem to obtain the *posterior probability*.

The experiment of recording piston ring failure in compressors is an example of where Bayes' Theorem might be used, and we derive Bayes' Theorem using this example. Suppose our piston rings are purchased from two manufacturers: 60% from manufacturer A and 40% from manufacturer B.

Let  $M_A$  denote the event that a part comes from manufacturer A, and  $M_B$  represent the event that a piston ring comes from manufacturer B. If we select a part at random from our supply of piston rings, we would assign probabilities to these events as follows:

$$P(M_A) = 0.6,$$

$$P(M_B) = 0.4.$$

These are our prior probabilities that the piston rings are from the individual manufacturers.

Say we are interested in knowing the probability that a piston ring that subsequently failed came from manufacturer A. This would be the posterior probability that it came from manufacturer A, given that the piston ring failed. The additional information we have about the piston ring is that it failed, and we use this to update our degree of belief that it came from manufacturer A.

Bayes' Theorem can be derived from the definition of conditional probability (Equation 2.5). Writing this in terms of our events, we are interested in the following probability:

$$P(M_A|F) = \frac{P(M_A \cap F)}{P(F)}, \quad (2.9)$$

where  $P(M_A|F)$  represents the posterior probability that the part came from manufacturer A, and  $F$  is the event that the piston ring failed. Using the Multiplication Rule (Equation 2.6), we can write the numerator of Equation 2.9 in terms of event  $F$  and our prior probability that the part came from manufacturer A, as follows

$$P(M_A|F) = \frac{P(M_A \cap F)}{P(F)} = \frac{P(M_A)P(F|M_A)}{P(F)}. \quad (2.10)$$

The next step is to find  $P(F)$ . The only way that a piston ring will fail is if: 1) it failed and it came from manufacturer A or 2) it failed and it came from manufacturer B. Thus, using the third axiom of probability, we can write

$$P(F) = P(M_A \cap F) + P(M_B \cap F).$$

Applying the Multiplication Rule as before, we have

$$P(F) = P(M_A)P(F|M_A) + P(M_B)P(F|M_B). \quad (2.11)$$

Substituting this for  $P(F)$  in Equation 2.10, we write the posterior probability as

$$P(M_A|F) = \frac{P(M_A)P(F|M_A)}{P(M_A)P(F|M_A) + P(M_B)P(F|M_B)}. \quad (2.12)$$

Note that we need to find the probabilities  $P(F|M_A)$  and  $P(F|M_B)$ . These are the probabilities that a piston ring will fail given it came from the corresponding manufacturer. These must be estimated in some way using available information (e.g., past failures). When we revisit Bayes' Theorem in the context of statistical pattern recognition (Chapter 9), these are the probabilities that are estimated to construct a certain type of classifier.

Equation 2.12 is Bayes' Theorem for a situation where only two outcomes are possible. In general, Bayes' Theorem can be written for any number of mutually exclusive events,  $E_1, \dots, E_k$ , whose union makes up the entire sample space. This is given below.

$$P(E_i|F) = \frac{P(E_i)P(F|E_i)}{P(E_1)P(F|E_1) + \dots + P(E_k)P(F|E_k)}. \quad (2.13)$$

## 2.4 Expectation

Expected values and variances are important concepts in statistics. They are used to describe distributions, to evaluate the performance of estimators, to obtain test statistics in hypothesis testing, and many other applications.

### Mean and Variance

The *mean* or *expected value* of a random variable is defined using the probability density (mass) function. It provides a measure of central tendency of the distribution. If we observe many values of the random variable and take the average of them, we would expect that value to be close to the mean. The expected value is defined below for the discrete case.

#### EXPECTED VALUE - DISCRETE RANDOM VARIABLES

$$\mu = E[X] = \sum_{i=1}^{\infty} x_i f(x_i). \quad (2.14)$$

We see from the definition that the expected value is a sum of all possible values of the random variable where each one is weighted by the probability that  $X$  will take on that value.

The *variance* of a discrete random variable is given by the following definition.

#### VARIANCE - DISCRETE RANDOM VARIABLES

For  $\mu < \infty$ ,

$$\sigma^2 = V(X) = E[(X - \mu)^2] = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i). \quad (2.15)$$

From Equation 2.15, we see that the variance is the sum of the squared distances, each one weighted by the probability that  $X = x_i$ . Variance is a measure of dispersion in the distribution. If a random variable has a large variance, then an observed value of the random variable is more likely to be far from the mean  $\mu$ . The standard deviation  $\sigma$  is the square root of the variance.

The mean and variance for continuous random variables are defined similarly, with the summation replaced by an integral. The mean and variance of a continuous random variable are given below.

#### EXPECTED VALUE - CONTINUOUS RANDOM VARIABLES

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx. \quad (2.16)$$

#### VARIANCE - CONTINUOUS RANDOM VARIABLES

For  $\mu < \infty$ ,

$$\sigma^2 = V(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx. \quad (2.17)$$

We note that Equation 2.17 can also be written as

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

Other expected values that are of interest in statistics are the **moments** of a random variable. These are the expectation of powers of the random variable. In general, we define the ***r*-th moment** as

$$\mu'_r = E[X^r], \quad (2.18)$$

and the ***r*-th central moment** as

$$\mu_r = E[(X - \mu)^r]. \quad (2.19)$$

The mean corresponds to  $\mu'_1$  and the variance is given by  $\mu_2$ .

## Skewness

The third central moment  $\mu_3$  is often called a measure of asymmetry or skewness in the distribution. The uniform and the normal distribution are examples of symmetric distributions. The gamma and the exponential are examples of skewed or asymmetric distributions. The following ratio is called the *coefficient of skewness*, which is often used to measure this characteristic:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}. \quad (2.20)$$

Distributions that are skewed to the left will have a negative coefficient of skewness, and distributions that are skewed to the right will have a positive value [Hogg and Craig, 1978]. The coefficient of skewness is zero for symmetric distributions. However, a coefficient of skewness equal to zero does not mean that the distribution must be symmetric.

## Kurtosis

Skewness is one way to measure a type of departure from normality. *Kurtosis* measures a different type of departure from normality by indicating the extent of the peak (or the degree of flatness near its center) in a distribution. The *coefficient of kurtosis* is given by the following ratio:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2}. \quad (2.21)$$

We see that this is the ratio of the fourth central moment divided by the square of the variance. If the distribution is normal, then this ratio is equal to 3. A ratio greater than 3 indicates more values in the neighborhood of the mean (is more peaked than the normal distribution). If the ratio is less than 3, then it is an indication that the curve is flatter than the normal.

Sometimes the *coefficient of excess kurtosis* is used as a measure of kurtosis. This is given by

$$\gamma_2' = \frac{\mu_4}{\mu_2^2} - 3. \quad (2.22)$$

In this case, distributions that are more peaked than the normal correspond to a positive value of  $\gamma_2'$ , and those with a flatter top have a negative coefficient of excess kurtosis.

## 2.5 Common Distributions

In this section, we provide a review of some useful probability distributions and briefly describe some applications to modeling data. Most of these distributions are used in later chapters, so we take this opportunity to define them and to fix our notation. We first cover two important discrete distributions: the binomial and the Poisson. These are followed by several continuous distributions: the uniform, the normal, the exponential, the gamma, the chi-square, the Weibull, the beta and the multivariate normal.

### Binomial

Let's say that we have an experiment, whose outcome can be labeled as a 'success' or a 'failure'. If we let  $X = 1$  denote a successful outcome and  $X = 0$  represent a failure, then we can write the probability mass function as

$$\begin{aligned} f(0) &= P(X = 0) = 1 - p, \\ f(1) &= P(X = 1) = p, \end{aligned} \quad (2.23)$$

where  $p$  represents the probability of a successful outcome. A random variable that follows the probability mass function in Equation 2.23 for  $0 < p < 1$  is called a Bernoulli random variable.

Now suppose we repeat this experiment for  $n$  trials, where each trial is independent (the outcome from one trial does not influence the outcome of another) and results in a success with probability  $p$ . If  $X$  denotes the number of successes in these  $n$  trials, then  $X$  follows the binomial distribution with parameters  $(n, p)$ . Examples of binomial distributions with different parameters are shown in [Figure 2.3](#).

To calculate a binomial probability, we use the following formula:

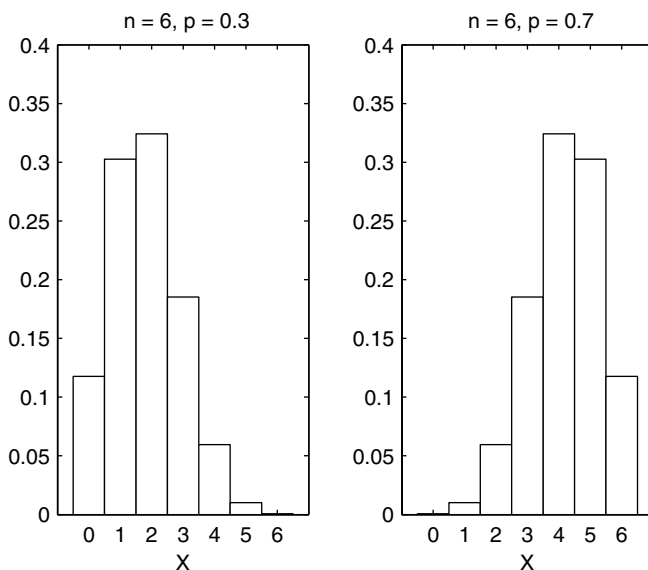
$$f(x; n, p) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}; \quad x = 0, 1, \dots, n. \quad (2.24)$$

The mean and variance of a binomial distribution are given by

$$E[X] = np,$$

and

$$V(X) = np(1 - p).$$



**FIGURE 2.3**

Examples of the binomial distribution for different success probabilities.

Some examples where the results of an experiment can be modeled by a binomial random variable are:

- A drug has probability 0.90 of curing a disease. It is administered to 100 patients, where the outcome for each patient is either cured or not cured. If  $X$  is the number of patients cured, then  $X$  is a binomial random variable with parameters (100, 0.90).
- The National Institute of Mental Health estimates that there is a 20% chance that an adult American suffers from a psychiatric disorder. Fifty adult Americans are randomly selected. If we let  $X$  represent the number who have a psychiatric disorder, then  $X$  takes on values according to the binomial distribution with parameters (50, 0.20).
- A manufacturer of computer chips finds that on the average 5% are defective. To monitor the manufacturing process, they take a random sample of size 75. If the sample contains more than five defective chips, then the process is stopped. The binomial distribution with parameters (75, 0.05) can be used to model the random variable  $X$ , where  $X$  represents the number of defective chips.

### Example 2.1

Suppose there is a 20% chance that an adult American suffers from a psychiatric disorder. We randomly sample 25 adult Americans. If we let  $X$  represent the number of people who have a psychiatric disorder, then  $X$  is a binomial random variable with parameters  $(25, 0.20)$ . We are interested in the probability that at most 3 of the selected people have such a disorder. We can use the MATLAB Statistics Toolbox function `binocdf` to determine  $P(X \leq 3)$ , as follows:

```
prob = binocdf(3,25,0.2);
```

We could also sum up the individual values of the probability mass function from  $X = 0$  to  $X = 3$ :

```
prob2 = sum(binopdf(0:3,25,0.2));
```

Both of these commands return a probability of 0.234. We now show how to generate the binomial distributions shown in [Figure 2.3](#).

```
% Get the values for the domain, x.
x = 0:6;
% Get the values of the probability mass function.
% First for n = 6, p = 0.3:
pdf1 = binopdf(x,6,0.3);
% Now for n = 6, p = 0.7:
pdf2 = binopdf(x,6,0.7);
```

Now we have the values for the probability mass function (or the heights of the bars). The plots are obtained using the following code.

```
% Do the plots.
subplot(1,2,1),bar(x,pdf1,1,'w')
title(' n = 6, p = 0.3')
xlabel('X'),ylabel('f(X)')
axis square
subplot(1,2,2),bar(x,pdf2,1,'w')
title(' n = 6, p = 0.7')
xlabel('X'),ylabel('f(X)')
axis square
```

□

### Poisson

A random variable  $X$  is a Poisson random variable with parameter  $\lambda$ ,  $\lambda > 0$ , if it follows the probability mass function given by

$$f(x; \lambda) = P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}; \quad x = 0, 1, \dots \quad (2.25)$$



The expected value and variance of a Poisson random variable are both  $\lambda$ , thus,

$$E[X] = \lambda,$$

and

$$V(X) = \lambda.$$

The Poisson distribution can be used in many applications. Examples of situations where a discrete random variable might follow a Poisson distribution are:

- the number of typographical errors on a page,
- the number of vacancies in a company during a month, or
- the number of defects in a length of wire.

The Poisson distribution is often used to approximate the binomial. When  $n$  is large and  $p$  is small (so  $np$  is moderate), then the number of successes occurring can be approximated by the Poisson random variable with parameter  $\lambda = np$ .

The Poisson distribution is also appropriate for some applications where events occur at points in time or space. We see it used in this context in Chapter 12, where we look at modeling spatial point patterns. Some other examples include the arrival of jobs at a business, the arrival of aircraft on a runway, and the breakdown of machines at a manufacturing plant. The number of events in these applications can be described by a *Poisson process*.

Let  $N(t)$ ,  $t \geq 0$ , represent the number of events that occur in the time interval  $[0, t]$ . For each interval  $[0, t]$ ,  $N(t)$  is a random variable that can take on values  $0, 1, 2, \dots$ . If the following conditions are satisfied, then the counting process  $\{N(t), t \geq 0\}$  is said to be a Poisson process with mean rate  $\lambda$  [Ross, 2000]:

1.  $N(0) = 0$ .
2. The process has independent increments.
3. The number  $N(t)$  of events in an interval of length  $t$  follows a Poisson distribution with mean  $\lambda t$ . Thus, for  $s \geq 0$ ,  $t \geq 0$ ,

$$P(N(t+s) - N(s) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}; \quad k = 0, 1, \dots \quad (2.26)$$

From the third condition, we know that the process has stationary increments. This means that the distribution of the number of events in an interval depends only on the length of the interval and not on the starting point. The

second condition specifies that the number of events in one interval does not affect the number of events in other intervals. The first condition states that the counting starts at time  $t = 0$ . The expected value of  $N(t)$  is given by

$$E[N(t)] = \lambda t.$$

### Example 2.2

In preparing this text, we executed the spell check command, and the editor reviewed the manuscript for typographical errors. In spite of this, some mistakes might be present. Assume that the number of typographical errors per page follows the Poisson distribution with parameter  $\lambda = 0.25$ . We calculate the probability that a page will have at least two errors as follows:

$$P(X \geq 2) = 1 - \{P(X = 0) + P(X = 1)\} = 1 - e^{-0.25} - e^{-0.25}0.25 \approx 0.0265.$$

We can get this probability using the MATLAB Statistics Toolbox function **poisscdf**. Note that  $P(X = 0) + P(X = 1)$  is the Poisson cumulative distribution function for  $a = 1$  (see Equation 2.4), which is why we use **1** as the argument to **poisscdf**.

```
prob = 1-poisscdf(1,0.25);
```

□

### Example 2.3

Suppose that accidents at a certain intersection occur in a manner that satisfies the conditions for a Poisson process with a rate of 2 per week ( $\lambda = 2$ ). What is the probability that at most 3 accidents will occur during the next 2 weeks? Using Equation 2.26, we have

$$P(N(2) \leq 3) = \sum_{k=0}^3 P(N(2) = k).$$

Expanding this out yields

$$P(N(2) \leq 3) = e^{-4} + 4e^{-4} + \frac{4^2}{2!}e^{-4} + \frac{4^3}{3!}e^{-4} \approx 0.4335.$$

As before, we can use the **poisscdf** function with parameter given by  $\lambda t = 2 \cdot 2$ .

```
prob = poisscdf(3,2*2);
```

□

## Uniform

Perhaps one of the most important distributions is the uniform distribution for continuous random variables. One reason is that the uniform  $(0, 1)$  distribution is used as the basis for simulating most random variables as we discuss in Chapter 4.

A random variable that is uniformly distributed over the interval  $(a, b)$  follows the probability density function given by

$$f(x; a, b) = \frac{1}{b-a}; \quad a < x < b. \quad (2.27)$$

The parameters for the uniform are the interval endpoints,  $a$  and  $b$ . The mean and variance of a uniform random variable are given by

$$E[X] = \frac{a+b}{2},$$

and

$$V(X) = \frac{(b-a)^2}{12}.$$

The cumulative distribution function for a uniform random variable is

$$F(x) = \begin{cases} 0; & x \leq a \\ \frac{x-a}{b-a}; & a < x < b \\ 1; & x \geq b. \end{cases} \quad (2.28)$$

### Example 2.4

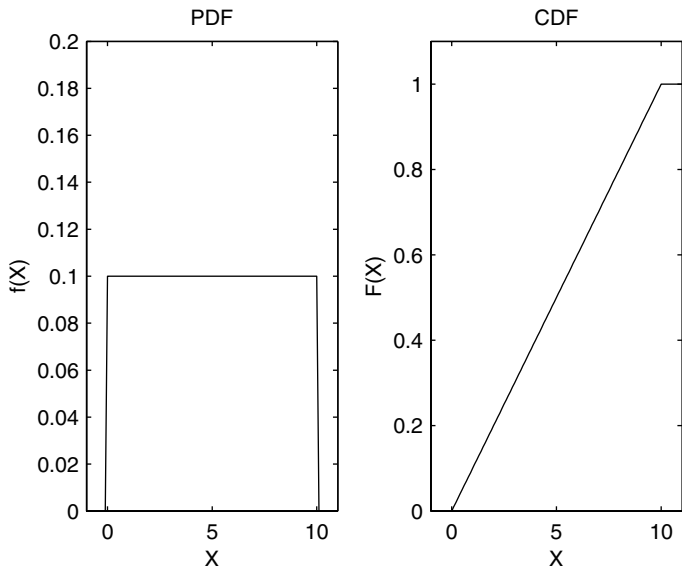
In this example, we illustrate the uniform probability density function over the interval  $(0, 10)$ , along with the corresponding cumulative distribution function. The MATLAB Statistics Toolbox functions **unifpdf** and **unifcdf** are used to get the desired functions over the interval.

```
% First get the domain over which we will
% evaluate the functions.
x = -1:.1:11;
% Now get the probability density function
% values at x.
pdf = unifpdf(x,0,10);
% Now get the cdf.
cdf = unifcdf(x,0,10);
```

Plots of the functions are provided in [Figure 2.4](#), where the probability density function is shown in the left plot and the cumulative distribution on the right. These plots are constructed using the following MATLAB commands.

```
% Do the plots.
subplot(1,2,1),plot(x,pdf)
title('PDF')
xlabel('X'),ylabel('f(X)')
axis([-1 11 0 0.2])
axis square
subplot(1,2,2),plot(x,cdf)
title('CDF')
xlabel('X'),ylabel('F(X)')
axis([-1 11 0 1.1])
axis square
```

□



**FIGURE 2.4** On the left is a plot of the probability density function for the uniform (0, 10). Note that the height of the curve is given by  $1/(b-a) = 1/10 = 0.10$ . The corresponding cumulative distribution function is shown on the right.

## Normal

A well known distribution in statistics and engineering is the normal distribution. Also called the Gaussian distribution, it has a continuous probability density function given by

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad (2.29)$$

where  $-\infty < x < \infty$ ;  $-\infty < \mu < \infty$ ;  $\sigma^2 > 0$ . The normal distribution is completely determined by its parameters ( $\mu$  and  $\sigma^2$ ), which are also the expected value and variance for a normal random variable. The notation  $X \sim N(\mu, \sigma^2)$  is used to indicate that a random variable  $X$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ . Several normal distributions with different parameters are shown in [Figure 2.5](#).

Some special properties of the normal distribution are given here.

- The value of the probability density function approaches zero as  $x$  approaches positive and negative infinity.
- The probability density function is centered at the mean  $\mu$ , and the maximum value of the function occurs at  $x = \mu$ .
- The probability density function for the normal distribution is symmetric about the mean  $\mu$ .

The special case of a standard normal random variable is one whose mean is zero ( $\mu = 0$ ), and whose standard deviation is one ( $\sigma = 1$ ). If  $X$  is normally distributed, then

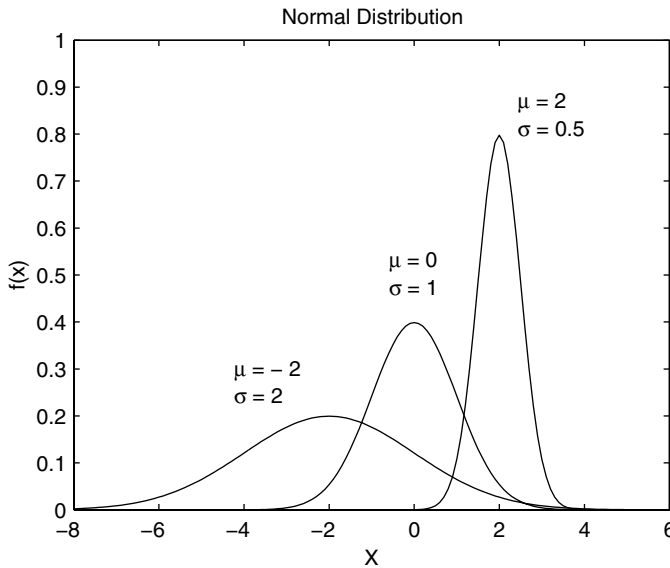
$$Z = \frac{X - \mu}{\sigma} \quad (2.30)$$

is a standard normal random variable.

Traditionally, the cumulative distribution function of a standard normal random variable is denoted by

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left\{-\frac{y^2}{2}\right\} dy. \quad (2.31)$$

The cumulative distribution function for a standard normal random variable can be calculated using the error function, denoted by *erf*. The relationship between these functions is given by



**FIGURE 2.5**

Examples of probability density functions for normally distributed random variables. Note that as the variance increases, the height of the probability density function at the mean decreases.

$$\Phi(z) = \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) + \frac{1}{2}. \quad (2.32)$$

The error function can be calculated in MATLAB using `erf(x)`. The MATLAB Statistics Toolbox has a function called `normcdf(x, mu, sigma)` that will calculate the cumulative distribution function for values in  $x$ . Its use is illustrated in the example given below.

### Example 2.5

Similar to the uniform distribution, the functions `normpdf` and `normcdf` are available in the MATLAB Statistics Toolbox for calculating the probability density function and cumulative distribution function for the normal. There is another special function called `normspec` that determines the probability that a random variable  $X$  assumes a value between two limits, where  $X$  is normally distributed with mean  $\mu$  and standard deviation  $\sigma$ . This function also plots the normal density, where the area between the specified limits is shaded. The syntax is shown below.

```
% Set up the parameters for the normal distribution.
mu = 5;
sigma = 2;
% Set up the upper and lower limits. These are in
% the two element vector 'specs'.
specs = [2, 8];
prob = normspec(specs, mu, sigma);
```

The resulting plot is shown in [Figure 2.6](#). Note that the default title and axes labels are shown, but these can be changed easily using the `title`, `xlabel`, and `ylabel` functions. You can also obtain tail probabilities by using `-Inf` as the first element of `specs` to designate no lower limit or `Inf` as the second element to indicate no upper limit.

□

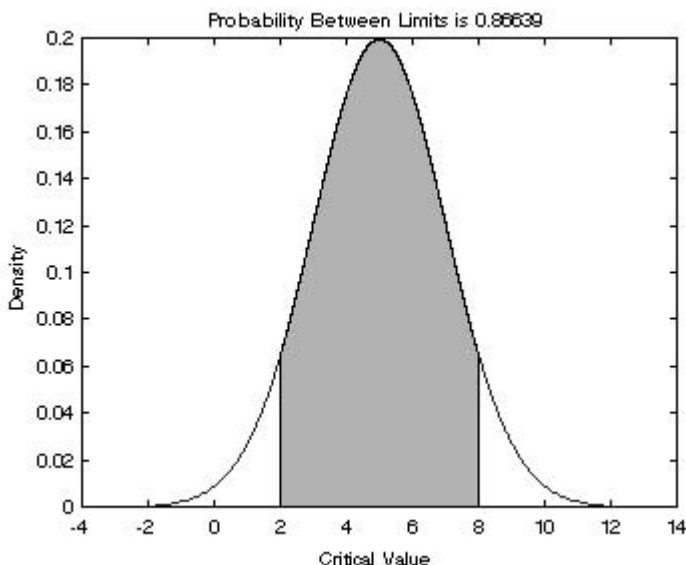


FIGURE 2.6

This shows the output from the function `normspec`. Note that it shades the area between the lower and upper limits that are specified as input arguments.

## Exponential

The exponential distribution can be used to model the amount of time until a specific event occurs or to model the time between independent events. Some examples where an exponential distribution could be used as the model are:

- the time until the computer locks up,
- the time between arrivals of telephone calls, or
- the time until a part fails.

The exponential probability density function with parameter  $\lambda$  is

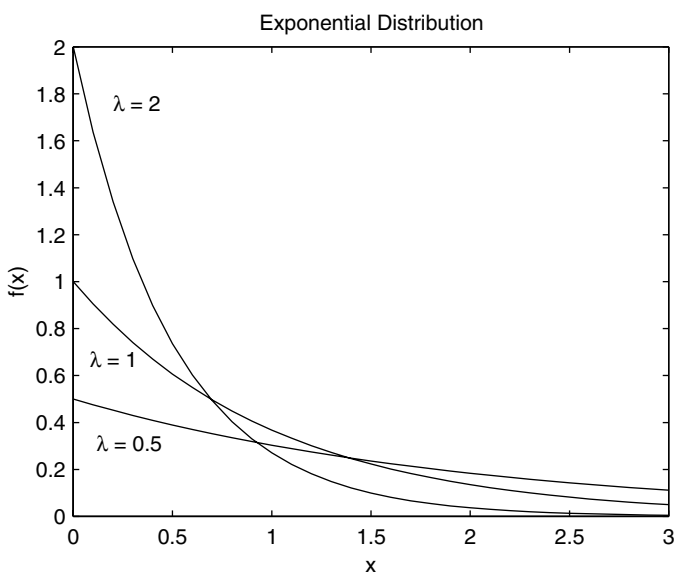
$$f(x;\lambda) = \lambda e^{-\lambda x}; \quad x \geq 0; \quad \lambda > 0. \quad (2.33)$$

The mean and variance of an exponential random variable are given by the following:

$$E[X] = \frac{1}{\lambda},$$

and

$$V(X) = \frac{1}{\lambda^2}.$$



**FIGURE 2.7**  
Exponential probability density functions for various values of  $\lambda$ .



The cumulative distribution function of an exponential random variable is given by

$$F(x) = \begin{cases} 0; & x < 0 \\ 1 - e^{-\lambda x}; & x \geq 0. \end{cases} \quad (2.34)$$

The exponential distribution is the only continuous distribution that has the memoryless property. This property describes the fact that the remaining lifetime of an object (whose lifetime follows an exponential distribution) does not depend on the amount of time it has already lived. This property is represented by the following equality, where  $s \geq 0$  and  $t \geq 0$ :

$$P(X > s + t | X > s) = P(X > t).$$

In words, this means that the probability that the object will operate for time  $s + t$ , given it has already operated for time  $s$ , is simply the probability that it operates for time  $t$ .

When the exponential is used to represent interarrival times, then the parameter  $\lambda$  is a rate with units of arrivals per time period. When the exponential is used to model the time until a failure occurs, then  $\lambda$  is the failure rate. Several examples of the exponential distribution are shown in [Figure 2.7](#).

### Example 2.6

The time between arrivals of vehicles at an intersection follows an exponential distribution with a mean of 12 seconds. What is the probability that the time between arrivals is 10 seconds or less? We are given the average interarrival time, so  $\lambda = 1/12$ . The required probability is obtained from Equation 2.34 as follows

$$P(X \leq 10) = 1 - e^{-(1/12)10} \approx 0.57.$$

You can calculate this using the MATLAB Statistics Toolbox function **expocdf** (**x**, **1/λ**). Note that this MATLAB function is based on a different definition of the exponential probability density function, which is given by

$$f(x; \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}; \quad x \geq 0; \quad \mu > 0. \quad (2.35)$$

In the Computational Statistics Toolbox, we include a function called **csexpoc**( $\mathbf{x}, \lambda$ ) that calculates the exponential cumulative distribution function using Equation 2.34.

□

## Gamma

The gamma probability density function with parameters  $\lambda > 0$  and  $t > 0$  is

$$f(x; \lambda, t) = \frac{\lambda e^{-\lambda x} (\lambda x)^{t-1}}{\Gamma(t)}; \quad x \geq 0, \quad (2.36)$$

where  $t$  is a shape parameter, and  $\lambda$  is the scale parameter. The gamma function  $\Gamma(t)$  is defined as

$$\Gamma(t) = \int_0^{\infty} e^{-y} y^{t-1} dy. \quad (2.37)$$

For integer values of  $t$ , Equation 2.37 becomes

$$\Gamma(t) = (t-1)!. \quad (2.38)$$

Note that for  $t = 1$ , the gamma density is the same as the exponential. When  $t$  is a positive integer, the gamma distribution can be used to model the amount of time one has to wait until  $t$  events have occurred, if the inter-arrival times are exponentially distributed.

The mean and variance of a gamma random variable are

$$E[X] = \frac{t}{\lambda},$$

and

$$V(X) = \frac{t}{\lambda^2}.$$

The cumulative distribution function for a gamma random variable is calculated using [Meeker and Escobar, 1998; Banks, et al., 2001]

$$F(x;\lambda, t) = \begin{cases} 0; & x \leq 0 \\ \frac{1}{\Gamma(t)} \int_0^{\lambda x} y^{t-1} e^{-y} dy; & x > 0. \end{cases} \quad (2.39)$$

Equation 2.39 can be evaluated easily in MATLAB using the **gammainc**( $\lambda * \mathbf{x}, \mathbf{t}$ ) function, where the above notation is used for the arguments.

### Example 2.7

We plot the gamma probability density function for  $\lambda = t = 1$  (this should look like the exponential),  $\lambda = t = 2$ , and  $\lambda = t = 3$ . You can use the MATLAB Statistics Toolbox function **gampdf**( $\mathbf{x}, \mathbf{t}, 1/\lambda$ ) or the function **csgammp**( $\mathbf{x}, \mathbf{t}, \lambda$ ).

```
% First get the domain over which to
% evaluate the functions.
x = 0:.1:3;
% Now get the functions values for
% different values of lambda.
y1 = gampdf(x,1,1/1);
y2 = gampdf(x,2,1/2);
y3 = gampdf(x,3,1/3);
% Plot the functions.
plot(x,y1,'r',x,y2,'g',x,y3,'b')
title('Gamma Distribution')
xlabel('X')
ylabel('f(x)')
```

The resulting curves are shown in [Figure 2.8](#).

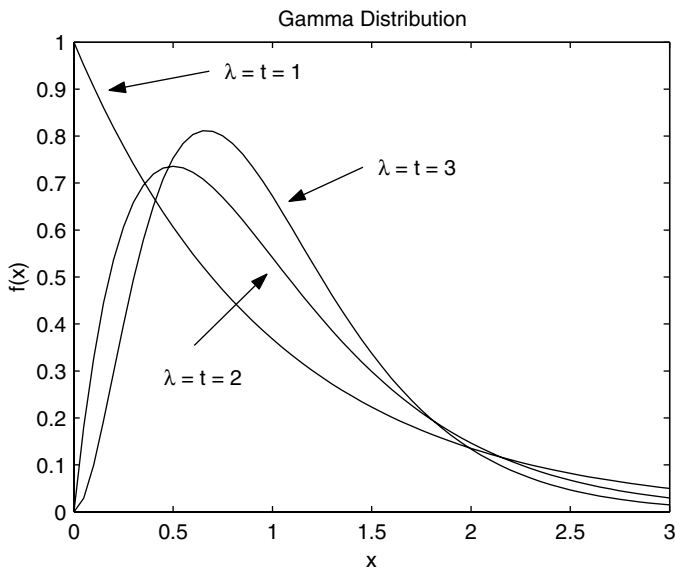


### Chi-Square

A gamma distribution where  $\lambda = 0.5$  and  $t = v/2$ , with  $v$  a positive integer, is called a chi-square distribution (denoted as  $\chi_v^2$ ) with  $v$  degrees of freedom. The chi-square distribution is used to derive the distribution of the sample variance and is important for goodness-of-fit tests in statistical analysis [Mood, Graybill, and Boes, 1974].

The probability density function for a chi-square random variable with  $v$  degrees of freedom is

$$f(x;v) = \frac{1}{\Gamma(v/2)} \left(\frac{1}{2}\right)^{v/2} x^{v/2-1} e^{-\frac{1}{2}x}; \quad x \geq 0. \quad (2.40)$$



**FIGURE 2.8**

We show three examples of the gamma probability density function. We see that when  $\lambda = t = 1$ , we have the same probability density function as the exponential with parameter  $\lambda = 1$ .

The mean and variance of a chi-square random variable can be obtained from the gamma distribution. These are given by

$$E[X] = v,$$

and

$$V(X) = 2v.$$

## Weibull

The Weibull distribution has many applications in engineering. In particular, it is used in reliability analysis. It can be used to model the distribution of the amount of time it takes for objects to fail. For the special case where  $v = 0$  and  $\beta = 1$ , the Weibull reduces to the exponential with  $\lambda = 1/\alpha$ .

The Weibull density for  $\alpha > 0$  and  $\beta > 0$  is given by

$$f(x;v, \alpha, \beta) = \left(\frac{\beta}{\alpha}\right)\left(\frac{x-v}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-v}{\alpha}\right)^{\beta}}; \quad x > v, \quad (2.41)$$

and the cumulative distribution is

$$F(x;v, \alpha, \beta) = \begin{cases} 0; & x \leq v \\ 1 - e^{-\left(\frac{x-v}{\alpha}\right)^{\beta}}; & x > v. \end{cases} \quad (2.42)$$

The location parameter is denoted by  $v$ , and the scale parameter is given by  $\alpha$ . The shape of the Weibull distribution is governed by the parameter  $\beta$ .

The mean and variance [Banks, et al., 2001] of a random variable from a Weibull distribution are given by

$$E[X] = v + \alpha\Gamma(1/\beta + 1),$$

and

$$V(X) = \alpha^2 \left\{ \Gamma(2/\beta + 1) - [\Gamma(1/\beta + 1)]^2 \right\}.$$

### Example 2.8

Suppose the time to failure of piston rings for stream-driven compressors can be modeled by the Weibull distribution with a location parameter of zero,  $\beta = 1/3$ , and  $\alpha = 500$ . We can find the mean time to failure using the expected value of a Weibull random variable, as follows

$$E[X] = v + \alpha\Gamma(1/\beta + 1) = 500 \times \Gamma(3 + 1) = 3000 \text{ hours.}$$

Let's say we want to know the probability that a piston ring will fail before 2000 hours. We can calculate this probability using

$$F(2000;0, 500, 1/3) = 1 - \exp\left\{-\left(\frac{2000}{500}\right)^{1/3}\right\} \approx 0.796.$$

□

You can use the MATLAB Statistics Toolbox function for applications where the location parameter is zero ( $v = 0$ ). This function is called

**weibcdf** (for the cumulative distribution function), and the input arguments are:  $(\mathbf{x}, \alpha^{-\beta}, \beta)$ . The reason for the different parameters is that MATLAB uses an alternate definition for the Weibull probability density function given by

$$f(x; a, b) = abx^{b-1}e^{-ax^b}; \quad x > 0. \quad (2.43)$$

Comparing this with Equation 2.41, we can see that  $v = 0$ ,  $a = \alpha^{-\beta}$  and  $b = \beta$ . You can also use the function **csweibc**  $(\mathbf{x}, v, \alpha, \beta)$  to evaluate the cumulative distribution function for a Weibull.

## Beta

The beta distribution is very flexible because it covers a range of different shapes depending on the values of the parameters. It can be used to model a random variable that takes on values over a bounded interval and assumes one of the shapes governed by the parameters. A random variable has a beta distribution with parameters  $\alpha > 0$  and  $\beta > 0$  if its probability density function is given by

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}; \quad 0 < x < 1, \quad (2.44)$$

where

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \quad (2.45)$$

The function  $B(\alpha, \beta)$  can be calculated in MATLAB using the **beta**  $(\alpha, \beta)$  function. The mean and variance of a beta random variable are

$$E[X] = \frac{\alpha}{\alpha + \beta},$$

and

$$V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

The cumulative distribution function for a beta random variable is given by integrating the beta probability density function as follows

$$F(x; \alpha, \beta) = \int_0^x \frac{1}{B(\alpha, \beta)} y^{\alpha-1} (1-y)^{\beta-1} dy. \quad (2.46)$$

The integral in Equation 2.46 is called the incomplete beta function. This can be calculated in MATLAB using the function `betainc(x, alpha, beta)`.

### EXAMPLE 2.9

We use the following MATLAB code to plot the beta density over the interval (0,1). We let  $\alpha = \beta = 0.5$  and  $\alpha = \beta = 3$ .

```
% First get the domain over which to evaluate
% the density function.
x = 0.01:.01:.99;
% Now get the values for the density function.
y1 = betapdf(x,0.5,0.5);
y2 = betapdf(x,3,3);
% Plot the results.
plot(x,y1,'r',x,y2,'g')
title('Beta Distribution')
xlabel('x')
ylabel('f(x)')
```

The resulting curves are shown in [Figure 2.9](#). You can use the MATLAB Statistics Toolbox function `betapdf(x, alpha, beta)`, as we did in the example, or the function `csbetap(x, alpha, beta)`.

□

## Multivariate Normal

So far, we have discussed several univariate distributions for discrete and continuous random variables. In this section, we describe one of the important and most commonly used multivariate densities: the multivariate normal distribution. This distribution is used throughout the rest of the text. Some examples of where we use it are in exploratory data analysis, in probability density estimation, and in statistical pattern recognition.

The probability density function for a general multivariate normal density for  $d$  dimensions is given by

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad (2.47)$$

where  $\mathbf{x}$  is a  $d$ -component column vector,  $\boldsymbol{\mu}$  is the  $d \times 1$  column vector of means, and  $\boldsymbol{\Sigma}$  is the  $d \times d$  covariance matrix. The superscript  $T$  represents the

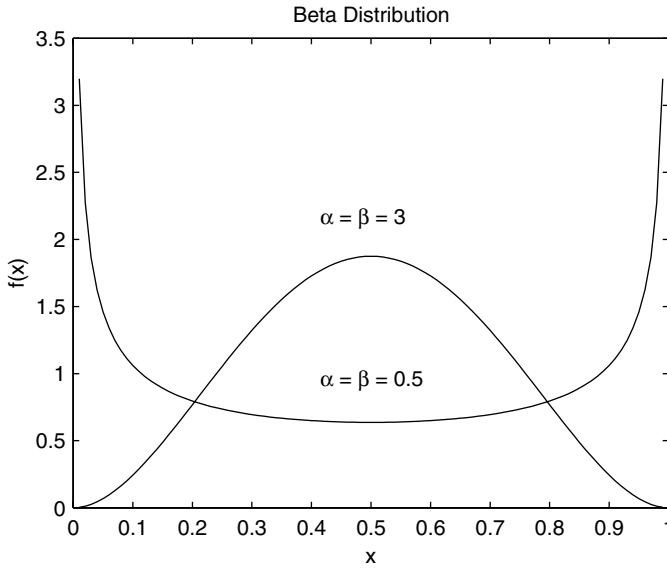


FIGURE 2.9.  
Beta probability density functions for various parameters.

transpose of an array, and the notation  $||$  denotes the determinant of a matrix.

The mean and covariance are calculated using the following formulas:

$$\boldsymbol{\mu} = E[\mathbf{x}] , \quad (2.48)$$

and

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] , \quad (2.49)$$

where the expected value of an array is given by the expected values of its components. Thus, if we let  $X_i$  represent the  $i$ -th component of  $\mathbf{x}$  and  $\mu_i$  the  $i$ -th component of  $\boldsymbol{\mu}$ , then the elements of Equation 2.48 can be written as

$$\mu_i = E[X_i] .$$

If  $\sigma_{ij}$  represents the  $ij$ -th element of  $\boldsymbol{\Sigma}$ , then the elements of the covariance matrix (Equation 2.49) are given by

$$\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] .$$



The covariance matrix is symmetric ( $\Sigma^T = \Sigma$ ) positive definite (all eigenvalues of  $\Sigma$  are greater than zero) for most applications of interest to statisticians and engineers.

We illustrate some properties of the multivariate normal by looking at the bivariate ( $d = 2$ ) case. The probability density function for a bivariate normal is represented by a bell-shaped surface. The center of the surface is determined by the mean  $\mu$  and the shape of the surface is determined by the covariance  $\Sigma$ . If the covariance matrix is diagonal (all of the off-diagonal elements are zero), and the diagonal elements are equal, then the shape is circular. If the diagonal elements are not equal, then we get an ellipse with the major axis vertical or horizontal. If the covariance matrix is not diagonal, then the shape is elliptical with the axes at an angle. Some of these possibilities are illustrated in the next example.

### Example 2.10

We first provide the following MATLAB function to calculate the multivariate normal probability density function and illustrate its use in the bivariate case. The function is called `csevalnorm`, and it takes input arguments `x`, `mu`, `cov_mat`. The input argument `x` is a matrix containing the points in the domain where the function is to be evaluated, `mu` is a  $d$ -dimensional row vector, and `cov_mat` is the  $d \times d$  covariance matrix.

```
function prob = csevalnorm(x,mu,cov_mat);
[n,d] = size(x);
% center the data points
x = x-ones(n,1)*mu;
a = (2*pi)^(d/2)*sqrt(det(cov_mat));
arg = diag(x*inv(cov_mat)*x');
prob = exp((-0.5)*arg);
prob = prob/a;
```

We now call this function for a bivariate normal centered at zero and covariance matrix equal to the identity matrix. The density surface for this case is shown in [Figure 2.10](#).

```
% Get the mean and covariance.
mu = zeros(1,2);
cov_mat = eye(2); % Identity matrix
% Get the domain.
% Should range (-4,4) in both directions.
[x,y] = meshgrid(-4:.2:4,-4:.2:4);
% Reshape into the proper format for the function.
X = [x(:),y(:)];
Z = csevalnorm(X,mu,cov_mat);
% Now reshape the matrix for plotting.
z = reshape(Z,size(x));
subplot(1,2,1) % plot the surface
```

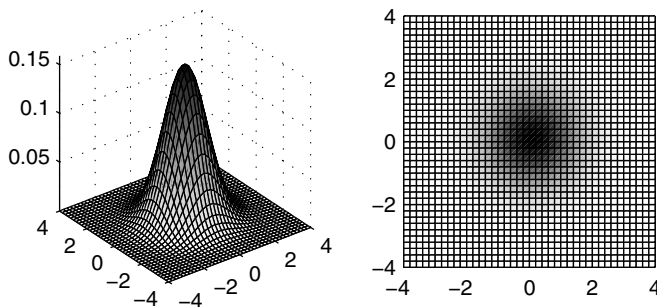


FIGURE 2.10

This figure shows a standard bivariate normal probability density function that is centered at the origin. The covariance matrix is given by the identity matrix. Notice that the shape of the surface looks circular. The plot on the right is for a viewpoint looking down on the surface.

```
surf(x,y,z),axis square, axis tight
title('BIVARIATE STANDARD NORMAL')
```

Next, we plot the surface for a bivariate normal centered at the origin with non-zero off-diagonal elements in the covariance matrix. Note the elliptical shape of the surface shown in [Figure 2.11](#).

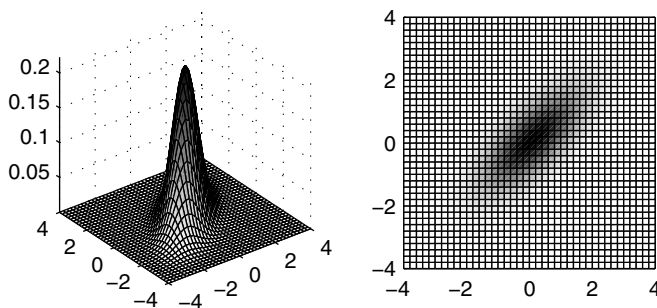


FIGURE 2.11

This shows a bivariate normal density where the covariance matrix has non-zero off-diagonal elements. Note that the surface has an elliptical shape. The plot on the right is for a viewpoint looking down on the surface.

```

subplot(1,2,2) % look down on the surface
pcolor(x,y,z),axis square
title('BIVARIATE STANDARD NORMAL')
% Now do the same thing for a covariance matrix
% with non-zero off-diagonal elements.
cov_mat = [1 0.7 ; 0.7 1];
Z = csevalnorm(X,mu,cov_mat);
z = reshape(Z,size(x));
subplot(1,2,1)
surf(x,y,z),axis square, axis tight
title('BIVARIATE NORMAL')
subplot(1,2,2)
pcolor(x,y,z),axis square
title('BIVARIATE NORMAL')

```

□

The probability that a point  $\mathbf{x} = (x_1, x_2)^T$  will assume a value in a region  $R$  can be found by integrating the bivariate probability density function over the region. Any plane that cuts the surface parallel to the  $x_1$ - $x_2$  plane intersects in an elliptic (or circular) curve, yielding a curve of constant density. Any plane perpendicular to the  $x_1$ - $x_2$  plane cuts the surface in a normal curve. This property indicates that in each dimension, the multivariate normal is a univariate normal distribution. This is discussed further in Chapter 5.

---

## 2.6 MATLAB Code

The MATLAB Statistics Toolbox has many functions for the more common distributions. It has functions for finding the value of the probability density (mass) function and the value of the cumulative distribution function. The reader is cautioned to remember that the definitions of the distributions (exponential, gamma, and Weibull) differ from what we describe in the text. For example, the exponential and the gamma distributions are parameterized differently in the MATLAB Statistics Toolbox. For a complete list of what is available in the toolbox for calculating probability density (mass) functions or cumulative distribution functions, see Appendix E.

The Computational Statistics Toolbox contains functions for several of the distributions, as defined in this chapter. In general, those functions that end in **p** correspond to the probability density (mass) function, and those ending with a **c** calculate the cumulative distribution function. [Table 2.1](#) provides a summary of the functions.

We note that a different function for evaluating the multivariate normal probability density function is available for download at

**TABLE 2.1**  
List of Functions from [Chapter 2](#) Included in the  
Computational Statistics Toolbox

Distribution	MATLAB Function
Beta	<code>csbetap</code> , <code>csbetac</code>
Binomial	<code>csbinop</code> , <code>csbinoc</code>
Chi-square	<code>cship</code> , <code>cschic</code>
Exponential	<code>csexpop</code> , <code>csexpoc</code>
Gamma	<code>csgammp</code> , <code>csgammc</code>
Normal - univariate	<code>csnormp</code> , <code>csnormc</code>
Normal - multivariate	<code>csevalnorm</code>
Poisson	<code>cspois</code> , <code>cspois</code>
Continuous Uniform	<code>csunifp</code> , <code>csunifc</code>
Weibull	<code>csweibp</code> , <code>csweibc</code>

<ftp://ftp.mathworks.com/pub/mathworks/>

under the `stats` directory. This function can be substituted for `csevalnorm`.



## 2.7 Further Reading

There are many excellent books on probability theory at the undergraduate and graduate levels. Ross [1994; 1997; 2000] is the author of several books on probability theory and simulation. These texts contain many examples and are appropriate for advanced undergraduate students in statistics, engineering and science. Rohatgi [1976] provides a solid theoretical introduction to probability theory. This text can be used by advanced undergraduate and beginning graduate students. It has recently been updated with many new examples and special topics [Rohatgi and Saleh, 2000]. For those who want to learn about probability, but do not want to be overwhelmed with the theory, then we recommend Durrett [1994].

At the graduate level, there is a book by Billingsley [1995] on probability and measure theory. He uses probability to motivate measure theory and then uses measure theory to generate more probability concepts. Another good reference is a text on probability and real analysis by Ash [1972]. This is suitable for graduate students in mathematics and statistics. For a book that can be used by graduate students in mathematics, statistics and engineering, see Port [1994]. This text provides a comprehensive treatment of the subject and can also be used as a reference by professional data analysts. Finally, Breiman [1992] provides an overview of probability theory that is accessible to statisticians and engineers.

## Exercises

- 2.1. Write a function using MATLAB's functions for numerical integration such as **quad** or **quadl** (MATLAB 6) that will find  $P(X \leq x)$  when the random variable is exponentially distributed with parameter  $\lambda$ . See **help** for information on how to use these functions.
- 2.2. Verify that the exponential probability density function with parameter  $\lambda$  integrates to 1. Use the MATLAB functions **quad** or **quadl** (MATLAB 6). See **help** for information on how to use these functions.
- 2.3. Radar and missile detection systems warn of enemy attacks. Suppose that a radar detection system has a probability 0.95 of detecting a missile attack.
  - a. What is the probability that one detection system will detect an attack? What distribution did you use?
  - b. Suppose three detection systems are located together in the same area and the operation of each system is independent of the others. What is the probability that at least one of the systems will detect the attack? What distribution did you use in this case?
- 2.4. When a random variable is equally likely to be either positive or negative, then the Laplacian or the double exponential distribution can be used to model it. The Laplacian probability density function for  $\lambda > 0$  is given by

$$f(x) = \frac{1}{2} \lambda e^{-\lambda|x|}; \quad -\infty < x < \infty.$$

- a. Derive the cumulative distribution function for the Laplacian.
  - b. Write a MATLAB function that will evaluate the Laplacian probability density function for given values in the domain.
  - c. Write a MATLAB function that will evaluate the Laplacian cumulative distribution function.
  - d. Plot the probability density function when  $\lambda = 1$ .
- 2.5. Suppose  $X$  follows the exponential distribution with parameter  $\lambda$ . Show that for  $s \geq 0$  and  $t \geq 0$ ,

$$P(X > s + t | X > s) = P(X > t).$$

- 2.6. The lifetime in years of a flat panel display is a random variable with the exponential probability density function given by

$$f(x;0.1) = 0.1e^{-0.1x}.$$

- a. What is the mean lifetime of the flat panel display?
  - b. What is the probability that the display fails within the first two years?
  - c. Given that the display has been operating for one year, what is the probability that it will fail within the next year?
- 2.7. The time to failure for a widget follows a Weibull distribution, with  $v = 0$ ,  $\beta = 1/2$ , and  $\alpha = 750$  hours.
- a. What is the mean time to failure of the widget?
  - b. What percentage of the widgets will fail by 2500 hours of operation? That is, what is the probability that a widget will fail within 2500 hours?
- 2.8. Let's say the probability of having a boy is 0.52. Using the Multiplication Rule, find the probability that a family's first and second children are boys. What is the probability that the first child is a boy and the second child is a girl?
- 2.9. Repeat Example 2.1 for  $n = 6$  and  $p = 0.5$ . What is the shape of the distribution?
- 2.10. Recall that in our piston ring example,  $P(M_A) = 0.6$  and  $P(M_B) = 0.4$ . From prior experience with the two manufacturers, we know that 2% of the parts supplied by manufacturer A are likely to fail and 6% of the parts supplied by manufacturer B are likely to fail. Thus,  $P(F|M_A) = 0.02$  and  $P(F|M_B) = 0.06$ . If we observe a piston ring failure, what is the probability that it came from manufacturer A?
- 2.11. Using the functions **fminbnd** or **fmin** (available in the standard MATLAB package), find the value for  $x$  where the maximum of the  $N(3, 1)$  probability density occurs. Note that you have to find the minimum of  $-f(x)$  to find the maximum of  $f(x)$  using these functions. Refer to the **help** files on these functions for more information on how to use them.
- 2.12. Using **normpdf** or **csnormp**, find the value of the probability density for  $N(0, 1)$  at  $\pm\infty$ . Use a small (large) value of  $x$  for  $-\infty$  ( $\infty$ ).
- 2.13. Verify Equation 2.38 using the MATLAB functions **factorial** and **gamma**.
- 2.14. Find the height of the curve for a normal probability density function at  $x = \mu$ , where  $\sigma = 0.5, 1, 2$ . What happens to the height of the curve as  $\sigma$  gets larger? Does the height change for different values of  $\mu$ ?
- 2.15. Write a function that calculates the Bayes' posterior probability given a vector of conditional probabilities and a vector of prior probabilities.

- 2.16. Compare the Poisson approximation to the actual binomial probability  $P(X = 4)$ , using  $n = 9$  and  $p = 0.1, 0.2, \dots, 0.9$ .
- 2.17. Using the function **normspec**, find the probability that the random variable defined in Example 2.5 assumes a value that is less than 3. What is the probability that the same random variable assumes a value that is greater than 5? Find these probabilities again using the function **normcdf**.
- 2.18. Find the probability for the Weibull random variable of Example 2.8 using the MATLAB Statistics Toolbox function **weibcdf** or the Computational Statistics Toolbox function **csweibc**.
- 2.19. The MATLAB Statistics Toolbox has a GUI demo called **disttool**. First view the **help** file on **disttool**. Then run the demo. Examine the probability density (mass) and cumulative distribution functions for the distributions discussed in the chapter.