5

Probability: The Mathematics of Randomness

The theory of probability is the mathematician's description of random variation. This chapter introduces enough probability to serve as a minimum background for making formal statistical inferences.

The chapter begins with a discussion of discrete random variables and their distributions. It next turns to continuous random variables and then probability plotting. Next, the simultaneous modeling of several random variables and the notion of independence are considered. Finally, there is a look at random variables that arise as functions of several others, and how randomness of the input variables is translated to the output variable.

5.1 (Discrete) Random Variables

The concept of a random (or chance) variable is introduced in general terms in this section. Then specialization to discrete cases is considered. The specification of a discrete probability distribution via a probability function or cumulative probability function is discussed. Next, summarization of discrete distributions in terms of (theoretical) means and variances is treated. Then the so-called binomial, geometric, and Poisson distributions are introduced as examples of useful discrete probability models.

5.1.1 Random Variables and Their Distributions

It is usually appropriate to think of a data value as subject to chance influences. In enumerative contexts, chance is commonly introduced into the data collection process through random sampling techniques. Measurement error is nearly always a factor in statistical engineering studies, and the many small, unnameable causes that work to produce it are conveniently thought of as chance phenomena. In analytical contexts, changes in system conditions work to make measured responses vary, and this is most often attributed to chance.

Definition 1

A **random variable** is a quantity that (prior to observation) can be thought of as dependent on chance phenomena. Capital letters near the end of the alphabet are typically used to stand for random variables.

Consider a situation (like that of Example 3 in Chapter 3) where the torques of bolts securing a machine component face plate are to be measured. The next measured value can be considered subject to chance influences and we thus term

$$Z = \text{the next torque recorded}$$

a random variable.

Following Definition 9 in Chapter 1, a distinction was made between discrete and continuous data. That terminology carries over to the present context and inspires two more definitions.

Definition 2

A **discrete random variable** is one that has isolated or separated possible values (rather than a continuum of available outcomes).

Definition 3

A **continuous random variable** is one that can be idealized as having an entire (continuous) interval of numbers as its set of possible values.

Random variables that are basically count variables clearly fall under Definition 2 and are discrete. It could be argued that all measurement variables are discrete—on the basis that all measurements are "to the nearest unit." But it is often mathematically convenient, and adequate for practical purposes, to treat them as continuous.

A random variable is, to some extent, a priori unpredictable. Therefore, in describing or modeling it, the important thing is to specify its set of potential values and the likelihoods associated with those possible values.

Definition 4

To specify a **probability distribution** for a random variable is to give its set of possible values and (in one way or another) consistently assign numbers between 0 and 1—called **probabilities**—as measures of the likelihood that the various numerical values will occur.

The methods used to specify discrete probability distributions are different from those used to specify continuous probability distributions. So the implications of Definition 4 are studied in two steps, beginning in this section with discrete distributions.

5.1.2 Discrete Probability Functions and Cumulative Probability Functions

The tool most often used to describe a discrete probability distribution is the **probability function**.

Definition 5

A **probability function** for a discrete random variable X, having possible values $x_1, x_2, \ldots$, is a nonnegative function $f(x)$, with $f(x_i)$ giving the probability that X takes the value x_i.

This text will use the notational convention that a capital P followed by an expression or phrase enclosed by brackets will be read "the probability" of that expression. In these terms, a probability function for X is a function f such that

Probability function for the discrete random variable X

$$f(x) = P[X = x]$$

That is, "$f(x)$ is the probability that (the random variable) X takes the value x."

Example 1

A Torque Requirement Random Variable

Consider again Example 3 in Chapter 3, where Brenny, Christensen, and Schneider measured bolt torques on the face plates of a heavy equipment component. With

$$Z = \text{the next measured torque for bolt 3 (recorded to the nearest integer)}$$

consider treating Z as a discrete random variable and giving a plausible probability function for it.

The relative frequencies for the bolt 3 torque measurements recorded in Table 3.4 on page 74 produce the relative frequency distribution in Table 5.1. This table shows, for example, that over the period the students were collecting data, about 15% of measured torques were 19 ft lb. If it is sensible to believe that the same system of causes that produced the data in Table 3.4 will operate

Example 1 *(continued)*

to produce the next bolt 3 torque, then it also makes sense to base a probability function for Z on the relative frequencies in Table 5.1. That is, the probability distribution specified in Table 5.2 might be used. (In going from the relative frequencies in Table 5.1 to proposed values for $f(z)$ in Table 5.2, there has been some slightly arbitrary rounding. This has been done so that probability values are expressed to two decimal places and now total to exactly 1.00.)

Table 5.1
Relative Frequency Distribution for Measured Bolt 3 Torques

z, Torque (ft lb)	Frequency	Relative Frequency
11	1	$1/34 \approx .02941$
12	1	$1/34 \approx .02941$
13	1	$1/34 \approx .02941$
14	2	$2/34 \approx .05882$
15	9	$9/34 \approx .26471$
16	3	$3/34 \approx .08824$
17	4	$4/34 \approx .11765$
18	7	$7/34 \approx .20588$
19	5	$5/34 \approx .14706$
20	1	$1/34 \approx .02941$
	34	1

Table 5.2
A Probability Function for Z

Torque z	Probability $f(z)$
11	.03
12	.03
13	.03
14	.06
15	.26
16	.09
17	.12
18	.20
19	.15
20	.03

The appropriateness of the probability function in Table 5.2 for describing Z depends essentially on the physical stability of the bolt-tightening process. But there is a second way in which relative frequencies can become obvious choices for probabilities. For example, think of treating the 34 torques represented in Table 5.1 as a population, from which $n = 1$ item is to be sampled at random, and

$$Y = \text{the torque value selected}$$

Then the probability function in Table 5.2 is also approximately appropriate for Y. This point is not so important in this specific example as it is in general: Where one value is to be selected at random from a population, an appropriate probability distribution is one that is equivalent to the population relative frequency distribution.

The probability distribution of a single value selected at random from a population

This text will usually express probabilities to two decimal places, as in Table 5.2. Computations may be carried to several more decimal places, but final probabilities will typically be reported only to two places. This is because numbers expressed to more than two places tend to look too impressive and be taken too seriously by the uninitiated. Consider for example the statement "There is a .097328 probability of booster engine failure" at a certain missile launch. This may represent the results of some very careful mathematical manipulations and be correct to six decimal places *in the context of the mathematical model used to obtain the value*. But it is doubtful that the model used is a good enough description of physical reality to warrant that much apparent precision. Two-decimal precision is about what is warranted in most engineering applications of simple probability.

Properties of a mathematically valid probability function

The probability function shown in Table 5.2 has two properties that are necessary for the mathematical consistency of a discrete probability distribution. The $f(z)$ values are each in the interval $[0, 1]$ and they total to 1. Negative probabilities or ones larger than 1 would make no practical sense. A probability of 1 is taken as indicating certainty of occurrence and a probability of 0 as indicating certainty of nonoccurrence. Thus, according to the model specified in Table 5.2, since the values of $f(z)$ sum to 1, the occurrence of one of the values 11, 12, 13, 14, 15, 16, 17, 18, 19, and 20 ft lb is certain.

A probability function $f(x)$ gives probabilities of occurrence for individual values. Adding the appropriate values gives probabilities associated with the occurrence of one of a specified type of value for X.

Example 1 *(continued)*

Consider using $f(z)$ defined in Table 5.2 to find

$$P[Z > 17] = P[\text{the next torque exceeds 17}]$$

Adding the $f(z)$ entries corresponding to possible values larger than 17 ft lb,

$$P[Z > 17] = f(18) + f(19) + f(20) = .20 + .15 + .03 = .38$$

The likelihood of the next torque being more than 17 ft lb is about 38%.

Example 1 (*continued*)

If, for example, specifications for torques were 16 ft lb to 21 ft lb, then the likelihood that the next torque measured will be within specifications is

$$\begin{aligned} P[16 \le Z \le 21] &= f(16) + f(17) + f(18) + f(19) + f(20) + f(21) \\ &= .09 + .12 + .20 + .15 + .03 + .00 \\ &= .59 \end{aligned}$$

In the torque measurement example, the probability function is given in tabular form. In other cases, it is possible to give a formula for $f(x)$.

Example 2

A Random Tool Serial Number

The last step of the pneumatic tool assembly process studied by Kraber, Rucker, and Williams (see Example 11 in Chapter 3) was to apply a serial number plate to the completed tool. Imagine going to the end of the assembly line at exactly 9:00 A.M. next Monday and observing the number plate first applied after 9:00. Suppose that

$$W = \text{the last digit of the serial number observed}$$

Suppose further that tool serial numbers begin with some code special to the tool model and end with consecutively assigned numbers reflecting how many tools of the particular model have been produced. The symmetry of this situation suggests that each possible value of W ($w = 0, 1, \ldots, 9$) is equally likely. That is, a plausible probability function for W is given by the formula

$$f(w) = \begin{cases} .1 & \text{for } w = 0, 1, 2, \ldots, 9 \\ 0 & \text{otherwise} \end{cases}$$

Another way of specifying a discrete probability distribution is sometimes used. That is to specify its **cumulative probability function**.

Definition 6

The **cumulative probability function** for a random variable X is a function $F(x)$ that for each number x gives the probability that X takes that value or a smaller one. In symbols,

$$\boxed{F(x) = P[X \le x]}$$

Since (for discrete distributions) probabilities are calculated by summing values of $f(x)$, for a discrete distribution,

Cumulative probability function for a discrete variable X

$$F(x) = \sum_{z \leq x} f(z)$$

(The sum is over possible values less than or equal to x.) In this discrete case, the graph of $F(x)$ will be a stair-step graph with jumps located at possible values and equal in size to the probabilities associated with those possible values.

Example 1 (*continued*)

Values of both the probability function and the cumulative probability function for the torque variable Z are given in Table 5.3. Values of $F(z)$ for other z are also easily obtained. For example,

$$F(10.7) = P[Z \leq 10.7] = 0$$
$$F(16.3) = P[Z \leq 16.3] = P[Z \leq 16] = F(16) = .50$$
$$F(32) = P[Z \leq 32] = 1.00$$

A graph of the cumulative probability function for Z is given in Figure 5.1. It shows the stair-step shape characteristic of cumulative probability functions for discrete distributions.

Table 5.3
Values of the Probability Function and Cumulative Probability Function for Z

z, Torque	$f(z) = P[Z = z]$	$F(z) = P[Z \leq z]$
11	.03	.03
12	.03	.06
13	.03	.09
14	.06	.15
15	.26	.41
16	.09	.50
17	.12	.62
18	.20	.82
19	.15	.97
20	.03	1.00

Example 1 (*continued*)

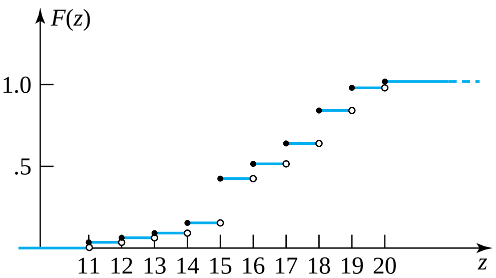


Figure 5.1 Graph of the cumulative probability function for Z

The information about a discrete distribution carried by its cumulative probability function is equivalent to that carried by the corresponding probability function. The cumulative version is sometimes preferred for table making, because round-off problems are more severe when adding several $f(x)$ terms than when taking the difference of two $F(x)$ values to get a probability associated with a consecutive sequence of possible values.

5.1.3 Summarization of Discrete Probability Distributions

Amost all of the devices for describing relative frequency (empirical) distributions in Chapter 3 have versions that can describe (theoretical) probability distributions.

For a discrete random variable with equally spaced possible values, a **probability histogram** gives a picture of the shape of the variable's distribution. It is made by centering a bar of height $f(x)$ over each possible value x. Probability histograms for the random variables Z and W in Examples 1 and 2 are given in Figure 5.2. Interpreting such probability histograms is similar to interpreting relative frequency histograms, except that the areas on them represent (theoretical) probabilities instead of (empirical) fractions of data sets.

It is useful to have a notion of mean value for a discrete random variable (or its probability distribution).

Definition 7

The **mean** or **expected value** of a discrete random variable X (sometimes called the mean of its probability distribution) is

$$EX = \sum_x x\, f(x) \tag{5.1}$$

EX is read as "the expected value of X," and sometimes the notation μ is used in place of EX.

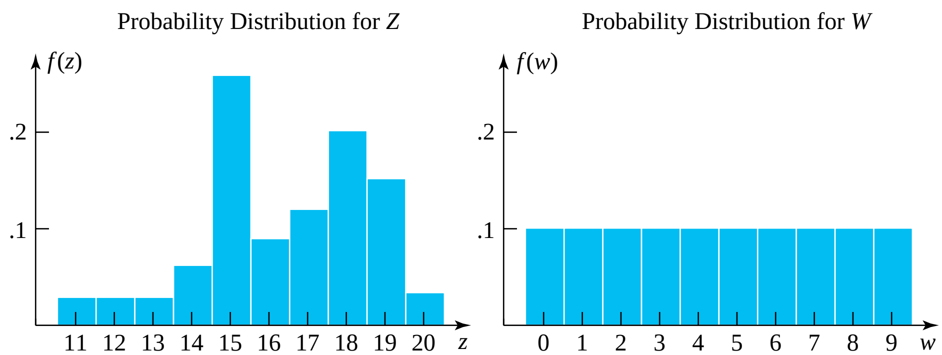


Figure 5.2 Probability histograms for Z and W (*Examples 1 and 2*)

(Remember the warning in Section 3.3 that μ would stand for both the mean of a population and the mean of a probability distribution.)

Example 1 (*continued*)

Returning to the bolt torque example, the expected (or theoretical mean) value of the next torque is

$$\begin{aligned} EZ &= \sum_z z\,f(z) \\ &= 11(.03) + 12(.03) + 13(.03) + 14(.06) + 15(.26) \\ &\quad + 16(.09) + 17(.12) + 18(.20) + 19(.15) + 20(.03) \\ &= 16.35 \text{ ft lb} \end{aligned}$$

This value is essentially the arithmetic mean of the bolt 3 torques listed in Table 3.4. (The slight disagreement in the third decimal place arises only because the relative frequencies in Table 5.1 were rounded slightly to produce Table 5.2.) This kind of agreement provides motivation for using the symbol μ, first seen in Section 3.3, as an alternative to EZ.

The mean of a discrete probability distribution has a balance point interpretation, much like that associated with the arithmetic mean of a data set. Placing (point) masses of sizes $f(x)$ at points x along a number line, EX is the center of mass of that distribution.

Example 2 *(continued)*

Considering again the serial number example, and the second part of Figure 5.2, if a balance point interpretation of expected value is to hold, EW had better turn out to be 4.5. And indeed,

$$EW = 0(.1) + 1(.1) + 2(.1) + \cdots + 8(.1) + 9(.1) = 45(.1) = 4.5$$

It was convenient to measure the spread of a data set (or its relative frequency distribution) with the variance and standard deviation. It is similarly useful to have notions of spread for a discrete probability distribution.

Definition 8

The **variance** of a discrete random variable X (or the variance of its distribution) is

$$\text{Var}\, X = \sum (x - EX)^2 f(x) \quad \left(= \sum x^2 f(x) - (EX)^2\right) \tag{5.2}$$

The **standard deviation** of X is $\sqrt{\text{Var}\, X}$. Often the notation σ^2 is used in place of Var X, and σ is used in place of $\sqrt{\text{Var}\, X}$.

The variance of a random variable is its expected (or mean) squared distance from the center of its probability distribution. The use of σ^2 to stand for both the variance of a population and the variance of a probability distribution is motivated on the same grounds as the double use of μ.

Example 1 *(continued)*

The calculations necessary to produce the bolt torque standard deviation are organized in Table 5.4. So

$$\sigma = \sqrt{\text{Var}\, Z} = \sqrt{4.6275} = 2.15 \text{ ft lb}$$

Except for a small difference due to round-off associated with the creation of Table 5.2, this standard deviation of the random variable Z is numerically the same as the population standard deviation associated with the bolt 3 torques in Table 3.4. (Again, this is consistent with the equivalence between the population relative frequency distribution and the probability distribution for Z.)

Table 5.4
Calculations for Var Z

z	$f(z)$	$(z-16.35)^2$	$(z-16.35)^2 f(z)$
11	.03	28.6225	.8587
12	.03	18.9225	.5677
13	.03	11.2225	.3367
14	.06	5.5225	.3314
15	.26	1.8225	.4739
16	.09	.1225	.0110
17	.12	.4225	.0507
18	.20	2.7225	.5445
19	.15	7.0225	1.0534
20	.03	13.3225	.3997
			Var $Z = 4.6275$

Example 2 *(continued)*

To illustrate the alternative for calculating a variance given in Definition 8, consider finding the variance and standard deviation of the serial number variable W. Table 5.5 shows the calculation of $\sum w^2 f(w)$.

Table 5.5
Calculations for $\sum w^2 f(w)$

w	$f(w)$	$w^2 f(w)$
0	.1	0.0
1	.1	.1
2	.1	.4
3	.1	.9
4	.1	1.6
5	.1	2.5
6	.1	3.6
7	.1	4.9
8	.1	6.4
9	.1	8.1
		28.5

Example 2 *(continued)*

Then

$$\text{Var}\, W = \sum w^2 f(w) - (EW)^2 = 28.5 - (4.5)^2 = 8.25$$

so that

$$\sqrt{\text{Var}\, W} = 2.87$$

Comparing the two probability histograms in Figure 5.2, notice that the distribution of W appears to be more spread out than that of Z. Happily, this is reflected in the fact that

$$\sqrt{\text{Var}\, W} = 2.87 > 2.15 = \sqrt{\text{Var}\, Z}$$

5.1.4 The Binomial and Geometric Distributions

Discrete probability distributions are sometimes developed from past experience with a particular physical phenomenon (as in Example 1). On the other hand, sometimes an easily manipulated set of mathematical assumptions having the potential to describe a variety of real situations can be put together. When those can be manipulated to derive generic distributions, those distributions can be used to model a number of different random phenomena. One such set of assumptions is that of **independent, identical success-failure trials**.

Independent identical success-failure trials

Many engineering situations involve repetitions of essentially the same "go–no go" (success-failure) scenario, where:

1. There is a *constant chance of a go/success outcome* on each repetition of the scenario (call this probability p).
2. The repetitions are *independent* in the sense that knowing the outcome of any one of them does not change assessments of chance related to any others.

Examples of this kind include the testing of items manufactured consecutively, where each will be classified as either conforming or nonconforming; observing motorists as they pass a traffic checkpoint and noting whether each is traveling at a legal speed or speeding; and measuring the performance of workers in two different workspace configurations and noting whether the performance of each is better in configuration A or configuration B.

In this context, there are two generic kinds of random variables for which deriving appropriate probability distributions is straightforward. The first is the case of a count of the repetitions out of n that yield a go/success result. That is, consider a variable

Binomial random variables

$$X = \text{the number of go/success results in } n \text{ independent identical success-failure trials}$$

X has the **binomial** (n, p) **distribution**.

Definition 9

The **binomial** (n, p) **distribution** is a discrete probability distribution with probability function

$$f(x) = \begin{cases} \dfrac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x} & \text{for } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{(5.3)}$$

for n a positive integer and $0 < p < 1$.

Equation (5.3) is completely plausible. In it there is one factor of p for each trial producing a go/success outcome and one factor of $(1 - p)$ for each trial producing a no go/failure outcome. And the $n!/x!\,(n - x)!$ term is a count of the number of patterns in which it would be possible to see x go/success outcomes in n trials. The name *binomial* distribution derives from the fact that the values $f(0), f(1), f(2), \ldots, f(n)$ are the terms in the expansion of

$$(p + (1 - p))^n$$

according to the binomial theorem.

Take the time to plot probability histograms for several different binomial distributions. It turns out that for $p < .5$, the resulting histogram is right-skewed. For $p > .5$, the resulting histogram is left-skewed. The skewness increases as p moves away from .5, and it decreases as n is increased. Four binomial probability histograms are pictured in Figure 5.3.

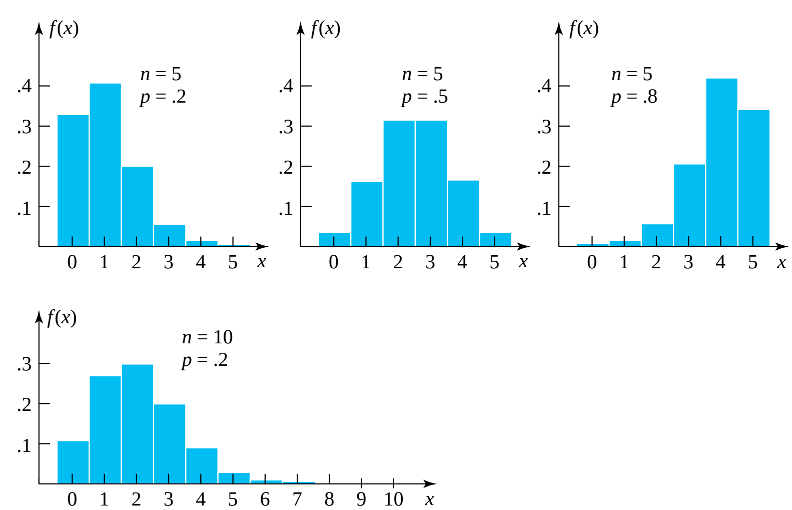


Figure 5.3 Four binomial probability histograms

Example 3

The Binomial Distribution and Counts of Reworkable Shafts

Consider again the situation of Example 12 in Chapter 3 and a study of the performance of a process for turning steel shafts. Early in that study, around 20% of the shafts were typically classified as "reworkable." Suppose that $p = .2$ is indeed a sensible figure for the chance that a given shaft will be reworkable. Suppose further that $n = 10$ shafts will be inspected, and the probability that at least two are classified as reworkable is to be evaluated.

Adopting a model of independent, identical success-failure trials for shaft conditions,

$$U = \text{the number of reworkable shafts in the sample of 10}$$

is a binomial random variable with $n = 10$ and $p = .2$. So

$$\begin{aligned} P[\text{at least two reworkable shafts}] &= P[U \geq 2] \\ &= f(2) + f(3) + \cdots + f(10) \\ &= 1 - (f(0) + f(1)) \\ &= 1 - \left(\frac{10!}{0!\,10!}(.2)^0(.8)^{10} + \frac{10!}{1!\,9!}(.2)^1(.8)^9 \right) \\ &= .62 \end{aligned}$$

(The trick employed here, to avoid plugging into the binomial probability function 9 times by recognizing that the $f(u)$'s have to sum up to 1, is a common and useful one.)

The .62 figure is only as good as the model assumptions that produced it. If an independent, identical success-failure trials description of shaft production fails to accurately portray physical reality, the .62 value is fine mathematics but possibly a poor description of what will actually happen. For instance, say that due to tool wear it is typical to see 40 shafts in specifications, then 10 reworkable shafts, a tool change, 40 shafts in specifications, and so on. In this case, the binomial distribution would be a very poor description of U, and the .62 figure largely irrelevant. (The independence-of-trials assumption would be inappropriate in this situation.)

The binomial distribution and simple random sampling

There is one important circumstance where a model of independent, identical success-failure trials is not exactly appropriate, but a binomial distribution can still be adequate for practical purposes—that is, in describing the results of simple random sampling from a dichotomous population. Suppose a population of size N contains a fraction p of type A objects and a fraction $(1 - p)$ of type B objects. If a simple random sample of n of these items is selected and

$$X = \text{the number of type A items in the sample}$$

strictly speaking, x is not a binomial random variable. But if n is a small fraction of N (say, less than 10%), and p is not too extreme (i.e., is not close to either 0 or 1), X is *approximately* binomial (n, p).

Example 4

Simple Random Sampling from a Lot of Hexamine Pellets

In the pelletizing machine experiment described in Example 14 in Chapter 3, Greiner, Grimm, Larson, and Lukomski found a combination of machine settings that allowed them to produce 66 conforming pellets out of a batch of 100 pellets. Treat that batch of 100 pellets as a population of interest and consider selecting a simple random sample of size $n = 2$ from it.

If one defines the random variable

$$V = \text{the number of conforming pellets in the sample of size 2}$$

the most natural probability distribution for V is obtained as follows. Possible values for V are 0, 1, and 2.

$$\begin{aligned}
f(0) &= P[V = 0] \\
&= P[\text{first pellet selected is nonconforming and} \\
&\qquad \text{subsequently the second pellet is also nonconforming}] \\
f(2) &= P[V = 2] \\
&= P[\text{first pellet selected is conforming and} \\
&\qquad \text{subsequently the second pellet selected is conforming}] \\
f(1) &= 1 - (f(0) + f(2))
\end{aligned}$$

Then think, "In the long run, the first selection will yield a nonconforming pellet about 34 out of 100 times. Considering only cases where this occurs, in the long run the next selection will also yield a nonconforming pellet about 33 out of 99 times." That is, a sensible evaluation of $f(0)$ is

$$f(0) = \frac{34}{100} \cdot \frac{33}{99} = .1133$$

Example 4 (*continued*)

Similarly,

$$f(2) = \frac{66}{100} \cdot \frac{65}{99} = .4333$$

and thus

$$f(1) = 1 - (.1133 + .4333) = 1 - .5467 = .4533$$

Now, V cannot be thought of as arising from exactly independent trials. For example, knowing that the first pellet selected was conforming would reduce most people's assessment of the chance that the second is also conforming from $\frac{66}{100}$ to $\frac{65}{99}$. Nevertheless, for most practical purposes, V can be thought of as *essentially* binomial with $n = 2$ and $p = .66$. To see this, note that

$$\frac{2!}{0!\,2!}(.34)^2(.66)^0 = .1156 \approx f(0)$$

$$\frac{2!}{1!\,1!}(.34)^1(.66)^1 = .4488 \approx f(1)$$

$$\frac{2!}{2!\,0!}(.34)^0(.66)^2 = .4356 \approx f(2)$$

Here, n is a small fraction of N, p is not too extreme, and a binomial distribution is a decent description of a variable arising from simple random sampling.

Calculation of the mean and variance for binomial random variables is greatly simplified by the fact that when the formulas (5.1) and (5.2) are used with the expression for binomial probabilities in equation (5.3), simple formulas result. For X a binomial (n, p) random variable,

Mean of the binomial (n, p) distribution

$$\mu = EX = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = np \qquad \textbf{(5.4)}$$

Further, it is the case that

Variance of the binomial (n, p) distribution

$$\sigma^2 = \text{Var}\, X = \sum_{x=0}^{n} (x - np)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = np(1-p) \qquad \textbf{(5.5)}$$

Example 3 (*continued*)

Returning to the machining of steel shafts, suppose that a binomial distribution with $n = 10$ and $p = .2$ is appropriate as a model for

$$U = \text{the number of reworkable shafts in the sample of 10}$$

Then, by formulas (5.4) and (5.5),

$$EU = (10)(.2) = 2 \text{ shafts}$$
$$\sqrt{\operatorname{Var} U} = \sqrt{10(.2)(.8)} = 1.26 \text{ shafts}$$

A second generic type of random variable associated with a series of independent, identical success-failure trials is

Geometric random variables

$$X = \text{the number of trials required to first obtain a go/success result}$$

X has the **geometric** (p) **distribution**.

Definition 10

The **geometric** (p) **distribution** is a discrete probability distribution with probability function

$$f(x) = \begin{cases} p(1-p)^{x-1} & \text{for } x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

for $0 < p < 1$.

Formula (5.6) makes good intuitive sense. In order for X to take the value x, there must be $x - 1$ consecutive no-go/failure results followed by a go/success. In formula (5.6), there are $x - 1$ terms $(1 - p)$ and one term p. Another way to see that formula (5.6) is plausible is to reason that for X as above and $x = 1, 2, \ldots$

$$\begin{aligned} 1 - F(x) &= 1 - P[X \le x] \\ &= P[X > x] \\ &= P[x \text{ no-go/failure outcomes in } x \text{ trials}] \end{aligned}$$

That is,

Simple relationship for the geometric (p) cumulative probability function

$$\boxed{1 - F(x) = (1-p)^x} \tag{5.7}$$

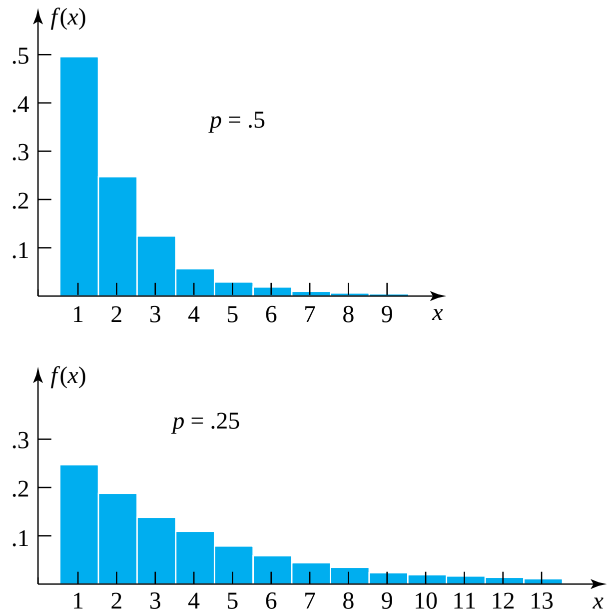


Figure 5.4 Two geometric probability histograms

by using the form of the binomial (x, p) probability function given in equation (5.3). Then for $x = 2, 3, \ldots,$ $f(x) = F(x) - F(x-1) = -(1 - F(x)) + (1 - F(x-1))$. This, combined with equation (5.7), gives equation (5.6).

The name *geometric* derives from the fact that the values $f(1), f(2), f(3), \ldots$ are terms in the geometric infinite series for

$$p \cdot \frac{1}{1-(1-p)}$$

The geometric distributions are discrete distributions with probability histograms exponentially decaying as x increases. Two different geometric probability histograms are pictured in Figure 5.4.

Example 5

The Geometric Distribution and Shorts in NiCad Batteries

In “A Case Study of the Use of an Experimental Design in Preventing Shorts in Nickel-Cadmium Cells” (*Journal of Quality Technology*, 1988), Ophir, El-Gad, and Snyder describe a series of experiments conducted in order to reduce the proportion of cells being scrapped by a battery plant because of internal shorts. The experimental program was successful in reducing the percentage of manufactured cells with internal shorts to around 1%.

Suppose that testing begins on a production run in this plant, and let

$$T = \text{the test number at which the first short is discovered}$$

A model for T (appropriate if the independent, identical success-failure trials description is apt) is geometric with $p = .01$. (p is the probability that any particular test yields a shorted cell.) Then, using equation (5.6),

$$\begin{aligned} P[\text{the first or second cell tested has the first short}] &= P[T = 1 \text{ or } T = 2] \\ &= f(1) + f(2) \\ &= (.01) + (.01)(1 - .01) \\ &= .02 \end{aligned}$$

Or, using equation (5.7),

$$\begin{aligned} P[\text{at least 50 cells are tested without finding a short}] &= P[T > 50] \\ &= (1 - .01)^{50} \\ &= .61 \end{aligned}$$

Like the binomial distributions, the geometric distributions have means and variances that are simple functions of the parameter p. That is, if X is geometric (p),

Mean of the geometric (p) distribution

$$\boxed{\mu = EX = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = \frac{1}{p}} \qquad \textbf{(5.8)}$$

and

Variance of the geometric (p) distribution

$$\boxed{\sigma^2 = \text{Var}\,X = \sum_{x=1}^{\infty} \left(x - \frac{1}{p}\right)^2 p(1-p)^{x-1} = \frac{1-p}{p^2}} \qquad \textbf{(5.9)}$$

Example 5 *(continued)*

In the context of battery testing, with T as before,

$$ET = \frac{1}{.01} = 100 \text{ batteries}$$

$$\sqrt{\text{Var}\,T} = \sqrt{\frac{(1 - .01)}{(.01)^2}} = 99.5 \text{ batteries}$$

Example 5 (*continued*)

Formula (5.8) is an intuitively appealing result. If there is only 1 chance in 100 of encountering a shorted battery at each test, it is sensible to expect to wait through 100 tests on average to encounter the first one.

5.1.5 The Poisson Distributions

As discussed in Section 3.4, it is often important to keep track of the total number of occurrences of some relatively rare phenomenon, where the physical or time unit under observation has the potential to produce many such occurrences. A case of floor tiles has potentially many total blemishes. In a one-second interval, there are potentially a large number of messages that can arrive for routing through a switching center. And a 1 cc sample of glass potentially contains a large number of imperfections.

So probability distributions are needed to describe random *counts of the number of occurrences of a relatively rare phenomenon across a specified interval of time or space*. By far the most commonly used theoretical distributions in this context are the **Poisson distributions**.

Definition 11

The **Poisson** (λ) **distribution** is a discrete probability distribution with probability function

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{(5.10)}$$

for $\lambda > 0$.

The form of equation (5.10) may initially seem unappealing. But it is one that has sensible mathematical origins, is manageable, and has proved itself empirically useful in many different "rare events" circumstances. One way to arrive at equation (5.10) is to think of a very large number of independent trials (opportunities for occurrence), where the probability of success (occurrence) on any one is very small and the product of the number of trials and the success probability is λ. One is then led to the binomial $(n, \frac{\lambda}{n})$ distribution. In fact, for large n, the binomial $(n, \frac{\lambda}{n})$ probability function approximates the one specified in equation (5.10). So one might think of the Poisson distribution for counts as arising through a mechanism that would present many tiny similar opportunities for independent occurrence or nonoccurrence throughout an interval of time or space.

The Poisson distributions are right-skewed distributions over the values $x = 0, 1, 2, \ldots,$ whose probability histograms peak near their respective λ's. Two different Poisson probability histograms are shown in Figure 5.5. λ is both the mean

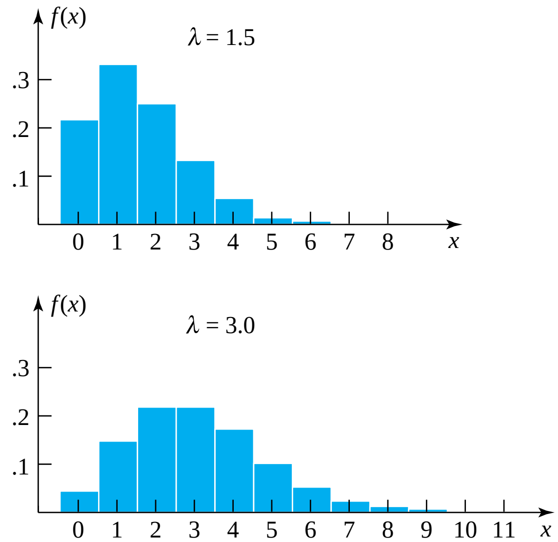


Figure 5.5 Two Poisson probability histograms

and the variance for the Poisson (λ) distribution. That is, if X has the Poisson (λ) distribution, then

Mean of the Poisson (λ) distribution

$$\mu = EX = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \tag{5.11}$$

and

Variance of the Poisson (λ) distribution

$$\operatorname{Var} X = \sum_{x=0}^{\infty} (x-\lambda)^2 \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \tag{5.12}$$

Fact (5.11) is helpful in picking out which Poisson distribution might be useful in describing a particular “rare events” situation.

Example 6

WWW

The Poisson Distribution and Counts of α-Particles

A classical data set of Rutherford and Geiger, reported in *Philosophical Magazine* in 1910, concerns the numbers of α-particles emitted from a small bar of polonium and colliding with a screen placed near the bar in 2,608 periods of 8 minutes each. The Rutherford and Geiger relative frequency distribution has mean 3.87 and a shape remarkably similar to that of the Poisson probability distribution with mean $\lambda = 3.87$.

Example 6 *(continued)*

In a duplication of the Rutherford/Geiger experiment, a reasonable probability function for describing

$S =$ the number of α-particles striking the screen in an additional 8-minute period

is then

$$f(s) = \begin{cases} \dfrac{e^{-3.87}(3.87)^s}{s!} & \text{for } s = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Using such a model, one has (for example)

$$\begin{aligned} &P[\text{at least 4 particles are recorded}] \\ &= P[S \geq 4] \\ &= f(4) + f(5) + f(6) + \cdots \\ &= 1 - (f(0) + f(1) + f(2) + f(3)) \\ &= 1 - \left(\frac{e^{-3.87}(3.87)^0}{0!} + \frac{e^{-3.87}(3.87)^1}{1!} + \frac{e^{-3.87}(3.87)^2}{2!} + \frac{e^{-3.87}(3.87)^3}{3!} \right) \\ &= .54 \end{aligned}$$

Example 7

Arrivals at a University Library

Stork, Wohlsdorf, and McArthur collected data on numbers of students entering the ISU library during various periods over a week's time. Their data indicate that between 12:00 and 12:10 P.M. on Monday through Wednesday, an average of around 125 students entered. Consider modeling

$M =$ the number of students entering the ISU library between 12:00 and 12:01 next Tuesday

Using a Poisson distribution to describe M, the reasonable choice of λ would seem to be

$$\lambda = \frac{125 \text{ students}}{10 \text{ minutes}}(1 \text{ minute}) = 12.5 \text{ students}$$

For this choice,

$$EM = \lambda = 12.5 \text{ students}$$

$$\sqrt{\text{Var}\, M} = \sqrt{\lambda} = \sqrt{12.5} = 3.54 \text{ students}$$

and, for example, the probability that between 10 and 15 students (inclusive) arrive at the library between 12:00 and 12:01 would be evaluated as

$$\begin{aligned}P[10 \le M \le 15] &= f(10) + f(11) + f(12) + f(13) + f(14) + f(15)\\ &= \frac{e^{-12.5}(12.5)^{10}}{10!} + \frac{e^{-12.5}(12.5)^{11}}{11!} + \frac{e^{-12.5}(12.5)^{12}}{12!}\\ &\quad + \frac{e^{-12.5}(12.5)^{13}}{13!} + \frac{e^{-12.5}(12.5)^{14}}{14!} + \frac{e^{-12.5}(12.5)^{15}}{15!}\\ &= .60\end{aligned}$$

Section 1 Exercises

1. A discrete random variable X can be described using the probability function

x	2	3	4	5	6
$f(x)$	.1	.2	.3	.3	.1

(a) Make a probability histogram for X. Also plot $F(x)$, the cumulative probability function for X.

(b) Find the mean and standard deviation of X.

2. In an experiment to evaluate a new artificial sweetener, ten subjects are all asked to taste cola from three unmarked glasses, two of which contain regular cola while the third contains cola made with the new sweetener. The subjects are asked to identify the glass whose content is different from the other two. If there is no difference between the taste of sugar and the taste of the new sweetener, the subjects would be just guessing.

(a) Make a table for a probability function for

$X =$ the number of subjects correctly identifying the artificially sweetened cola

under this hypothesis of no difference in taste.

(b) If seven of the ten subjects correctly identify the artificial sweetener, is this outcome strong evidence of a taste difference? Explain.

3. Suppose that a small population consists of the $N = 6$ values 2, 3, 4, 4, 5, and 6.

(a) Sketch a relative frequency histogram for this population and compute the population mean, μ, and standard deviation, σ.

(b) Now let $X =$ the value of a single number selected at random from this population. Sketch a probability histogram for this variable X and compute EX and $\operatorname{Var} X$.

(c) Now think of drawing a simple random sample of size $n = 2$ from this small population. Make tables giving the probability distributions of the random variables

$$\overline{X} = \text{the sample mean}$$
$$S^2 = \text{the sample variance}$$

(There are 15 different possible unordered samples of 2 out of 6 items. Each of the 15 possible samples is equally likely to be chosen and has its own corresponding $\bar{x}$ and s^2.) Use the tables and make probability histograms for these random variables. Compute $E\overline{X}$ and $\operatorname{Var}\overline{X}$. How do these compare to μ and σ^2?

4. Sketch probability histograms for the binomial distributions with $n = 5$ and $p = .1, .3, .5, .7$, and $.9$. On each histogram, mark the location of the mean and indicate the size of the standard deviation.

5. Suppose that an eddy current nondestructive evaluation technique for identifying cracks in critical metal parts has a probability of around .20 of detecting a single crack of length .003 in. in a certain material. Suppose further that $n = 8$ specimens of this material, each containing a single crack of length .003 in., are inspected using this technique. Let W be the number of these cracks that are detected. Use an appropriate probability model and evaluate the following:
   (a) $P[W = 3]$
   (b) $P[W \le 2]$
   (c) EW
   (d) Var W
   (e) the standard deviation of W

6. In the situation described in Exercise 5, suppose that a series of specimens, each containing a single crack of length .003 in., are inspected. Let Y be the number of specimens inspected in order to obtain the first crack detection. Use an appropriate probability model and evaluate all of the following:
   (a) $P[Y = 5]$
   (b) $P[Y \le 4]$
   (c) EY
   (d) Var Y
   (e) the standard deviation of Y

7. Sketch probability histograms for the Poisson distributions with means $\lambda = .5, 1.0, 2.0$, and 4.0. On each histogram, mark the location of the mean and indicate the size of the standard deviation.

8. A process for making plate glass produces an average of four seeds (small bubbles) per 100 square feet. Use Poisson distributions and assess probabilities that
   (a) a particular piece of glass 5 ft $\times$ 10 ft will contain more than two seeds.
   (b) a particular piece of glass 5 ft $\times$ 5 ft will contain no seeds.

9. Transmission line interruptions in a telecommunications network occur at an average rate of one per day.
   (a) Use a Poisson distribution as a model for

   $X =$ the number of interruptions in the next five-day work week

   and assess $P[X = 0]$.
   (b) Now consider the random variable

   $Y =$ the number of weeks in the next four in which there are no interruptions

   What is a reasonable probability model for Y? Assess $P[Y = 2]$.

10. Distinguish clearly between the subjects of *probability* and *statistics*. Is one field a subfield of the other?

11. What is the difference between a relative frequency distribution and a probability distribution?

5.2 Continuous Random Variables

It is often convenient to think of a random variable as not discrete but rather continuous in the sense of having a whole (continuous) interval for its set of possible values. The devices used to describe continuous probability distributions differ from the tools studied in the last section. So the first tasks here are to introduce the notion of a probability density function, to show its relationship to the cumulative probability function for a continuous random variable, and to show how it is used to find the mean and variance for a continuous distribution. After this, several standard continuous distributions useful in engineering applications of probability theory will be discussed. That is, the normal (or Gaussian) exponential and Weibull distributions are presented.

5.2.1 Probability Density Functions and Cumulative Probability Functions

The methods used to specify and describe probability distributions have parallels in mechanics. When considering continuous probability distributions, the analogy to mechanics becomes especially helpful. In mechanics, the properties of a continuous mass distribution are related to the possibly varying density of the mass across its region of location. Amounts of mass in particular regions are obtained from the density by integration.

The concept in probability theory corresponding to mass density in mechanics is *probability density*. To specify a continuous probability distribution, one needs to describe "how thick" the probability is in the various parts of the set of possible values. The formal definition is

Definition 12

A **probability density function** for a continuous random variable X is a nonnegative function $f(x)$ with

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{5.13}$$

and such that for all $a \leq b$, one is willing to assign $P[a \leq X \leq b]$ according to

$$\boxed{P[a \leq X \leq b] = \int_a^b f(x)\,dx} \tag{5.14}$$

A generic probability density function is pictured in Figure 5.6. In keeping with equations (5.13) and (5.14), the plot of $f(x)$ does not dip below the x axis, the total area under the curve $y = f(x)$ is 1, and areas under the curve above particular intervals give probabilities corresponding to those intervals.

Mechanics analogy for probability density

In direct analogy to what is done in mechanics, if $f(x)$ is indeed the "density of probability" around x, then the probability in an interval of small length dx around x is approximately $f(x)\,dx$. (In mechanics, if $f(x)$ is *mass* density around x, then the *mass* in an interval of small length dx around x is approximately $f(x)\,dx$.) Then to get a probability between a and b, one needs to sum up such $f(x)\,dx$ values. $\int_a^b f(x)\,dx$ is exactly the limit of $\sum f(x)\,dx$ values as dx gets small. (In mechanics, $\int_a^b f(x)\,dx$ is the mass between a and b.) So the expression (5.14) is reasonable.

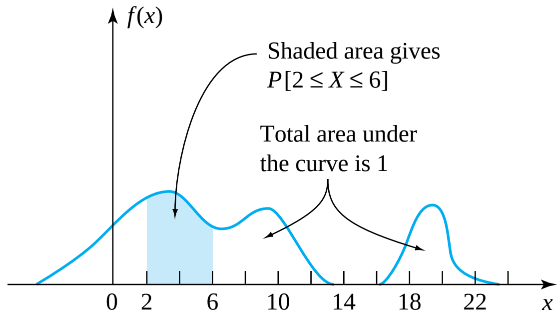


Figure 5.6 A generic probability density function

Example 8

The Random Time Until a First Arc in the Bob Drop Experiment

Consider once again the bob drop experiment first described in Section 1.4 and revisited in Example 4 in Chapter 4. In any use of the apparatus, the bob is almost certainly not released exactly “in sync” with the 60 cycle current that produces the arcs and marks on the paper tape. One could think of a random variable

$$Y = \text{the time elapsed (in seconds) from bob release until the first arc}$$

as continuous with set of possible values $(0, \frac{1}{60})$.

What is a plausible probability density function for Y? The symmetry of this situation suggests that probability density should be constant over the interval $(0, \frac{1}{60})$ and 0 outside the interval. That is, for any two values y_1 and y_2 in $(0, \frac{1}{60})$, the probability that Y takes a value within a small interval around y_1 of length dy (i.e., $f(y_1)\,dy$ approximately) should be the same as the probability that Y takes a value within a small interval around y_2 of the same length dy (i.e., $f(y_2)\,dy$ approximately). This forces $f(y_1) = f(y_2)$, so there must be a constant probability density on $(0, \frac{1}{60})$.

Now if $f(y)$ is to have the form

$$f(y) = \begin{cases} c & \text{for } 0 < y < \frac{1}{60} \\ 0 & \text{otherwise} \end{cases}$$

for some constant c (i.e., is to be as pictured in Figure 5.7), in light of equation (5.13), it must be that

$$1 = \int_{-\infty}^{\infty} f(y)\,dy = \int_{-\infty}^{0} 0\,dy + \int_{0}^{1/60} c\,dy + \int_{1/60}^{\infty} 0\,dy = \frac{c}{60}$$

That is, $c = 60$, and thus,

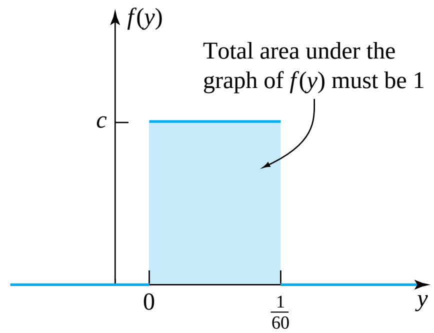


Figure 5.7 Probability density function for Y (time elapsed before arc)

$$f(y) = \begin{cases} 60 & \text{for } 0 < y < \frac{1}{60} \\ 0 & \text{otherwise} \end{cases} \tag{5.15}$$

If the function specified by equation (5.15) is adopted as a probability density for Y, it is then (for example) possible to calculate that

$$P\left[Y \le \frac{1}{100}\right] = \int_{-\infty}^{1/100} f(y)\,dy = \int_{-\infty}^{0} 0\,dy + \int_{0}^{1/100} 60\,dy = .6$$

For X a continuous random variable, $P[X = a] = 0$

One point about continuous probability distributions that may at first seem counterintuitive concerns the probability associated with a continuous random variable assuming a particular prespecified value (say, a). Just as the mass a continuous mass distribution places at a single point is 0, so also is $P[X = a] = 0$ for a continuous random variable X. This follows from equation (5.14), because

$$P[a \le X \le a] = \int_{a}^{a} f(x)\,dx = 0$$

One consequence of this mathematical curiosity is that when working with continuous random variables, you don't need to worry about whether or not inequality signs you write are strict inequality signs. That is, if X is continuous,

$$P[a \le X \le b] = P[a < X \le b] = P[a \le X < b] = P[a < X < b]$$

Definition 6 gave a perfectly general definition of the cumulative probability function for a random variable (which was specialized in Section 5.1 to the case of a discrete variable). Here equation (5.14) can be used to express the cumulative probability function for a continuous random variable in terms of an integral of its probability density. That is, for X continuous with probability density $f(x)$,

Cumulative probability function for a continuous variable

$$\boxed{F(x) = P[X \le x] = \int_{-\infty}^{x} f(t)\,dt} \tag{5.16}$$

$F(x)$ is obtained from $f(x)$ by integration, and applying the fundamental theorem of calculus to equation (5.16)

Another relationship between F(x) and f(x)

$$\boxed{\frac{d}{dx}F(x) = f(x)} \tag{5.17}$$

That is, $f(x)$ is obtained from $F(x)$ by differentiation.

Example 8 (*continued*)

The cumulative probability function for Y, the elapsed time from bob release until first arc, is easily obtained from equation (5.15). For $y \le 0$,

$$F(y) = P[Y \le y] = \int_{-\infty}^{y} f(t)\,dt = \int_{-\infty}^{y} 0\,dt = 0$$

and for $0 < y \le \frac{1}{60}$,

$$F(y) = P[Y \le y] = \int_{-\infty}^{y} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{y} 60\,dt = 0 + 60y = 60y$$

and for $y > \frac{1}{60}$,

$$F(y) = P[Y \le y] = \int_{-\infty}^{y} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{1/60} 60\,dt + \int_{1/60}^{y} 0\,dt = 1$$

That is,

$$F(y) = \begin{cases} 0 & \text{if } y \le 0 \\ 60y & \text{if } 0 < y \le 1/60 \\ 1 & \text{if } \frac{1}{60} < y \end{cases}$$

A plot of $F(y)$ is given in Figure 5.8. Comparing Figure 5.8 to Figure 5.7 shows that indeed the graph of $F(y)$ has slope 0 for $y < 0$ and $y > \frac{1}{60}$ and slope 60 for $0 < y < \frac{1}{60}$. That is, $f(y)$ is the derivative of $F(y)$, as promised by equation (5.17).

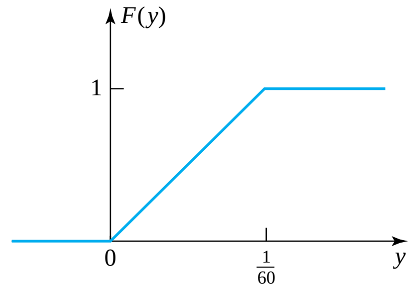


Figure 5.8 Cumulative probability function for Y (time elapsed before arc)

Figure 5.8 is typical of cumulative probability functions for continuous distributions. The graphs of such cumulative probability functions are *continuous* in the sense that they are unbroken curves.

5.2.2 Means and Variances for Continuous Distributions

A plot of the probability density $f(x)$ is a kind of idealized histogram. It has the same kind of visual interpretations that have already been applied to relative frequency histograms and probability histograms. Further, it is possible to define a mean and variance for a continuous probability distribution. These numerical summaries are used in the same way that means and variances are used to describe data sets and discrete probability distributions.

Definition 13

The **mean** or **expected value** of a continuous random variable X (sometimes called the mean of its probability distribution) is

$$EX = \int_{-\infty}^{\infty} x\, f(x)\, dx \tag{5.18}$$

As for discrete random variables, the notation μ is sometimes used in place of EX.

Formula (5.18) is perfectly plausible from at least two perspectives. First, the probability in a small interval around x of length dx is approximately $f(x)\, dx$. So multiplying this by x and summing as in Definition 7, one has $\sum x\, f(x)\, dx$, and formula (5.18) is exactly the limit of such sums as dx gets small. And second, in mechanics the center of mass of a continuous mass distribution is of the form given in equation (5.18) except for division by a total mass, which for a probability distribution is 1.

Example 8 (*continued*)

Thinking of the probability density in Figure 5.7 as an idealized histogram and thinking of the balance point interpretation of the mean, it is clear that EY had better turn out to be $\frac{1}{120}$ for the elapsed time variable. Happily, equations (5.18) and (5.15) give

$$\mu = EY = \int_{-\infty}^{\infty} y\, f(y)\, dy = \int_{-\infty}^{0} y \cdot 0\, dy + \int_{0}^{1/60} y \cdot 60\, dy + \int_{1/60}^{\infty} y \cdot 0\, dy$$

$$= 30y^2 \Big|_0^{1/60} = \frac{1}{120} \text{ sec}$$

"Continuization" of the formula for the variance of a discrete random variable produces a definition of the variance of a continuous random variable.

Definition 14

The **variance** of a continuous random variable X (sometimes called the variance of its probability distribution) is

$$\operatorname{Var} X = \int_{-\infty}^{\infty} (x - EX)^2 f(x)\, dx \quad \left(= \int_{-\infty}^{\infty} x^2 f(x)\, dx - (EX)^2 \right) \tag{5.19}$$

The **standard deviation** of X is $\sqrt{\operatorname{Var} X}$. Often the notation σ^2 is used in place of $\operatorname{Var} X$, and σ is used in place of $\sqrt{\operatorname{Var} X}$.

Example 8 (*continued*)

Return for a final time to the bob drop and the random variable Y. Using formula (5.19) and the form of Y's probability density,

$$\sigma^2 = \operatorname{Var} Y = \int_{-\infty}^{0} \left(y - \frac{1}{120}\right)^2 \cdot 0\, dy + \int_{0}^{1/60} \left(y - \frac{1}{120}\right)^2 \cdot 60\, dy$$

$$+ \int_{1/60}^{\infty} \left(y - \frac{1}{120}\right)^2 \cdot 0\, dy = \frac{60\left(y - \frac{1}{120}\right)^3}{3} \Bigg|_0^{1/60}$$

$$= \frac{1}{3}\left(\frac{1}{120}\right)^2$$

So the standard deviation of Y is

$$\sigma = \sqrt{\operatorname{Var} Y} = \sqrt{\frac{1}{3}\left(\frac{1}{120}\right)^2} = .0048 \text{ sec}$$

5.2.3 The Normal Probability Distributions

Just as there are a number of standard discrete distributions commonly applied to engineering problems, there are also a number of standard continuous probability distributions. This text has already alluded to the **normal** or **Gaussian distributions** and made use of their properties in producing normal plots. It is now time to introduce them formally.

Definition 15

> The **normal** or **Gaussian** (μ, σ^2) **distribution** is a continuous probability distribution with probability density
>
> $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \qquad \text{for all } x \tag{5.20}$$
>
> for $\sigma > 0$.

It is not necessarily obvious, but formula (5.20) does yield a legitimate probability density, in that the total area under the curve $y = f(x)$ is 1. Further, it is also the case that

Normal distribution mean and variance

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

and

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

That is, the parameters μ and σ^2 used in Definition 15 are indeed, respectively, the mean and variance (as defined in Definitions 13 and 14) of the distribution.

Figure 5.9 is a graph of the probability density specified by formula (5.20). The bell-shaped curve shown there is symmetric about $x = \mu$ and has inflection points at $\mu - \sigma$ and $\mu + \sigma$. The exact form of formula (5.20) has a number of theoretical origins. It is also a form that turns out to be empirically useful in a great variety of applications.

In theory, probabilities for the normal distributions can be found directly by integration using formula (5.20). Indeed, readers with pocket calculators that are preprogrammed to do numerical integration may find it instructive to check some of the calculations in the examples that follow, by straightforward use of formulas (5.14) and (5.20). But the freshman calculus methods of evaluating integrals via antidifferentiation will fail when it comes to the normal densities. They do not have antiderivatives that are expressible in terms of elementary functions. Instead, special normal probability tables are typically used.

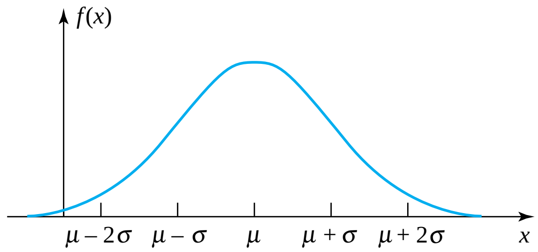


Figure 5.9 Graph of a normal probability density function

The use of tables for evaluating normal probabilities depends on the following relationship. If X is normally distributed with mean μ and variance σ^2,

$$P[a \leq X \leq b] = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\,dx = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz \qquad \textbf{(5.21)}$$

where the second inequality follows from the change of variable or substitution $z = \frac{x-\mu}{\sigma}$. Equation (5.21) involves an integral of the normal density with $\mu = 0$ and $\sigma = 1$. It says that evaluation of all normal probabilities can be reduced to the evaluation of normal probabilities for that special case.

Definition 16

The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**.

Relation between normal (μ, σ^2) probabilities and standard normal probabilities

The relationship between normal (μ, σ^2) and standard normal probabilities is illustrated in Figure 5.10. Once one realizes that probabilities for all normal distributions can be had by tabulating probabilities for only the standard normal distribution, it is a relatively simple matter to use techniques of numerical integration to produce a standard normal table. The one that will be used in this text (other forms are possible) is given in Table B.3. It is a table of the **standard normal cumulative probability function**. That is, for values z located on the table's margins, the entries in the table body are

$$\Phi(z) = F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$$

(Φ is routinely used to stand for the standard normal cumulative probability function, instead of the more generic F.)

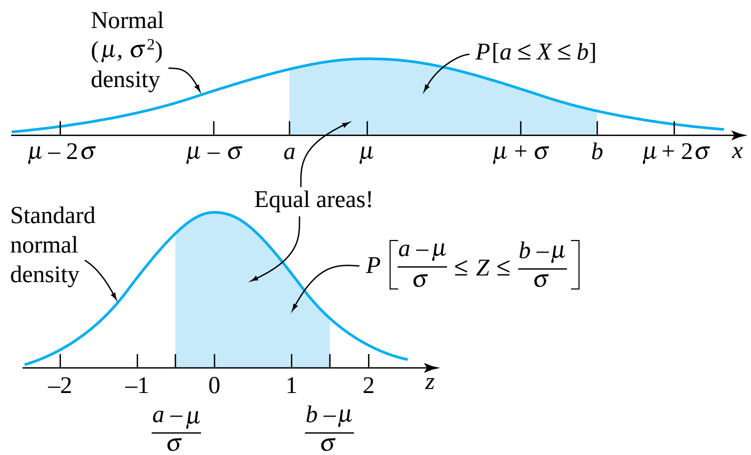


Figure 5.10 Illustration of the relationship between normal (μ, σ^2) and standard normal probabilities

Example 9

Standard Normal Probabilities

Suppose that Z is a standard normal random variable. We will find some probabilities for Z using Table B.3.

By a straight table look-up,

$$P[Z < 1.76] = \Phi(1.76) = .96$$

(The tabled value is .9608, but in keeping with the earlier promise to state final probabilities to only two decimal places, the tabled value was rounded to get .96.) After two table look-ups and a subtraction,

$$\begin{aligned} P[.57 < Z < 1.32] &= P[Z < 1.32] - P[Z \le .57] \\ &= \Phi(1.32) - \Phi(.57) \\ &= .9066 - .7157 \\ &= .19 \end{aligned}$$

And a single table look-up and a subtraction yield a right-tail probability like

$$P[Z > -.89] = 1 - P[Z \le -.89] = 1 - .1867 = .81$$

As the table was used in these examples, probabilities for values z located on the table's margins were found in the table's body. The process can be run in

Example 9
(continued)

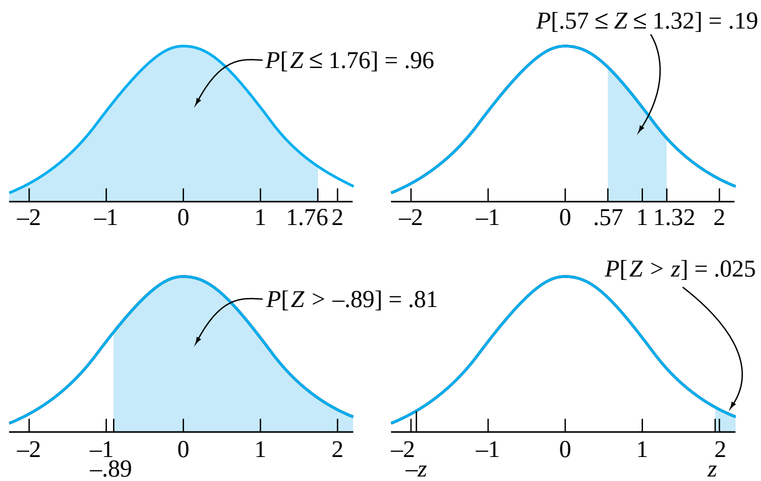


Figure 5.11 Standard normal probabilities for Example 9

reverse. Probabilities located in the table's body can be used to specify values z on the margins. For example, consider locating a value z such that

$$P[-z < Z < z] = .95$$

z will then put probability $\frac{1-.95}{2} = .025$ in the right tail of the standard normal *distribution*—i.e., be such that $\Phi(z) = .975$. Locating .975 in the table body, one sees that $z = 1.96$.

Figure 5.11 illustrates all of the calculations for this example.

The last part of Example 9 amounts to finding the .975 quantile for the standard normal distribution. In fact, the reader is now in a position to understand the origin of Table 3.10 (see page 89). The standard normal quantiles there were found by looking in the body of Table B.3 for the relevant probabilities and then locating corresponding z's on the margins.

In mathematical symbols, for $\Phi(z)$, the standard normal cumulative probability function, and $Q_z(p)$, **the standard normal quantile function**,

$$\left.\begin{aligned} \Phi(Q_z(p)) &= p \\ Q_z(\Phi(z)) &= z \end{aligned}\right\} \qquad \textbf{(5.22)}$$

Relationships (5.22) mean that Q_z and Φ are inverse functions. (In fact, the relationship $Q = F^{-1}$ is not just a standard normal phenomenon but is true in general for continuous distributions.)

Relationship (5.21) shows how to use the standard normal cumulative probability function to find general normal probabilities. For X normal (μ, σ^2) and a value

x associated with X, one converts to units of standard deviations above the mean via

z-value for a value x of a normal (μ, σ^2) random variable

$$\boxed{z = \frac{x - \mu}{\sigma}} \tag{5.23}$$

and then consults the standard normal table using z instead of x.

Example 10

Net Weights of Jars of Baby Food

J. Fisher, in his article "Computer Assisted Net Weight Control" (*Quality Progress*, June 1983), discusses the filling of food containers by weight. In the article, there is a reasonably bell-shaped histogram of individual net weights of jars of strained plums with tapioca. The mean of the values portrayed is about 137.2 g, and the standard deviation is about 1.6 g. The declared (or label) weight on jars of this product is 135.0 g.

Suppose that it is adequate to model

$$W = \text{the next strained plums and tapioca fill weight}$$

with a normal distribution with $\mu = 137.2$ and $\sigma = 1.6$. And further suppose the probability that the next jar filled is below declared weight (i.e., $P[W < 135.0]$) is of interest. Using formula (5.23), $w = 135.0$ is converted to units of standard deviations above μ (converted to a z-value) as

$$z = \frac{135.0 - 137.2}{1.6} = -1.38$$

Then, using Table B.3,

$$P[W < 135.0] = \Phi(-1.38) = .08$$

This model puts the chance of obtaining a below-nominal fill level at about 8%.

As a second example, consider the probability that W is within 1 gram of nominal (i.e., $P[134.0 < W < 136.0]$). Using formula (5.23), both $w_1 = 134.0$ and $w_2 = 136.0$ are converted to z-values or units of standard deviations above the mean as

$$z_1 = \frac{134.0 - 137.2}{1.6} = -2.00$$

$$z_2 = \frac{136.0 - 137.2}{1.6} = -.75$$

Example 10 (*continued*)

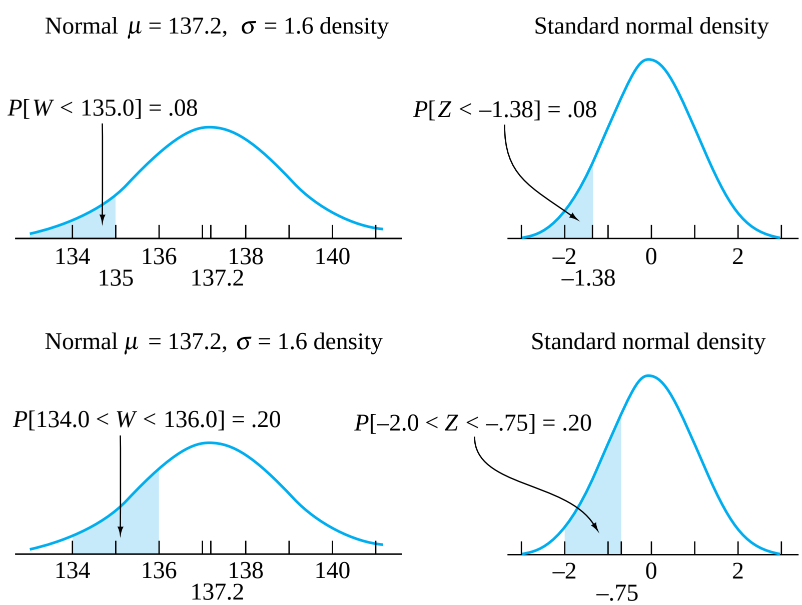


Figure 5.12 Normal probabilities for Example 10

So then

$$P[134.0 < W < 136.0] = \Phi(-.75) - \Phi(-2.00) = .2266 - .0228 = .20$$

The preceding two probabilities and their standard normal counterparts are shown in Figure 5.12.

The calculations for this example have consisted of starting with all of the quantities on the right of formula (5.23) and going from the margin of Table B.3 to its body to find probabilities for W. An important variant on this process is to instead go from the body of the table to its margins to obtain z, and then—given only two of the three quantities on the right of formula (5.23)—to solve for the third.

For example, suppose that it is easy to adjust the aim of the filling process (i.e., the mean μ of W) and one wants to decrease the probability that the next jar is below the declared weight of 135.0 to .01 by increasing μ. What is the minimum μ that will achieve this (assuming that σ remains at 1.6 g)?

Figure 5.13 shows what to do. μ must be chosen in such a way that $w = 135.0$ becomes the .01 quantile of the normal distribution with mean μ and standard deviation $\sigma = 1.6$. Consulting either Table 3.10 or Table B.3, it is easy to determine that the .01 quantile of the standard normal distribution is

$$z = Q_z(.01) = -2.33$$

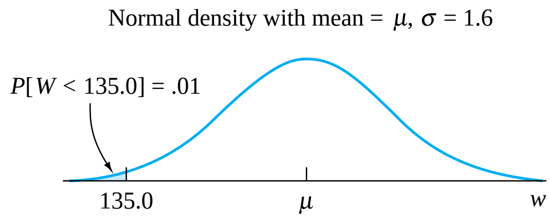


Figure 5.13 Normal distribution and $P[W < 135.0] = .01$

So in light of equation (5.23) one wants

$$-2.33 = \frac{135.0 - \mu}{1.6}$$

i.e.,

$$\mu = 138.7 \text{ g}$$

An increase of about $138.7 - 137.2 = 1.5$ g in fill level aim is required.

In practical terms, the reduction in $P[W < 135.0]$ is bought at the price of increasing the average *give-away cost* associated with filling jars so that on average they contain much more than the nominal contents. In some applications, this type of cost will be prohibitive. There is another approach open to a process engineer. That is to reduce the variation in fill level through acquiring more precise filling equipment. In terms of equation (5.23), instead of increasing μ one might consider paying the cost associated with reducing σ. The reader is encouraged to verify that a reduction in σ to about .94 g would also produce $P[W < 135.0] = .01$ without any change in μ.

As Example 10 illustrates, equation (5.23) is the fundamental relationship used in problems involving normal distributions. One way or another, three of the four entries in equation (5.23) are specified, and the fourth must be obtained.

5.2.4 The Exponential Distributions (*Optional*)

Section 5.1 discusses the fact that the Poisson distributions are often used as models for the number of occurrences of a relatively rare phenomenon in a specified interval of time. The same mathematical theory that suggests the appropriateness of the Poisson distributions in that context also suggests the usefulness of the **exponential distributions** for describing waiting times until occurrences.

Definition 17

The **exponential** (α) **distribution** is a continuous probability distribution with probability density

$$f(x) = \begin{cases} \frac{1}{\alpha} e^{-x/\alpha} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(5.24)}$$

for $\alpha > 0$.

Figure 5.14 shows plots of $f(x)$ for three different values of α. Expression (5.24) is extremely convenient, and it is not at all difficult to show that α is both the mean and the standard deviation of the exponential (α) distribution. That is,

Mean of the exponential (α) distribution

$$\mu = EX = \int_0^\infty x \frac{1}{\alpha} e^{-x/\alpha}\, dx = \alpha$$

and

Variance of the exponential (α) distribution

$$\sigma^2 = \operatorname{Var} X = \int_0^\infty (x - \alpha)^2 \frac{1}{\alpha} e^{-x/\alpha}\, dx = \alpha^2$$

Further, the exponential (α) distribution has a simple cumulative probability function,

Exponential (α) cumulative probability function

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 - e^{-x/\alpha} & \text{if } x > 0 \end{cases} \quad \textbf{(5.25)}$$

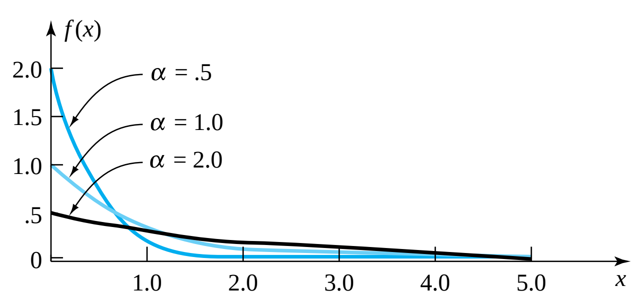


Figure 5.14 Three exponential probability densities

Example 11
(Example 7 revisited)

The Exponential Distribution and Arrivals at a University Library

Recall that Stork, Wohlsdorf, and McArthur found the arrival rate of students at the ISU library between 12:00 and 12:10 P.M. early in the week to be about 12.5 students per minute. That translates to a $\frac{1}{12.5} = .08$ min average waiting time between student arrivals.

Consider observing the ISU library entrance beginning at exactly noon next Tuesday and define the random variable

$$T = \text{the waiting time (in minutes) until the first student passes through the door}$$

A possible model for T is the exponential distribution with $\alpha = .08$. Using it, the probability of waiting more than 10 seconds ($\frac{1}{6}$ min) for the first arrival is

$$P\left[T > \frac{1}{6}\right] = 1 - F\left(\frac{1}{6}\right) = 1 - \left(1 - e^{-1/6(.08)}\right) = .12$$

This result is pictured in Figure 5.15.

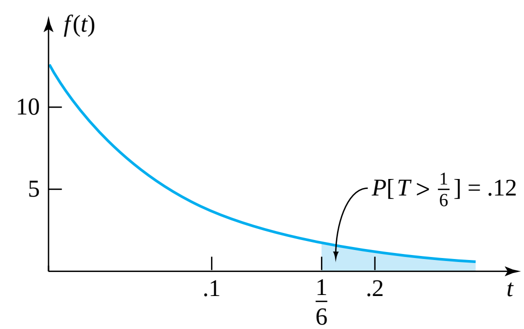


Figure 5.15 Exponential probability for Example 11

Geometric and exponential distributions

The exponential distribution is the continuous analog of the geometric distribution in several respects. For one thing, both the geometric probability function and the exponential probability density decline exponentially in their arguments x. For another, they both possess a kind of **memoryless property**. If the first success in a series of independent identical success-failure trials is known not to have occurred through trial t_0, then the additional number of trials (beyond t_0) needed to produce the first success is a geometric (p) random variable (as was the total number of trials required from the beginning). Similarly, if an exponential (α) waiting time is known not to have been completed by time t_0, then the additional waiting time to completion is exponential (α). This memoryless property is related to the *force-of-mortality function* of the distribution being constant. The force-of-mortality function for a distribution is a concept of reliability theory discussed briefly in Appendix A.4.

5.2.5 The Weibull Distributions (*Optional*)

The Weibull distributions generalize the exponential distributions and provide much more flexibility in terms of distributional shape. They are extremely popular with engineers for describing the strength properties of materials and the life lengths of manufactured devices. The most natural way to specify these distributions is through their cumulative probability functions.

Definition 18

The **Weibull** (α, β) **distribution** is a continuous probability distribution with cumulative probability function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-(x/\alpha)^{\beta}} & \text{if } x \geq 0 \end{cases} \tag{5.26}$$

for parameters $\alpha > 0$ and $\beta > 0$.

Beginning from formula (5.26), it is possible to determine properties of the Weibull distributions. Differentiating formula (5.26) produces the Weibull (α, β) probability density

Weibull (α, β) probability density

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{\beta}{\alpha^{\beta}} x^{\beta-1} e^{-(x/\alpha)^{\beta}} & \text{if } x > 0 \end{cases} \tag{5.27}$$

This in turn can be shown to yield the mean

Weibull (α, β) mean

$$\mu = EX = \alpha\Gamma\left(1 + \tfrac{1}{\beta}\right) \tag{5.28}$$

and variance

Weibull (α, β) variance

$$\sigma^2 = \operatorname{Var} X = \alpha^2 \left[\Gamma\left(1 + \tfrac{2}{\beta}\right) - \left(\Gamma\left(1 + \tfrac{1}{\beta}\right)\right)^2\right] \tag{5.29}$$

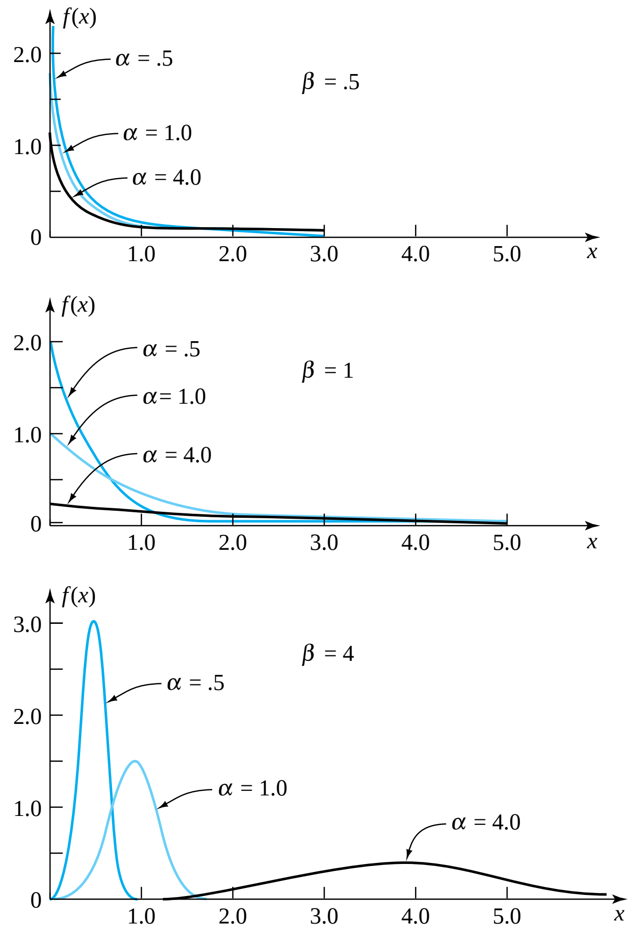


Figure 5.16 Nine Weibull probability densities

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ is the *gamma function* of advanced calculus. (For integer values n, $\Gamma(n) = (n-1)!$.) These formulas for $f(x)$, μ, and σ^2 are not particularly illuminating. So it is probably most helpful to simply realize that β controls the shape of the Weibull distribution and that α controls the scale. Figure 5.16 shows plots of $f(x)$ for several (α, β) pairs.

Note that $\beta = 1$ gives the special case of the exponential distributions. For small β, the distributions are decidedly right-skewed, but for β larger than about 3.6, they actually become left-skewed. Regarding distribution location, the form of the distribution mean given in equation (5.28) is not terribly revealing. It is perhaps more helpful that the median for the Weibull (α, β) distribution is

Weibull (α, β) median

$$\boxed{Q(.5) = \alpha e^{-(.3665/\beta)}} \qquad \textbf{(5.30)}$$

So, for example, for large shape parameter β the Weibull median is essentially α. And formulas (5.28) through (5.30) show that for fixed β the Weibull mean, median, and standard deviation are all proportional to the scale parameter α.

Example 12

The Weibull Distribution and the Strength of a Ceramic Material

The report "Review of Workshop on Design, Analysis and Reliability Prediction for Ceramics—Part II" by E. Lenoe (*Office of Naval Research Far East Scientific Bulletin*, 1987) suggests that tensile strengths (MPa) of .95 mm rods of HIPped UBE SN-10 with 2.5% yttria material can be described by a Weibull distribution with $\beta = 8.8$ and median 428 MPa. Let

$$S = \text{measured tensile strength of an additional rod (MPa)}$$

Under the assumption that S can be modeled using a Weibull distribution with the suggested characteristics, suppose that $P[S \leq 400]$ is needed. Using equation (5.30),

$$428 = \alpha e^{-(.3665/8.8)}$$

Thus, the Weibull scale parameter is

$$\alpha = 446$$

Then, using equation (5.26),

$$P[S \leq 400] = 1 - e^{-(400/446)^{8.8}} = .32$$

Figure 5.17 illustrates this probability calculation.

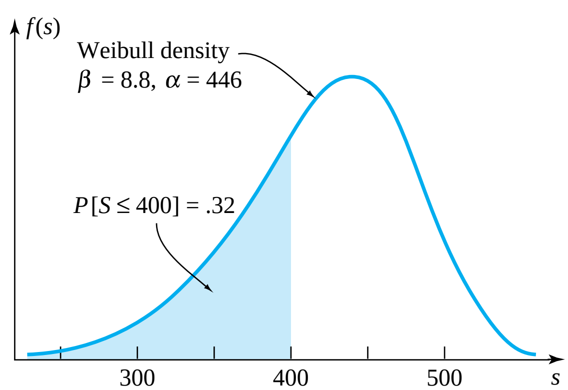


Figure 5.17 Weibull density and $P[S \leq 400]$

Section 2 Exercises

1. The random number generator supplied on a calculator is not terribly well chosen, in that values it generates are not adequately described by a distribution uniform on the interval $(0, 1)$. Suppose instead that a probability density

$$f(x) = \begin{cases} k(5 - x) & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

is a more appropriate model for X = the next value produced by this random number generator.

(a) Find the value of k.
(b) Sketch the probability density involved here.
(c) Evaluate $P[.25 < X < .75]$.
(d) Compute and graph the cumulative probability function for X, $F(x)$.
(e) Calculate EX and the standard deviation of X.

2. Suppose that Z is a standard normal random variable. Evaluate the following probabilities involving Z:

(a) $P[Z < -.62]$ (b) $P[Z > 1.06]$
(c) $P[-.37 < Z < .51]$ (d) $P[|Z| \le .47]$
(e) $P[|Z| > .93]$ (f) $P[-3.0 < Z < 3.0]$

Now find numbers # such that the following statements involving Z are true:

(g) $P[Z \le \#] = .90$ (h) $P[|Z| < \#] = .90$
(i) $P[|Z| > \#] = .03$

3. Suppose that X is a normal random variable with mean 43.0 and standard deviation 3.6. Evaluate the following probabilities involving X:

(a) $P[X < 45.2]$ (b) $P[X \le 41.7]$
(c) $P[43.8 < X \le 47.0]$ (d) $P[|X - 43.0| \le 2.0]$
(e) $P[|X - 43.0| > 1.7]$

Now find numbers # such that the following statements involving X are true:

(f) $P[X < \#] = .95$ (g) $P[X \ge \#] = .30$
(h) $P[|X - 43.0| > \#] = .05$

4. The diameters of bearing journals ground on a particular grinder can be described as normally distributed with mean 2.0005 in. and standard deviation .0004 in.

(a) If engineering specifications on these diameters are 2.0000 in. $\pm$ .0005 in., what fraction of these journals are in specifications?
(b) What adjustment to the grinding process (holding the process standard deviation constant) would increase the fraction of journal diameters that will be in specifications? What appears to be the best possible fraction of journal diameters inside $\pm$ .0005 in. specifications, given the $\sigma = .0004$ in. apparent precision of the grinder?
(c) Suppose consideration was being given to purchasing a more expensive/newer grinder, capable of holding tighter tolerances on the parts it produces. What σ would have to be associated with the new machine in order to guarantee that (when perfectly adjusted so that $\mu = 2.0000$) the grinder would produce diameters with at least 95% meeting 2.0000 in. $\pm$ .0005 in. specifications?

5. The mileage to first failure for a model of military personnel carrier can be modeled as exponential with mean 1,000 miles.

(a) Evaluate the probability that a vehicle of this type gives less than 500 miles of service before first failure. Evaluate the probability that it gives at least 2,000 miles of service before first failure.
(b) Find the .05 quantile of the distribution of mileage to first failure. Then find the .90 quantile of the distribution.

6. Some data analysis shows that lifetimes, x (in 10^6 revolutions before failure), of certain ball bearings can be modeled as Weibull with $\beta = 2.3$ and $\alpha = 80$.

(a) Make a plot of the Weibull density (5.27) for this situation. (Plot for x between 0 and 200. Standard statistical software packages like MINITAB will have routines for evaluating this density. In MINITAB look under the "Calc/Probability Distributions/Weibull" menu.)
(b) What is the median bearing life?
(c) Find the .05 and .95 quantiles of bearing life.

5.3 Probability Plotting (*Optional*)

Calculated probabilities are only as relevant in a given application as are the distributions used to produce them. It is thus important to have data-based methods to assess the relevance of a given continuous distribution to a given application. The basic logic for making such tools was introduced in Section 3.2. Suppose you have data consisting of n realizations of a random variable X, say $x_1 \le x_2 \le \cdots \le x_n$ and want to know whether a probability density with the same shape as $f(x)$ might adequately describe X. To investigate, it is possible to make and interpret a probability plot consisting of n ordered pairs

Ordered pairs making a probability plot

$$\left(x_i, Q\left(\frac{i-.5}{n}\right)\right)$$

where x_i is the ith smallest data value (the $\left(\frac{i-.5}{n}\right)$ quantile of the data set) and $Q\left(\frac{i-.5}{n}\right)$ is the $\left(\frac{i-.5}{n}\right)$ quantile of the probability distribution specified by $f(x)$.

This section will further discuss the importance of this method. First, some additional points about probability plotting are made in the familiar context where $f(x)$ is the standard normal density (i.e., in the context of normal plotting). Then the general applicability of the idea is illustrated by using it in assessing the appropriateness of exponential and Weibull models. In the course of the discussion, the importance of probability plotting to process capability studies and life data analysis will be indicated.

5.3.1 More on Normal Probability Plots

Definition 15 gives the form of the normal or Gaussian probability density with mean μ and variance σ^2. The discussion that follows the definition shows that all normal distributions have the same essential shape. Thus, a theoretical Q-Q plot using standard normal quantiles can be used to judge whether or not there is *any* normal probability distribution that seems a sensible model.

Example 13 **Weights of Circulating U.S. Nickels**

Ash, Davison, and Miyagawa studied characteristics of U.S. nickels. They obtained the weights of 100 nickels to the nearest .01 g. They found those to have a mean of 5.002 g and a standard deviation of .055 g. Consider the weight of another nickel taken from a pocket, say, U. It is sensible to think that $EU \approx 5.002$ g and $\sqrt{\text{Var}\, U} \approx .055$ g. Further, it would be extremely convenient if a normal distribution could be used to describe U. Then, for example, normal distribution calculations with $\mu = 5.002$ g and $\sigma = .055$ g could be used to assess

$$P[U > 5.05] = P[\text{the nickel weighs over 5.05 g}]$$

A way of determining whether or not the students' data support the use of a normal model for U is to make a normal probability plot. Table 5.6 presents the data collected by Ash, Davison, and Miyagawa. Table 5.7 shows some of the calculations used to produce the normal probability plot in Figure 5.18.

Table 5.6
Weights of 100 U.S. Nickels

Weight (g)	Frequency	Weight (g)	Frequency
4.81	1	5.00	12
4.86	1	5.01	10
4.88	1	5.02	7
4.89	1	5.03	7
4.91	2	5.04	5
4.92	2	5.05	4
4.93	3	5.06	4
4.94	2	5.07	3
4.95	6	5.08	2
4.96	4	5.09	3
4.97	5	5.10	2
4.98	4	5.11	1
4.99	7	5.13	1

Table 5.7
Example Calculations for a Normal Plot of Nickel Weights

i	$\left(\frac{i-.5}{100}\right)$	x_i	$Q_z\left(\frac{i-.5}{100}\right)$
1	.005	4.81	−2.576
2	.015	4.86	−2.170
3	.025	4.88	−1.960
4	.035	4.89	−1.812
5	.045	4.91	−1.695
6	.055	4.91	−1.598
7	.065	4.92	−1.514
⋮	⋮	⋮	⋮
98	.975	5.10	1.960
99	.985	5.11	2.170
100	.995	5.13	2.576

Example 13
(continued)

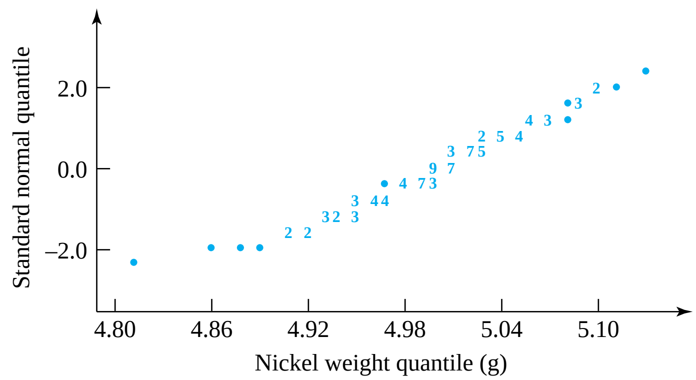


Figure 5.18 Normal plot of nickel weights

At least up to the resolution provided by the graphics in Figure 5.18, the plot is pretty linear for weights above, say, 4.90 g. However, there is some indication that the shape of the lower end of the weight distribution differs from that of a normal distribution. Real nickels seem to be more likely to be light than a normal model would predict. Interestingly enough, the four nickels with weights under 4.90 g were all minted in 1970 or before (these data were collected in 1988). This suggests the possibility that the shape of the lower end of the weight distribution is related to wear patterns and unusual damage (particularly the extreme lower tail represented by the single 1964 coin with weight 4.81 g).

But whatever the origin of the shape in Figure 5.18, its message is clear. For most practical purposes, a normal model for the random variable

$$U = \text{the weight of a nickel taken from a pocket}$$

will suffice. Bear in mind, though, that such a distribution will tend to slightly overstate probabilities associated with larger weights and understate probabilities associated with smaller weights.

Much was made in Section 3.2 of the fact that linearity on a Q-Q plot indicates equality of distribution shape. But to this point, no use has been made of the fact that when there is near-linearity on a Q-Q plot, the *nature of the linear relationship* gives information regarding the relative location and spread of the two distributions involved. This can sometimes provide a way to choose sensible parameters of a theoretical distribution for describing the data set.

For example, a normal probability plot can be used not only to determine whether some normal distribution might describe a random variable but also to graphically pick out *which one* might be used. For a roughly linear normal plot, *Reading a mean and standard deviation from a normal plot*

1. the horizontal coordinate corresponding to a vertical coordinate of 0 provides a mean for a normal distribution fit to the data set, and

2. the reciprocal of the slope provides a standard deviation (this is the difference between the horizontal coordinates of points with vertical coordinates differing by 1).

Example 14

Normal Plotting and Thread Lengths of U-bolts

Table 5.8 gives thread lengths produced in the manufacture of some U-bolts for the auto industry. The measurements are in units of .001 in. over nominal. The particular bolts that gave the measurements in Table 5.8 were sampled from a single machine over a 20-minute period.

Figure 5.19 gives a normal plot of the data. It indicates that (allowing for the fact that the relatively crude measurement scale employed is responsible for the discrete/rough appearance of the plot) a normal distribution might well have been a sensible probability model for the random variable

$$L = \text{the actual thread length of an additional U-bolt manufactured in the same time period}$$

The line eye-fit to the plot further suggests appropriate values for the mean and standard deviation: $\mu \approx 10.8$ and $\sigma \approx 2.1$. (Direct calculation with the data in Table 5.8 gives a sample mean and standard deviation of, respectively, $\bar{l} \approx 10.9$ and $s \approx 1.9$.)

Table 5.8
Measured Thread Lengths for 25 U-Bolts

Thread Length (.001 in. over Nominal)	Tally	Frequency
6	|	1
7		0
8	|||	3
9		0
10	||||	4
11	~~||||~~ ~~||||~~	10
12		0
13	~~||||~~ |	6
14	|	1

Example 14 (*continued*)

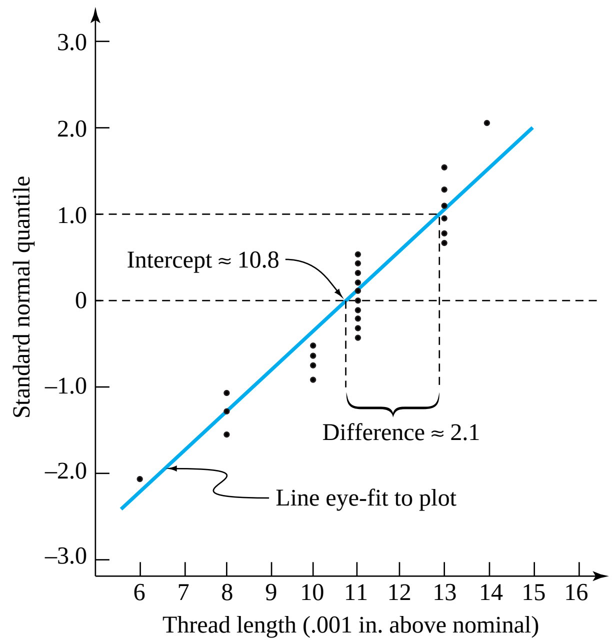


Figure 5.19 Normal plot of thread lengths and eye-fit line

In manufacturing contexts like the previous example, it is common to use the fact that an approximate standard deviation can easily be read from the (reciprocal) slope of a normal plot to obtain a graphical tool for assessing process potential. That is, the primary limitation on the performance of an industrial machine or process is typically the basic precision or short-term variation associated with it. Suppose a dimension of the output of such a process or machine over a short period is approximately normally distributed with standard deviation σ. Then, since for any normal random variable X with mean μ and standard deviation σ,

$$P[\mu - 3\sigma < X < \mu + 3\sigma] > .99$$

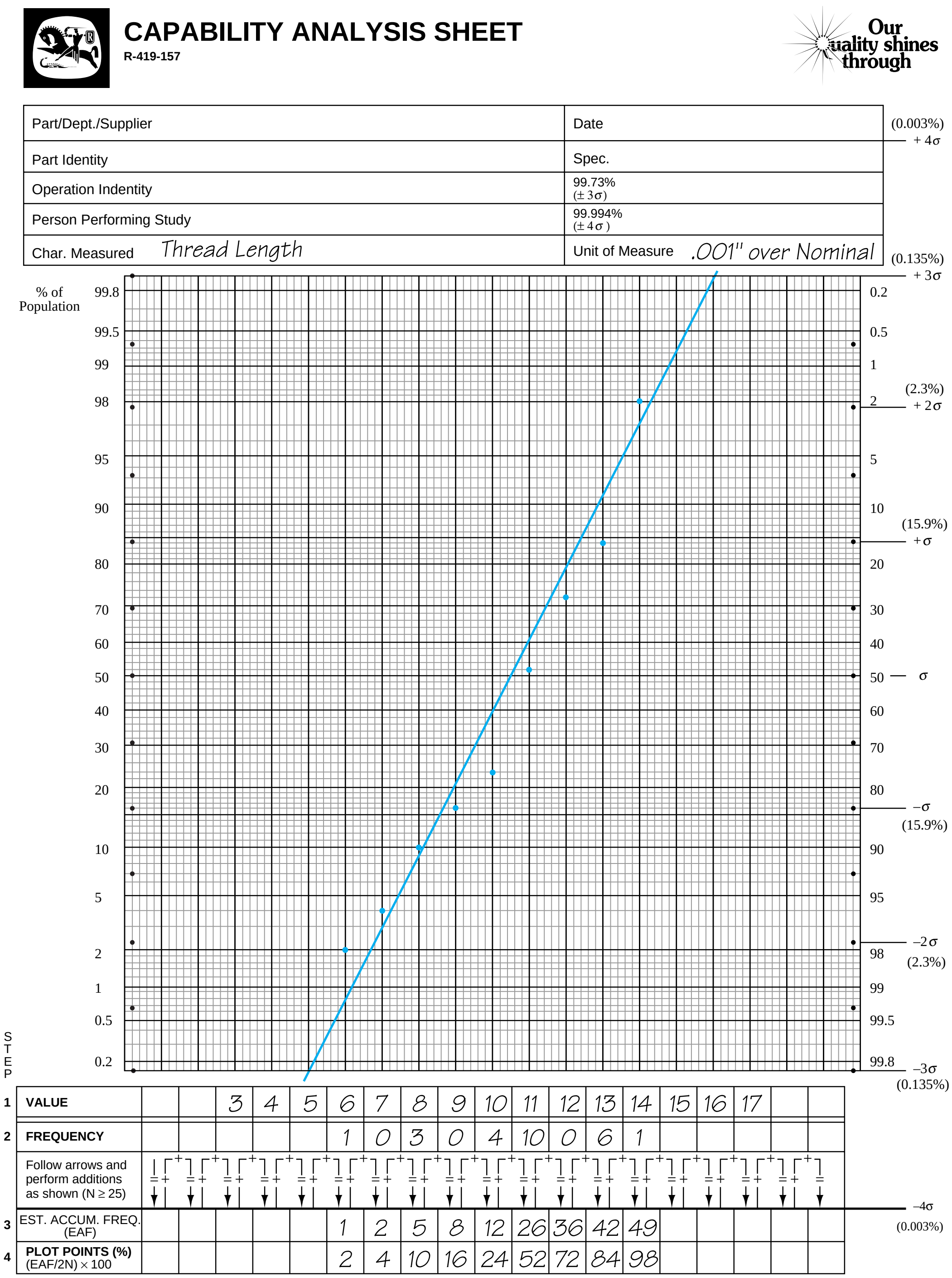
6σ as a process capability

it makes some sense to use 6σ $(= (\mu + 3\sigma) - (\mu - 3\sigma))$ as a measure of **process capability**. And it is easy to read such a capability figure off a normal plot. Many companies use specially prepared *process capability analysis forms* (which are in essence pieces of normal probability paper) for this purpose.

Example 14 (*continued*)

Figure 5.20 is a plot of the thread length data from Table 5.8, made on a common capability analysis sheet. Using the plot, it is very easy, even for someone with limited quantitative background (and perhaps even lacking a basic understanding of the concept of a standard deviation), to arrive at the figure

$$\text{Process capability} \approx 16 - 5 = 11(.001 \text{ in.})$$


CAPABILITY ANALYSIS SHEET
R-419-157
Our Quality shines through

Part/Dept./Supplier	Date
Part Identity	Spec.
Operation Indentity	99.73% (± 3σ)
Person Performing Study	99.994% (± 4σ)
Char. Measured Thread Length	Unit of Measure .001" over Nominal

STEP																				
1	VALUE			3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
2	FREQUENCY						1	0	3	0	4	10	0	6	1					
3	EST. ACCUM. FREQ. (EAF)						1	2	5	8	12	26	36	42	49					
4	PLOT POINTS (%) (EAF/2N) × 100						2	4	10	16	24	52	72	84	98					

Figure 5.20 Thread length data plotted on a capability analysis form (used with permission of Reynolds Metals Company)

5.3.2 Probability Plots for Exponential and Weibull Distributions

To illustrate the application of probability plotting to distributions that are not normal (Gaussian), the balance of this section considers its use with first exponential and then general Weibull models.

Example 15

Service Times at a Residence Hall Depot Counter and Exponential Probability Plotting

Jenkins, Milbrath, and Worth studied service times at a residence hall "depot" counter. Figure 5.21 gives the times (in seconds) required to complete 65 different postage stamp sales at the counter.

The shape of the stem-and-leaf diagram is reminiscent of the shape of the exponential probability densities shown in Figure 5.14. So if one defines the random variable

$$T = \text{the next time required to complete a postage stamp sale at the depot counter}$$

an exponential distribution might somehow be used to describe T.

0	
0	8 8 8 9 9
1	0 0 0 0 0 2 2 2 2 3 3 4 4 4 4 4 4
1	5 6 7 7 7 7 8 8 9 9
2	0 1 2 2 2 2 3 4
2	6 7 8 9 9 9
3	0 2 2 2 4 4
3	6 6 7 7
4	2 3
4	5 6 7 8 8
5	
5	
6	
6	
7	0
7	
8	
8	7

Figure 5.21 Stem-and-leaf plot of service times

The exponential distributions introduced in Definition 17 all have the same essential shape. Thus the exponential distribution with $\alpha = 1$ is a convenient representative of that shape. A plot of $\alpha = 1$ exponential quantiles versus corresponding service time quantiles will give a tool for comparing the empirical shape to the theoretical exponential shape.

For an exponential distribution with mean $\alpha = 1$,

$$F(x) = 1 - e^{-x} \qquad \text{for } x > 0$$

So for $0 < p < 1$, setting $F(x) = p$ and solving,

$$x = -\ln(1 - p)$$

That is, $-\ln(1-p) = Q(p)$, the p quantile of this distribution. Thus, for data $x_1 \le x_2 \le \cdots \le x_n$, an exponential probability plot can be made by plotting the ordered pairs

Points to plot for an exponential probability plot

$$\left(x_i, -\ln\left(1 - \frac{i - .5}{n}\right)\right) \qquad \textbf{(5.31)}$$

Figure 5.22 is a plot of the points in display (5.31) for the service time data. It shows remarkable linearity. Except for the fact that the third- and fourth-largest service times (both 48 seconds) appear to be somewhat smaller than might be predicted based on the shape of the exponential distribution, the empirical service time distribution corresponds quite closely to the exponential distribution shape.

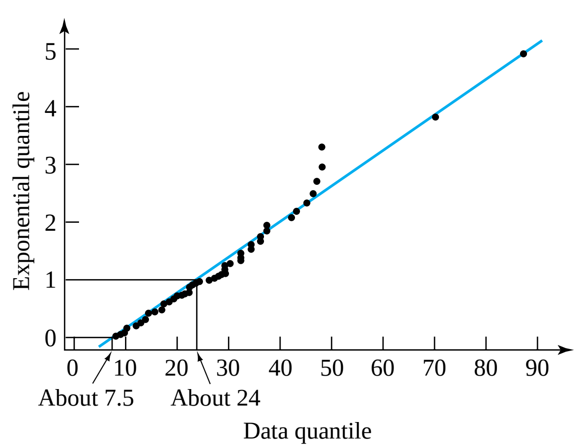


Figure 5.22 Exponential probability plot and eye-fit line for the service times

Example 15 *(continued)*

As was the case in normal-plotting, the character of the linearity in Figure 5.22 also carries some valuable information that can be applied to the modeling of the random variable T. The positioning of the line sketched onto the plot indicates the appropriate location of an exponentially shaped distribution for T, and the slope of the line indicates the appropriate spread for that distribution.

As introduced in Definition 17, the exponential distributions have positive density $f(x)$ for positive x. One might term 0 a *threshold value* for the distributions defined there. In Figure 5.22 the threshold value $(0 = Q(0))$ for the exponential distribution with $\alpha = 1$ corresponds to a service time of roughly 7.5 seconds. This means that to model a variable related to T with a distribution exactly of the form given in Definition 17, it is

$$S = T - 7.5$$

that should be considered.

Further, a change of one unit on the vertical scale in the plot corresponds to a change on the horizontal scale of roughly

$$24 - 7.5 = 16.5 \text{ sec}$$

That is, an exponential model for S ought to have an associated spread that is 16.5 times that of the exponential distribution with $\alpha = 1$.

So ultimately, the data in Figure 5.21 lead via exponential probability plotting to the suggestion that

$$\begin{aligned} S &= T - 7.5 \\ &= \text{the excess of the next time required to complete a postage stamp sale} \\ & \text{ over a threshold value of 7.5 seconds} \end{aligned}$$

be described with the density

$$f(s) = \begin{cases} \dfrac{1}{16.5} e^{-(s/16.5)} & \text{for } s > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{(5.32)}$$

Probabilities involving T can be computed by first expressing them in terms of S and then using expression (5.32). If for some reason a density for T itself is desired, simply shift the density in equation (5.32) to the right 7.5 units to obtain the density

$$f(t) = \begin{cases} \dfrac{1}{16.5} e^{-((t-7.5)/16.5)} & \text{for } t > 7.5 \\ 0 & \text{otherwise} \end{cases}$$

Figure 5.23 shows probability densities for both S and T.

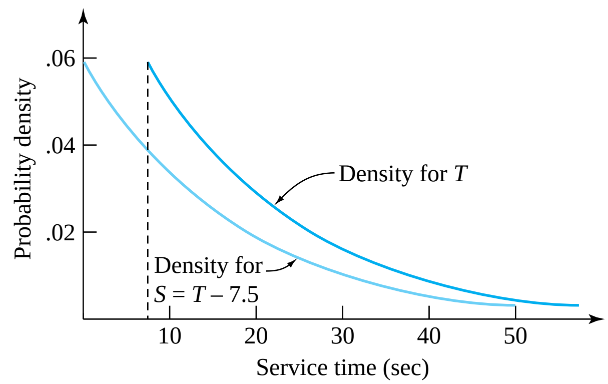


Figure 5.23 Probability densities for both S and T

To summarize the preceding example: Because of the relatively simple form of the exponential $\alpha = 1$ cumulative probability function, it is easy to find quantiles for this distribution. When these are plotted against corresponding quantiles of a data set, an exponential probability plot is obtained. On this plot, linearity indicates exponential shape, the horizontal intercept of a linear plot indicates an appropriate threshold value, and the reciprocal of the slope indicates an appropriate value for the exponential parameter α.

Much the same story can be told for the Weibull distributions for any fixed β. That is, using the form (5.26) of the Weibull cumulative probability function, it is straightforward to argue that for data $x_1 \le x_2 \le \cdots \le x_n$, a plot of the ordered pairs

Points to plot for a fixed β Weibull plot

$$\left(x_i, \left(-\ln\left(1 - \frac{i - .5}{n}\right)\right)^{1/\beta} \right) \tag{5.33}$$

is a tool for investigating whether a variable might be described using a Weibull-shaped distribution *for the particular β in question*. On such a plot, linearity indicates Weibull shape β, the horizontal intercept indicates an appropriate threshold value, and the reciprocal of the slope indicates an appropriate value for the parameter α.

Although the kind of plot indicated by display (5.33) is easy to make and interpret, it is *not* the most common form of probability plotting associated with the Weibull distributions. In order to plot the points in display (5.33), a value of β is input (and a threshold and scale parameter are read off the graph). In most engineering applications of the Weibull distributions, what is needed (instead of a method that inputs β and can be used to identify a threshold and α) is a method that tacitly inputs the 0 threshold implicit in Definition 18 and can be used to identify α and β. This is particularly true in applications to reliability, where the useful life or time to failure of some device is the variable of interest. It is similarly true in applications to material science, where intrinsically positive material properties like yield strength are under study.

It is possible to develop a probability plotting method that allows identification of values for both α and β in Definition 18. The trick is to work on a log scale. That is, if X is a random variable with the Weibull (α, β) distribution, then for $x > 0$,

$$F(x) = 1 - e^{-(x/\alpha)^{\beta}}$$

so that with $Y = \ln(X)$

$$\begin{aligned} P[Y \leq y] &= P[X \leq e^{y}] \\ &= 1 - e^{-(e^{y}/\alpha)^{\beta}} \end{aligned}$$

So for $0 < p < 1$, setting $p = P[Y \leq y]$ gives

$$p = 1 - e^{-(e^{y}/\alpha)^{\beta}}$$

After some algebra this implies

$$\beta y - \beta \ln(\alpha) = \ln\left(-\ln(1-p)\right) \qquad \textbf{(5.34)}$$

Now y is (by design) the p quantile of the distribution of $Y = \ln(X)$. So equation (5.34) says that $\ln(-\ln(1-p))$ is a linear function of $\ln(X)$'s quantile function. The slope of that relationship is β. Further, equation (5.34) shows that when $\ln(-\ln(1-p)) = 0$, the quantile function of $\ln(X)$ has the value $\ln(\alpha)$. So exponentiation of the horizontal intercept gives α. Thus, for data $x_1 \leq x_2 \leq \cdots \leq x_n$, one is led to consider a plot of ordered pairs

Points to plot for a 0-threshold Weibull plot

$$\boxed{\left(\ln x_i, \ln\left(-\ln\left(1 - \frac{i - .5}{n}\right)\right)\right)} \qquad \textbf{(5.35)}$$

Reading α and β from a 0-threshold Weibull plot

If data in hand are consistent with a (0-threshold) Weibull (α, β) model, a reasonably linear plot with

1. slope β and
2. horizontal axis intercept equal to $\ln(\alpha)$

may be expected.

Example 16

Electrical Insulation Failure Voltages and Weibull Plotting

The data given in the stem-and-leaf plot of Figure 5.24 are failure voltages (in kv/mm) for a type of electrical cable insulation subjected to increasing voltage

3	
3	9.4
4	5.3
4	9.2, 9.4
5	1.3, 2.0, 3.2, 3.2, 4.9
5	5.5, 7.1, 7.2, 7.5, 9.2
6	1.0, 2.4, 3.8, 4.3
6	7.3, 7.7

Figure 5.24 Stem-and-leaf plot of insulation failure voltages

stress. They were taken from *Statistical Models and Methods for Lifetime Data* by J. F. Lawless.

Consider the Weibull modeling of

$$R = \text{the voltage at which one additional specimen of this insulation will fail}$$

Table 5.9 shows some of the calculations needed to use display (5.35) to produce Figure 5.25. The near-linearity of the plot in Figure 5.25 suggests that a (0-threshold) Weibull distribution might indeed be used to describe R. A Weibull shape parameter of roughly

$$\beta \approx \text{slope of the fitted line} \approx \frac{1-(-4)}{4.19-3.67} \approx 9.6$$

is indicated. Further, a scale parameter α with

$$\ln(\alpha) \approx \text{horizontal intercept} \approx 4.08$$

and thus

$$\alpha \approx 59$$

appears appropriate.

Example 16 (*continued*)

Table 5.9
Example Calculations for a 0-Threshold Weibull Plot of Failure Voltages

i	$x_i = i$th Smallest Voltage	$\ln(x_i)$	$p = (i - .5)/20$	$\ln(-\ln(1-p))$
1	39.4	3.67	.025	−3.68
2	45.3	3.81	.075	−2.55
3	49.2	3.90	.125	−2.01
4	49.4	3.90	.175	−1.65
⋮	⋮	⋮	⋮	⋮
19	67.3	4.21	.925	.95
20	67.7	4.22	.975	1.31

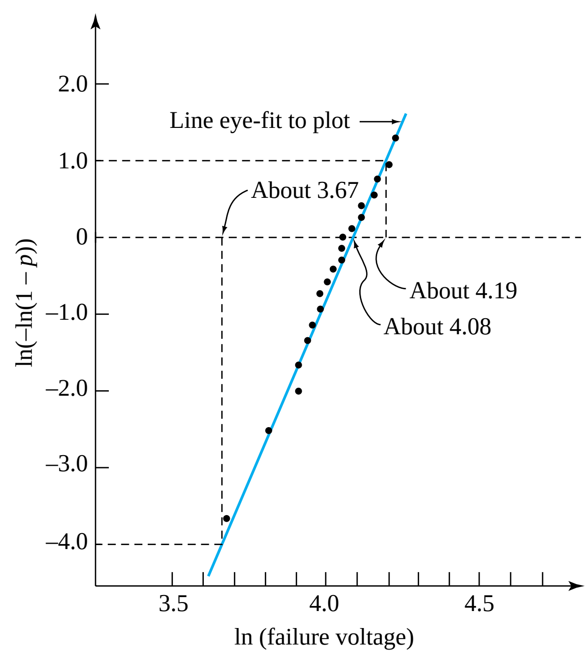


Figure 5.25 0-threshold Weibull plot for insulation failure voltages

Plotting form (5.35) is quite popular in reliability and materials applications. It is common to see such Weibull plots made on special **Weibull paper** (see Figure 5.26). This is graph paper whose scales are constructed so that instead of using plotting positions (5.35) on regular graph paper, one can use plotting positions

$$\left(x_i, \frac{i - .5}{n}\right)$$

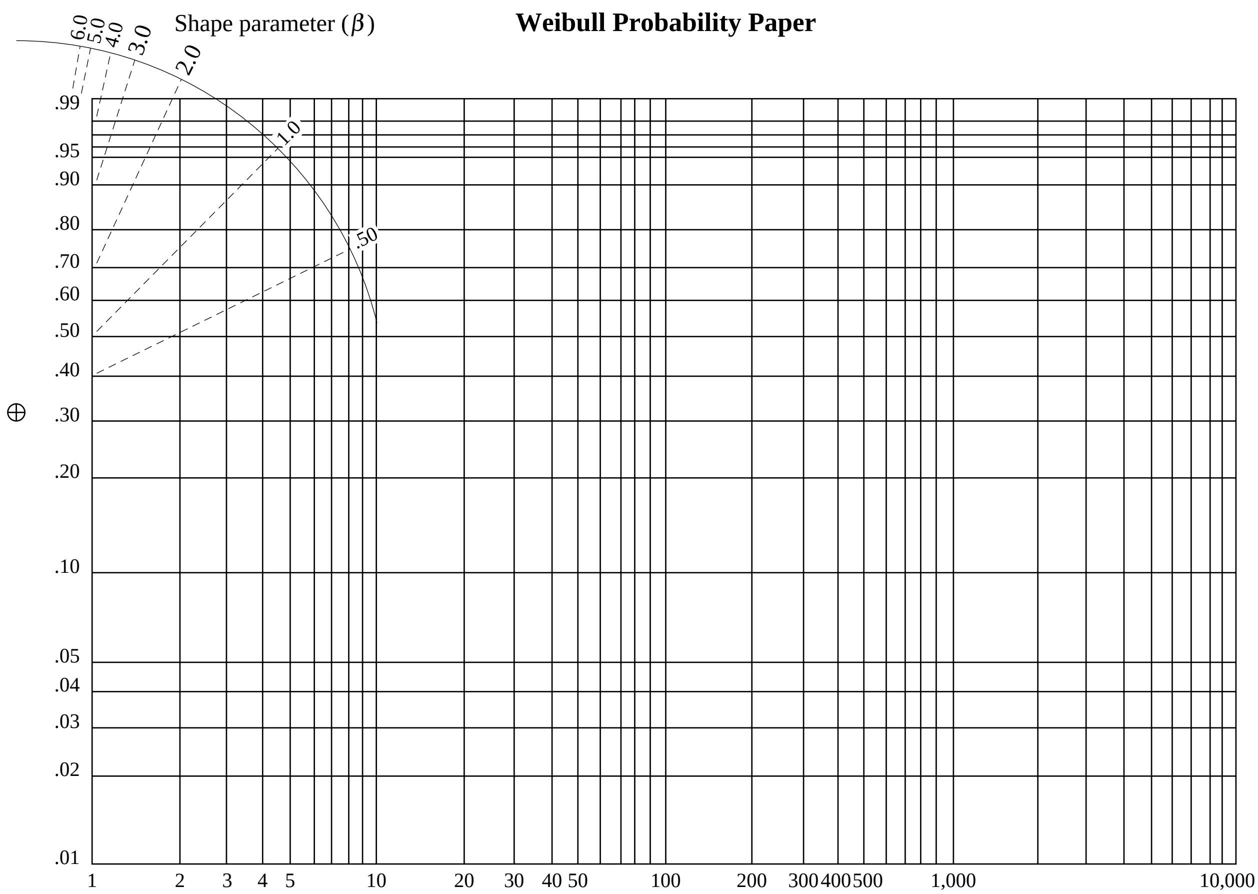


Figure 5.26 Weibull probability paper

for data $x_1 \leq x_2 \leq \cdots \leq x_n$. (The determination of β is even facilitated through the inclusion of the protractor in the upper left corner.) Further, standard statistical packages often have built-in facilities for Weibull plotting of this type.

It should be emphasized that the idea of probability plotting is a quite general one. Its use has been illustrated here only with normal, exponential, and Weibull distributions. But remember that for any probability density $f(x)$, theoretical Q-Q plotting provides a tool for assessing whether the distributional shape portrayed by $f(x)$ might be used in the modeling of a random variable.

Section 3 Exercises

1. What is the practical usefulness of the technique of probability plotting?

2. Explain how an approximate mean μ and standard deviation σ can be read off a plot of standard normal quantiles versus data quantiles.

3. Exercise 3 of Section 3.2 refers to the chemical process yield data of J. S. Hunter given in Exercise 1 of Section 3.1. There you were asked to make a normal plot of those data.

(a) If you have not already done so, use a computer package to make a version of the normal plot.

(b) Use your plot to derive an approximate mean and a standard deviation for the chemical process yields.

4. The article "Statistical Investigation of the Fatigue Life of Deep Groove Ball Bearings" by J. Leiblein and M. Zelen (*Journal of Research of the National Bureau of Standards*, 1956) contains the data given below on the lifetimes of 23 ball bearings. The units are 10^6 revolutions before failure.

17.88, 28.92, 33.00, 41.52, 42.12, 45.60,
48.40, 51.84, 51.96, 54.12, 55.56, 67.80,
68.64, 68.64, 68.88, 84.12, 93.12, 98.64,
105.12, 105.84, 127.92, 128.04, 173.40

(a) Use a normal plot to assess how well a normal distribution fits these data. Then determine if bearing load life can be better represented by a normal distribution if life is expressed on the log scale. (Take the natural logarithms of these data and make a normal plot.) What mean and standard deviation would you use in a normal description of log load life? For these parameters, what are the .05 quantiles of ln(life) and of life?

(b) Use the method of display (5.35) and investigate whether the Weibull distribution might be used to describe bearing load life. If a Weibull description is sensible, read appropriate parameter values from the plot. Then use the form of the Weibull cumulative probability function given in Section 5.2 to find the .05 quantile of the bearing load life distribution.

5. The data here are from the article "Fiducial Bounds on Reliability for the Two-Parameter Negative Exponential Distribution," by F. Grubbs (*Technometrics*, 1971). They are the mileages at first failure for 19 military personnel carriers.

162, 200, 271, 320, 393, 508, 539, 629,
706, 777, 884, 1008, 1101, 1182, 1462,
1603, 1984, 2355, 2880

(a) Make a histogram of these data. How would you describe its shape?

(b) Plot points (5.31) and make an exponential probability plot for these data. Does it appear that the exponential distribution can be used to model the mileage to failure of this kind of vehicle? In Example 15, a threshold service time of 7.5 seconds was suggested by a similar exponential probability plot. Does the present plot give a strong indication of the need for a threshold mileage larger than 0 if an exponential distribution is to be used here?

5.4 Joint Distributions and Independence

Most applications of probability to engineering statistics involve not one but several random variables. In some cases, the application is intrinsically multivariate. It then makes sense to think of more than one process variable as subject to random influences and to evaluate probabilities associated with them in combination. Take, for example, the assembly of a ring bearing with nominal inside diameter 1.00 in. on a rod with nominal diameter .99 in. If

$$X = \text{the ring bearing inside diameter}$$

$$Y = \text{the rod diameter}$$

one might be interested in

$$P[X < Y] = P[\text{there is an interference in assembly}]$$

which involves *both* variables.

But even when a situation is univariate, samples larger than size 1 are essentially always used in engineering applications. The n data values in a sample are usually thought of as subject to chance causes and their simultaneous behavior must then be modeled. The methods of Sections 5.1 and 5.2 are capable of dealing with only a single random variable at a time. They must be generalized to create methods for describing several random variables simultaneously.

Entire books are written on various aspects of the simultaneous modeling of many random variables. This section can give only a brief introduction to the topic. Considering first the comparatively simple case of jointly discrete random variables, the topics of joint and marginal probability functions, conditional distributions, and independence are discussed primarily through reference to simple bivariate examples. Then the analogous concepts of joint and marginal probability density functions, conditional distributions, and independence for jointly continuous random variables are introduced. Again, the discussion is carried out primarily through reference to a bivariate example.

5.4.1 Describing Jointly Discrete Random Variables

For several discrete variables the device typically used to specify probabilities is a **joint probability function**. The two-variable version of this is defined next.

Definition 19

A **joint probability function** for discrete random variables X and Y is a nonnegative function $f(x, y)$, giving the probability that (simultaneously) X takes the value x and Y takes the value y. That is,

$$f(x, y) = P[X = x \text{ and } Y = y]$$

Example 17
(*Example 1 revisited*)

The Joint Probability Distribution of Two Bolt Torques

Return again to the situation of Brenny, Christensen, and Schneider and the measuring of bolt torques on the face plates of a heavy equipment component to the nearest integer. With

$$X = \text{the next torque recorded for bolt 3}$$

$$Y = \text{the next torque recorded for bolt 4}$$

Example 17 *(continued)*

the data displayed in Table 3.4 (see page 74) and Figure 3.9 suggest, for example, that a sensible value for $P[X = 18 \text{ and } Y = 18]$ might be $\frac{1}{34}$, the relative frequency of this pair in the data set. Similarly, the assignments

$$P[X = 18 \text{ and } Y = 17] = \frac{2}{34}$$

$$P[X = 14 \text{ and } Y = 9] = 0$$

also correspond to observed relative frequencies.

If one is willing to accept the whole set of relative frequencies defined by the students' data as defining probabilities for X and Y, these can be collected conveniently in a two-dimensional table specifying a joint probability function for X and Y. This is illustrated in Table 5.10. (To avoid clutter, 0 entries in the table have been left blank.)

Table 5.10
$f(x, y)$ for the Bolt Torque Problem

$y \backslash x$	11	12	13	14	15	16	17	18	19	20
20								2/34	2/34	1/34
19							2/34			
18			1/34	1/34			1/34	1/34	1/34	
17					2/34	1/34	1/34	2/34		
16				1/34	2/34	2/34			2/34	
15	1/34	1/34			3/34					
14					1/34			2/34		
13					1/34					

Properties of a joint probability function for X and Y

The probability function given in tabular form in Table 5.10 has two properties that are necessary for mathematical consistency. These are that the $f(x, y)$ values are each in the interval $[0, 1]$ and that they total to 1. By summing up just some of the $f(x, y)$ values, probabilities associated with X and Y being configured in patterns of interest are obtained.

Example 17 *(continued)*

Consider using the joint distribution given in Table 5.10 to evaluate

$$P[X \geq Y],$$

$$P[|X - Y| \leq 1],$$

$$\text{and } P[X = 17]$$

Take first $P[X \geq Y]$, the probability that the measured bolt 3 torque is at least as big as the measured bolt 4 torque. Figure 5.27 indicates with asterisks which possible combinations of x and y lead to bolt 3 torque at least as large as the bolt 4 torque. Referring to Table 5.10 and adding up those entries corresponding to the cells that contain asterisks,

$$\begin{aligned} P[X \geq Y] &= f(15,13) + f(15,14) + f(15,15) + f(16,16) \\ &\quad + f(17,17) + f(18,14) + f(18,17) + f(18,18) \\ &\quad + f(19,16) + f(19,18) + f(20,20) \\ &= \frac{1}{34} + \frac{1}{34} + \frac{3}{34} + \frac{2}{34} + \cdots + \frac{1}{34} = \frac{17}{34} \end{aligned}$$

Similar reasoning allows evaluation of $P[|X - Y| \leq 1]$—the probability that the bolt 3 and 4 torques are within 1 ft lb of each other. Figure 5.28 shows combinations of x and y with an absolute difference of 0 or 1. Then, adding probabilities corresponding to these combinations,

$$\begin{aligned} P[|X - Y| \leq 1] &= f(15,14) + f(15,15) + f(15,16) + f(16,16) \\ &\quad + f(16,17) + f(17,17) + f(17,18) + f(18,17) \\ &\quad + f(18,18) + f(19,18) + f(19,20) + f(20,20) = \frac{18}{34} \end{aligned}$$

$y \backslash x$	11	12	13	14	15	16	17	18	19	20
20										*
19									*	*
18								*	*	*
17							*	*	*	*
16						*	*	*	*	*
15					*	*	*	*	*	*
14				*	*	*	*	*	*	*
13			*	*	*	*	*	*	*	*

Figure 5.27 Combinations of bolt 3 and bolt 4 torques with $x \geq y$

$y \backslash x$	11	12	13	14	15	16	17	18	19	20
20									*	*
19								*	*	*
18							*	*	*	
17						*	*	*		
16					*	*	*			
15				*	*	*				
14			*	*	*					
13		*	*	*						

Figure 5.28 Combinations of bolt 3 and bolt 4 torques with $|x - y| \leq 1$

Example 17 (*continued*)

Finally, $P[X = 17]$, the probability that the measured bolt 3 torque is 17 ft lb, is obtained by adding down the $x = 17$ column in Table 5.10. That is,

$$P[X = 17] = f(17, 17) + f(17, 18) + f(17, 19) = \frac{1}{34} + \frac{1}{34} + \frac{2}{34} = \frac{4}{34}$$

Finding marginal probability functions using a bivariate joint probability function

In bivariate problems like the present one, one can add down columns in a two-way table giving $f(x, y)$ to get values for the probability function of X, $f_X(x)$. And one can add across rows in the same table to get values for the probability function of Y, $f_Y(y)$. One can then write these sums in the *margins* of the two-way table. So it should not be surprising that probability distributions for individual random variables obtained from their joint distribution are called **marginal distributions**. A formal statement of this terminology in the case of two discrete variables is next.

Definition 20

The individual probability functions for discrete random variables X and Y with joint probability function $f(x, y)$ are called **marginal probability functions**. They are obtained by summing $f(x, y)$ values over all possible values of the other variable. In symbols, the marginal probability function for X is

$$f_X(x) = \sum_y f(x, y)$$

and the marginal probability function for Y is

$$f_Y(y) = \sum_x f(x, y)$$

Example 17 (*continued*)

Table 5.11 is a copy of Table 5.10, augmented by the addition of marginal probabilities for X and Y. Separating off the margins from the two-way table produces tables of marginal probabilities in the familiar format of Section 5.1. For example, the marginal probability function of Y is given separately in Table 5.12.

Table 5.11
Joint and Marginal Probabilities for X and Y

$y \setminus x$	11	12	13	14	15	16	17	18	19	20	$f_Y(y)$
20								2/34	2/34	1/34	5/34
19							2/34				2/34
18			1/34	1/34			1/34	1/34	1/34		5/34
17					2/34	1/34	1/34	2/34			6/34
16				1/34	2/34	2/34			2/34		7/34
15	1/34	1/34			3/34						5/34
14					1/34			2/34			3/34
13					1/34						1/34
$f_X(x)$	1/34	1/34	1/34	2/34	9/34	3/34	4/34	7/34	5/34	1/34	

Table 5.12
Marginal Probability Function for Y

y	$f_Y(y)$
13	1/34
14	3/34
15	5/34
16	7/34
17	6/34
18	5/34
19	2/34
20	5/34

Getting marginal probability functions from joint probability functions raises the natural question whether the process can be reversed. That is, if $f_X(x)$ and $f_Y(y)$ are known, is there then exactly one choice for $f(x, y)$? The answer to this question is "No." Figure 5.29 shows two quite different bivariate joint distributions that nonetheless possess the same marginal distributions. The marked difference between the distributions in Figure 5.29 has to do with the *joint*, rather than individual, behavior of X and Y.

5.4.2 Conditional Distributions and Independence for Discrete Random Variables

When working with several random variables, it is often useful to think about what is expected of one of the variables, given the values assumed by all others. For

Distribution 1

$y \backslash x$	1	2	3	
3	.4	0	0	.4
2	0	.4	0	.4
1	0	0	.2	.2
	.4	.4	.2	

Distribution 2

$y \backslash x$	1	2	3	
3	.16	.16	.08	.4
2	.16	.16	.08	.4
1	.08	.08	.04	.2
	.4	.4	.2	

Figure 5.29 Two different joint distributions with the same marginal distributions

example, in the bolt (X) torque situation, a technician who has just loosened bolt 3 and measured the torque as 15 ft lb ought to have expectations for bolt 4 torque (Y) somewhat different from those described by the marginal distribution in Table 5.12. After all, returning to the data in Table 3.4 that led to Table 5.10, the relative frequency distribution of bolt 4 torques for those components with bolt 3 torque of 15 ft lb is as in Table 5.13. Somehow, knowing that $X = 15$ ought to make a probability distribution for Y like the relative frequency distribution in Table 5.13 more relevant than the marginal distribution given in Table 5.12.

Table 5.13
Relative Frequency Distribution for Bolt 4 Torques When Bolt 3 Torque Is 15 ft lb

y, Torque (ft lb)	Relative Frequency
13	$1/9$
14	$1/9$
15	$3/9$
16	$2/9$
17	$2/9$

The theory of probability makes allowance for this notion of "distribution of one variable knowing the values of others" through the concept of conditional distributions. The two-variable version of this is defined next.

Definition 21

For discrete random variables X and Y with joint probability function $f(x, y)$, the **conditional probability function of X given $Y = y$** is the function of x

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{\sum_x f(x, y)}$$

The **conditional probability function of Y given $X = x$** is the function of y

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{\sum_{y} f(x, y)}$$

Comparing Definitions 20 and 21

The conditional probability function for X given $Y = y$

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)} \tag{5.36}$$

and

The conditional probability function for Y given $X = x$

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \tag{5.37}$$

Finding conditional distributions from a joint probability function

And formulas (5.36) and (5.37) are perfectly sensible. Equation (5.36) says that starting from $f(x, y)$ given in a two-way table and looking only at the row specified by $Y = y$, the appropriate (conditional) distribution for X is given by the probabilities in that row (the $f(x, y)$ values) divided by their sum ($f_Y(y) = \sum_x f(x, y)$), so that they are renormalized to total to 1. Similarly, equation (5.37) says that looking only at the column specified by $X = x$, the appropriate conditional distribution for Y is given by the probabilities in that column divided by their sum.

Example 17 (*continued*)

To illustrate the use of equations (5.36) and (5.37), consider several of the conditional distributions associated with the joint distribution for the bolt 3 and bolt 4 torques, beginning with the conditional distribution for Y given that $X = 15$.

From equation (5.37),

$$f_{Y|X}(y \mid 15) = \frac{f(15, y)}{f_X(15)}$$

Referring to Table 5.11, the marginal probability associated with $X = 15$ is $\frac{9}{34}$. So dividing values in the $X = 15$ column of that table by $\frac{9}{34}$, leads to the conditional distribution for Y given in Table 5.14. Comparing this to Table 5.13, indeed formula (5.37) produces a conditional distribution that agrees with intuition.

Example 17 *(continued)*

Table 5.14
The Conditional Probability Function for Y Given $X = 15$

y	$f_{Y\|X}(y \mid 15)$
13	$\left(\frac{1}{34}\right) \div \left(\frac{9}{34}\right) = \frac{1}{9}$
14	$\left(\frac{1}{34}\right) \div \left(\frac{9}{34}\right) = \frac{1}{9}$
15	$\left(\frac{3}{34}\right) \div \left(\frac{9}{34}\right) = \frac{3}{9}$
16	$\left(\frac{2}{34}\right) \div \left(\frac{9}{34}\right) = \frac{2}{9}$
17	$\left(\frac{2}{34}\right) \div \left(\frac{9}{34}\right) = \frac{2}{9}$

Next consider $f_{Y|X}(y \mid 18)$ specified by

$$f_{Y|X}(y \mid 18) = \frac{f(18, y)}{f_X(18)}$$

Consulting Table 5.11 again leads to the conditional distribution for Y given that $X = 18$, shown in Table 5.15. Tables 5.14 and 5.15 confirm that the conditional distributions of Y given $X = 15$ and given $X = 18$ are quite different. For example, knowing that $X = 18$ would on the whole make one expect Y to be larger than when $X = 15$.

Table 5.15
The Conditional Probability Function for Y Given $X = 18$

y	$f_{Y\|X}(y \mid 18)$
14	2/7
17	2/7
18	1/7
20	2/7

To make sure that the meaning of equation (5.36) is also clear, consider the conditional distribution of the bolt 3 torque (X) given that the bolt 4 torque is 20

$(Y = 20)$. In this situation, equation (5.36) gives

$$f_{X|Y}(x \mid 20) = \frac{f(x, 20)}{f_Y(20)}$$

(Conditional probabilities for X are the values in the $Y = 20$ row of Table 5.11 divided by the marginal $Y = 20$ value.) Thus, $f_{X|Y}(x \mid 20)$ is given in Table 5.16.

Table 5.16
The Conditional Probability Function for X Given $Y = 20$

x	$f_{X\|Y}(x \mid 20)$
18	$\left(\frac{2}{34}\right) \div \left(\frac{5}{34}\right) = \frac{2}{5}$
19	$\left(\frac{2}{34}\right) \div \left(\frac{5}{34}\right) = \frac{2}{5}$
20	$\left(\frac{1}{34}\right) \div \left(\frac{5}{34}\right) = \frac{1}{5}$

The bolt torque example has the feature that the conditional distributions for Y given various possible values for X differ. Further, these are not generally the same as the marginal distribution for Y. X provides some information about Y, in that depending upon its value there are differing probability assessments for Y. Contrast this with the following example.

Example 18

Random Sampling Two Bolt 4 Torques

Suppose that the 34 bolt 4 torques obtained by Brenny, Christensen, and Schneider and given in Table 3.4 are written on slips of paper and placed in a hat. Suppose further that the slips are mixed, one is selected, the corresponding torque is noted, and the slip is replaced. Then the slips are again mixed, another is selected, and the second torque is noted. Define the two random variables

$$U = \text{the value of the first torque selected}$$

and

$$V = \text{the value of the second torque selected}$$

Example 18 *(continued)*

Intuition dictates that (in contrast to the situation of X and Y in Example 17) the variables U and V don't furnish any information about each other. Regardless of what value U takes, the relative frequency distribution of bolt 4 torques in the hat is appropriate as the (conditional) probability distribution for V, and vice versa. That is, not only do U and V share the common marginal distribution given in Table 5.17 but it is also the case that for all u and v, both

$$f_{U|V}(u \mid v) = f_U(u) \tag{5.38}$$

and

$$f_{V|U}(v \mid u) = f_V(v) \tag{5.39}$$

Equations (5.38) and (5.39) say that the marginal probabilities in Table 5.17 also serve as conditional probabilities. They also specify how joint probabilities for U and V must be structured. That is, rewriting the left-hand side of equation (5.38) using expression (5.36),

$$\frac{f(u, v)}{f_V(v)} = f_U(u)$$

That is,

$$f(u, v) = f_U(u)\, f_V(v) \tag{5.40}$$

(The same logic applied to equation (5.39) also leads to equation (5.40).) Expression (5.40) says that joint probability values for U and V are obtained by multiplying corresponding marginal probabilities. Table 5.18 gives the joint probability function for U and V.

Table 5.17
The Common Marginal Probability Function for U and V

u or v	$f_U(u)$ or $f_V(v)$
13	1/34
14	3/34
15	5/34
16	7/34
17	6/34
18	5/34
19	2/34
20	5/35

Table 5.18
Joint Probabilities for U and V

$v \setminus u$	13	14	15	16	17	18	19	20	$f_V(v)$
20	$\frac{5}{(34)^2}$	$\frac{15}{(34)^2}$	$\frac{25}{(34)^2}$	$\frac{35}{(34)^2}$	$\frac{30}{(34)^2}$	$\frac{25}{(34)^2}$	$\frac{10}{(34)^2}$	$\frac{25}{(34)^2}$	$5/34$
19	$\frac{2}{(34)^2}$	$\frac{6}{(34)^2}$	$\frac{10}{(34)^2}$	$\frac{14}{(34)^2}$	$\frac{12}{(34)^2}$	$\frac{10}{(34)^2}$	$\frac{4}{(34)^2}$	$\frac{10}{(34)^2}$	$2/34$
18	$\frac{5}{(34)^2}$	$\frac{15}{(34)^2}$	$\frac{25}{(34)^2}$	$\frac{35}{(34)^2}$	$\frac{30}{(34)^2}$	$\frac{25}{(34)^2}$	$\frac{10}{(34)^2}$	$\frac{25}{(34)^2}$	$5/34$
17	$\frac{6}{(34)^2}$	$\frac{18}{(34)^2}$	$\frac{30}{(34)^2}$	$\frac{42}{(34)^2}$	$\frac{36}{(34)^2}$	$\frac{30}{(34)^2}$	$\frac{12}{(34)^2}$	$\frac{30}{(34)^2}$	$6/34$
16	$\frac{7}{(34)^2}$	$\frac{21}{(34)^2}$	$\frac{35}{(34)^2}$	$\frac{49}{(34)^2}$	$\frac{42}{(34)^2}$	$\frac{35}{(34)^2}$	$\frac{14}{(34)^2}$	$\frac{35}{(34)^2}$	$7/34$
15	$\frac{5}{(34)^2}$	$\frac{15}{(34)^2}$	$\frac{25}{(34)^2}$	$\frac{35}{(34)^2}$	$\frac{30}{(34)^2}$	$\frac{25}{(34)^2}$	$\frac{10}{(34)^2}$	$\frac{25}{(34)^2}$	$5/34$
14	$\frac{3}{(34)^2}$	$\frac{9}{(34)^2}$	$\frac{15}{(34)^2}$	$\frac{21}{(34)^2}$	$\frac{18}{(34)^2}$	$\frac{15}{(34)^2}$	$\frac{6}{(34)^2}$	$\frac{15}{(34)^2}$	$3/34$
13	$\frac{1}{(34)^2}$	$\frac{3}{(34)^2}$	$\frac{5}{(34)^2}$	$\frac{7}{(34)^2}$	$\frac{6}{(34)^2}$	$\frac{5}{(34)^2}$	$\frac{2}{(34)^2}$	$\frac{5}{(34)^2}$	$1/34$
$f_U(u)$	$1/34$	$3/34$	$5/34$	$7/34$	$6/34$	$5/34$	$2/34$	$5/34$	

Example 18 suggests that the intuitive notion that several random variables are unrelated might be formalized in terms of all conditional distributions being equal to their corresponding marginal distributions. Equivalently, it might be phrased in terms of joint probabilities being the products of corresponding marginal probabilities. The formal mathematical terminology is that of **independence** of the random variables. The definition for the two-variable case is next.

Definition 22

Discrete random variables X and Y are called **independent** if their joint probability function $f(x, y)$ is the product of their respective marginal probability functions. That is, independence means that

$$f(x, y) = f_X(x)\, f_Y(y) \qquad \text{for all } x, y \tag{5.41}$$

If formula (5.41) does not hold, the variables X and Y are called **dependent**.

(Formula (5.41) does imply that conditional distributions are all equal to their corresponding marginals, so that the definition does fit its "unrelatedness" motivation.)

U and V in Example 18 are independent, whereas X and Y in Example 17 are dependent. Further, the two joint distributions depicted in Figure 5.29 give an example of a highly dependent joint distribution (the first) and one of independence (the second) that have the same marginals.

Independence of observations in statistical studies

The notion of independence is a fundamental one. When it is sensible to model random variables as independent, *great mathematical simplicity results*. Where engineering data are being collected in an analytical context, and care is taken to make sure that all obvious physical causes of **carryover effects** that might influence successive observations **are minimal**, an assumption of independence between observations is often appropriate. And in enumerative contexts, **relatively small** (compared to the population size) **simple random samples** yield observations that can typically be considered as at least approximately independent.

Example 18
(continued)

Again consider putting bolt torques on slips of paper in a hat. The method of torque selection described earlier for producing U and V is not simple random sampling. Simple random sampling as defined in Section 2.2 is *without-replacement* sampling, not the *with-replacement* sampling method used to produce U and V. Indeed, if the first slip is not replaced before the second is selected, the probabilities in Table 5.18 are not appropriate for describing U and V. For example, if no replacement is done, since only one slip is labeled 13 ft lb, one clearly wants

$$f(13, 13) = P[U = 13 \text{ and } V = 13] = 0$$

not the value

$$f(13, 13) = \frac{1}{(34)^2}$$

indicated in Table 5.18. Put differently, if no replacement is done, one clearly wants to use

$$f_{V|U}(13 \mid 13) = 0$$

rather than the value

$$f_{V|U}(13 \mid 13) = f_V(13) = \frac{1}{34}$$

which would be appropriate if sampling is done with replacement. Simple random sampling doesn't lead to exactly independent observations.

But suppose that instead of containing 34 slips labeled with torques, the hat contained 100×34 slips labeled with torques with relative frequencies as in Table 5.17. Then even if sampling is done without replacement, the probabilities developed earlier for U and V (and placed in Table 5.18) remain at least *approximately* valid. For example, with 3,400 slips and using without-replacement sampling,

$$f_{V|U}(13 \mid 13) = \frac{99}{3{,}399}$$

is appropriate. Then, using the fact that

$$f_{V|U}(v \mid u) = \frac{f(u, v)}{f_U(u)}$$

so that

$$f(u, v) = f_{V|U}(v \mid u)\, f_U(u)$$

without replacement, the assignment

$$f(13, 13) = \frac{99}{3{,}399} \cdot \frac{1}{34}$$

is appropriate. But the point is that

$$\frac{99}{3{,}399} \approx \frac{1}{34}$$

and so

$$f(13, 13) \approx \frac{1}{34} \cdot \frac{1}{34}$$

For this hypothetical situation where the population size $N = 3{,}400$ is much larger than the sample size $n = 2$, independence is a suitable approximate description of observations obtained using simple random sampling.

Where several variables are both independent and have the same marginal distributions, some additional jargon is used.

Definition 23

If random variables $X_1, X_2, \ldots, X_n$ all have the same marginal distribution and are independent, they are termed **iid** or **independent and identically distributed**.

For example, the joint distribution of U and V given in Table 5.18 shows U and V to be iid random variables.

When can observations be modeled as iid?

The standard statistical examples of iid random variables are successive measurements taken from a stable process and the results of random sampling with replacement from a single population. The question of whether an iid model is appropriate in a statistical application thus depends on whether or not the data-generating mechanism being studied can be thought of as conceptually equivalent to these.

5.4.3 Describing Jointly Continuous Random Variables (*Optional*)

All that has been said about joint description of discrete random variables has its analog for continuous variables. Conceptually and computationally, however, the jointly continuous case is more challenging. Probability density functions replace probability functions, and multivariate calculus substitutes for simple arithmetic. So most readers will be best served in the following introduction to multivariate continuous distributions by reading for the main ideas and not getting bogged down in details.

The counterpart of a joint probability function, the device that is commonly used to specify probabilities for several continuous random variables, is a **joint probability density**. The two-variable version of this is defined next.

Definition 24

A **joint probability density** for continuous random variables X and Y is a nonnegative function $f(x, y)$ with

$$\int\int f(x,y)\,dx\,dy = 1$$

and such that for any region $\mathcal{R}$, one is willing to assign

$$P[(X,Y) \in \mathcal{R}] = \int\int_{\mathcal{R}} f(x,y)\,dx\,dy \tag{5.42}$$

Instead of summing values of a probability function to find probabilities for a discrete distribution, equation (5.42) says (as in Section 5.2) to integrate a probability density. The new complication here is that the integral is two-dimensional. But it is still possible to draw on intuition developed in mechanics, remembering that this is exactly the sort of thing that is done to specify mass distributions in several dimensions. (Here, mass is probability, and the total mass is 1.)

Example 19
(*Example 15 revisited*)

Residence Hall Depot Counter Service Time and a Continuous Joint Distribution

Consider again the depot service time example. As Section 5.3 showed, the students' data suggest an exponential model with $\alpha = 16.5$ for the random variable

S = the excess (over a 7.5 sec threshold) time required to complete the next sale

Imagine that the true value of S will be measured with a (very imprecise) analog stopwatch, producing the random variable

$$R = \text{the measured excess service time}$$

Consider the function of two variables

$$f(s, r) = \begin{cases} \frac{1}{16.5} e^{-s/16.5} \frac{1}{\sqrt{2\pi(.25)}} e^{-(r-s)^2/2(.25)} & \text{for } s > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.43}$$

as a potential joint probability density for S and R. Figure 5.30 provides a representation of $f(s, r)$, sketched as a surface in three-dimensional space.

As defined in equation (5.43), $f(s, r)$ is nonnegative, and its integral (the volume underneath the surface sketched in Figure 5.30 over the region in the (s, r)-plane where s is positive) is

$$\begin{aligned} \int\int f(s, r)\, ds\, dr &= \int_0^\infty \int_{-\infty}^\infty \frac{1}{16.5\sqrt{2\pi(.25)}} e^{-(s/16.5)-((r-s)^2/2(.25))}\, dr\, ds \\ &= \int_0^\infty \frac{1}{16.5} e^{-s/16.5} \left\{ \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi(.25)}} e^{-(r-s)^2/2(.25)}\, dr \right\} ds \\ &= \int_0^\infty \frac{1}{16.5} e^{-s/16.5}\, ds \\ &= 1 \end{aligned}$$

(The integral in braces is 1 because it is the integral of a normal density with

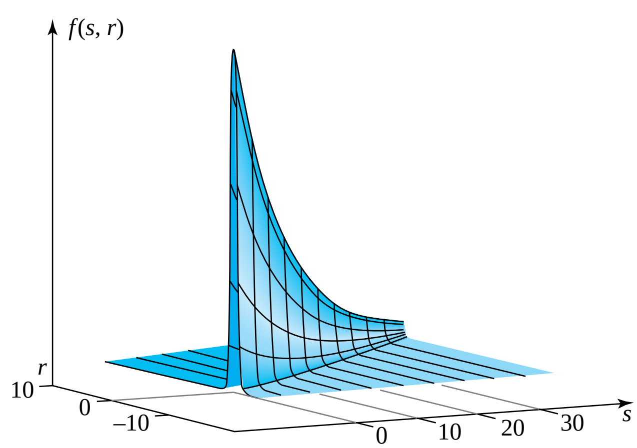


Figure 5.30 A joint probability density for S and R

Example 19 (*continued*)

mean s and standard deviation .5.) Thus, equation (5.43) specifies a mathematically legitimate joint probability density.

To illustrate the use of a joint probability density in finding probabilities, first consider evaluating $P[R > S]$. Figure 5.31 shows the region in the (s, r)-plane where $f(s, r) > 0$ and $r > s$. It is over this region that one must integrate in order to evaluate $P[R > S]$. Then,

$$\begin{aligned}
P[R > S] &= \iint_{\mathcal{R}} f(s, r)\, ds\, dr \\
&= \int_0^\infty \int_s^\infty f(s, r)\, dr\, ds \\
&= \int_0^\infty \frac{1}{16.5} e^{-s/16.5} \left\{ \int_s^\infty \frac{1}{\sqrt{2\pi(.25)}} e^{-(r-s)^2/2(.25)}\, dr \right\} ds \\
&= \int_0^\infty \frac{1}{16.5} e^{-s/16.5} \left\{ \frac{1}{2} \right\} ds \\
&= \frac{1}{2}
\end{aligned}$$

(once again using the fact that the integral in braces is a normal (mean s and standard deviation .5) probability).

As a second example, consider the problem of evaluating $P[S > 20]$. Figure 5.32 shows the region over which $f(s, r)$ must be integrated in order to evaluate $P[S > 20]$. Then,

$$\begin{aligned}
P[S > 20] &= \iint_{\mathcal{R}} f(s, r)\, ds\, dr \\
&= \int_{20}^\infty \int_{-\infty}^\infty f(s, r)\, dr\, ds \\
&= \int_{20}^\infty \frac{1}{16.5} e^{-s/16.5} \left\{ \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi(.25)}} e^{-(r-s)^2/2(.25)}\, dr \right\} ds \\
&= \int_{20}^\infty \frac{1}{16.5} e^{-s/16.5}\, ds \\
&= e^{-20/16.5} \\
&\approx .30
\end{aligned}$$

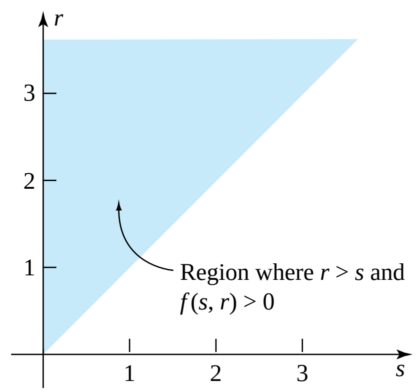


Figure 5.31 Region where $f(s, r) > 0$ and $r > s$

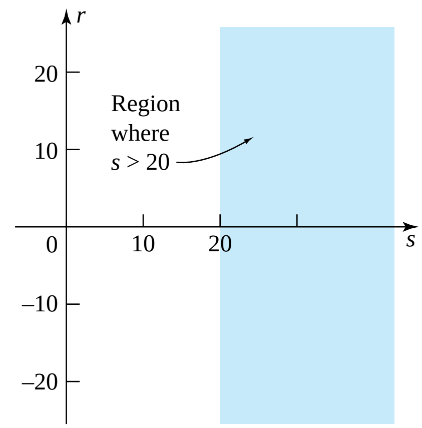


Figure 5.32 Region where $f(s, r) > 0$ and $s > 20$

The last part of the example essentially illustrates the fact that for X and Y with joint density $f(x, y)$,

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(t, y)\, dy\, dt$$

This is a statement giving the cumulative probability function for X. Differentiation with respect to x shows that a marginal probability density for X is obtained from the joint density by integrating out y. Putting this in the form of a definition gives the following.

Definition 25 The individual probability densities for continuous random variables X and Y with joint probability density $f(x, y)$ are called **marginal probability densities**. They are obtained by integrating $f(x, y)$ over all possible values of the other variable. In symbols, the marginal probability density function for X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \tag{5.44}$$

and the marginal probability density function for Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{5.45}$$

Compare Definitions 20 and 25 (page 282). The same kind of thing is done for jointly continuous variables to find marginal distributions as for jointly discrete variables, except that integration is substituted for summation.

Example 19 *(continued)*

Starting with the joint density specified by equation (5.43), it is possible to arrive at reasonably explicit expressions for the marginal densities for S and R. First considering the density of S, Definition 25 declares that for $s > 0$,

$$f_S(s) = \int_{-\infty}^{\infty} \frac{1}{16.5} e^{-s/16.5} \left\{ \frac{1}{\sqrt{2\pi(.25)}} e^{-(r-s)^2/2(.25)} \right\} dr$$

$$= \frac{1}{16.5} e^{-s/16.5}$$

Further, since $f(s, r)$ is 0 for negative s, if $s < 0$,

$$f_S(s) = \int_{-\infty}^{\infty} 0 \, dr = 0$$

That is, the form of $f(s, r)$ was chosen so that (as suggested by Example 15) S has an exponential distribution with mean $\alpha = 16.5$.

The determination of $f_R(r)$ is conceptually no different than the determination of $f_S(s)$, but the details are more complicated. Some work (involving completion of a square in the argument of the exponential function and recognition of an integral as a normal probability) will show the determined reader that for any r,

$$f_R(r) = \int_0^{\infty} \frac{1}{16.5\sqrt{2\pi(.25)}} e^{-(s/16.5)-((r-s)^2/2(.25))} \, ds$$

$$= \frac{1}{16.5} \left(1 - \Phi\left(\frac{1}{33} - 2r\right)\right) \exp\left(\frac{1}{2{,}178} - \frac{r}{16.5}\right) \qquad \textbf{(5.46)}$$

where, as usual, Φ is the standard normal cumulative probability function. A graph of $f_R(r)$ is given in Figure 5.33.

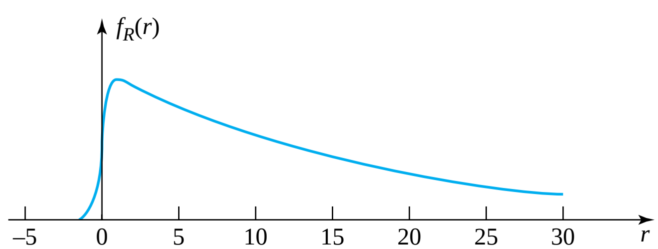


Figure 5.33 Marginal probability density for R

The marginal density for R derived from equation (5.43) does not belong to any standard family of distributions. Indeed, there is generally no guarantee that the process of finding marginal densities from a joint density will produce expressions for the densities even as explicit as that in display (5.46).

5.4.4 Conditional Distributions and Independence for Continuous Random Variables (*Optional*)

In order to motivate the definition for conditional distributions derived from a joint probability density, consider again Definition 21 (page 284). For jointly discrete variables X and Y, the conditional distribution for X given $Y = y$ is specified by holding y fixed and treating $f(x, y)$ as a probability function for X after appropriately renormalizing it—i.e., seeing that its values total to 1. The analogous operation for two jointly continuous variables is described next.

Definition 26

For continuous random variables X and Y with joint probability density $f(x, y)$, the **conditional probability density function of** X **given** $Y = y$**,** is the function of x

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{\displaystyle\int_{-\infty}^{\infty} f(x, y)\, dx}$$

The **conditional probability density function of** Y **given** $X = x$ is the function of y

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{\displaystyle\int_{-\infty}^{\infty} f(x, y)\, dy}$$

Definitions 25 and 26 lead to

Conditional probability density for X given $Y = y$

$$\boxed{f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}} \tag{5.47}$$

and

Conditional probability density for Y given $X = x$

$$\boxed{f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}} \tag{5.48}$$

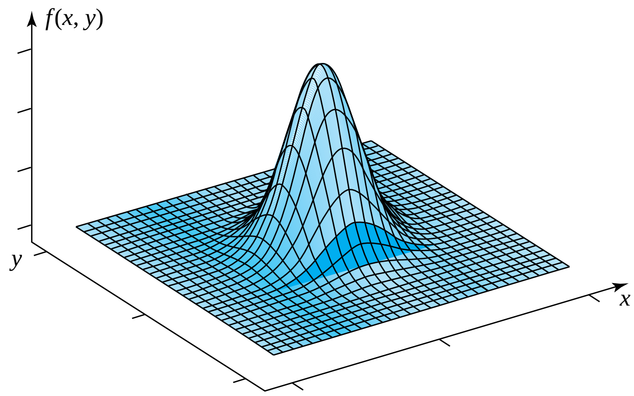


Figure 5.34 A Joint probability density $f(x, y)$ and the shape of a conditional density for X given a value of Y

Geometry of conditional densities

Expressions (5.47) and (5.48) are formally identical to the expressions (5.36) and (5.37) relevant for discrete variables. The geometry indicated by equation (5.47) is that the shape of $f_{X|Y}(x \mid y)$ as a function of x is determined by cutting the $f(x, y)$ surface in a graph like that in Figure 5.34 with the $Y = y$-plane. In Figure 5.34, the divisor in equation (5.47) is the area of the shaded figure above the (x, y)-plane below the $f(x, y)$ surface on the $Y = y$ plane. That division serves to produce a function of x that will integrate to 1. (Of course, there is a corresponding geometric story told for the conditional distribution of Y given $X = x$ in expression (5.48)).

Example 19 (*continued*)

In the service time example, it is fairly easy to recognize the conditional distribution of R given $S = s$ as having a familiar form. For $s > 0$, applying expression (5.48),

$$f_{R|S}(r \mid s) = \frac{f(s, r)}{f_S(s)} = f(s, r) \div \left(\frac{1}{16.5} e^{-s/16.5} \right)$$

which, using equation (5.43), gives

$$f_{R|S}(r \mid s) = \frac{1}{\sqrt{2\pi(.25)}} e^{-(r-s)^2/2(.25)} \qquad \textbf{(5.49)}$$

That is, given that $S = s$, the conditional distribution of R is normal with mean s and standard deviation .5.

This realization is consistent with the bell-shaped cross sections of $f(s, r)$ shown in Figure 5.30. The form of $f_{R|S}(r \mid s)$ given in equation (5.49) says that the measured excess service time is the true excess service time plus a normally distributed measurement error that has mean 0 and standard deviation .5.

It is evident from expression (5.49) (or from the way the positions of the bell-shaped contours on Figure 5.30 vary with s) that the variables S and R ought to be called dependent. After all, knowing that $S = s$ gives the value of R except for a normal error of measurement with mean 0 and standard deviation .5. On the other hand, had it been the case that all conditional distributions of R given $S = s$ were the same (and equal to the marginal distribution of R), S and R should be called independent. The notion of unchanging conditional distributions, all equal to their corresponding marginal, is equivalently and more conveniently expressed in terms of the joint probability density factoring into a product of marginals. The formal version of this for two variables is next.

Definition 27

Continuous random variables X and Y are called **independent** if their joint probability density function $f(x, y)$ is the product of their respective marginal probability densities. That is, independence means that

$$f(x, y) = f_X(x)\, f_Y(y) \qquad \text{for all } x, y \tag{5.50}$$

If expression (5.50) does not hold, the variables X and Y are called **dependent**.

Expression (5.50) is formally identical to expression (5.41), which appeared in Definition 22 for discrete variables. The type of factorization given in these expressions provides great mathematical convenience.

It remains in this section to remark that the concept of *iid random variables* introduced in Definition 23 is as relevant to continuous cases as it is to discrete ones. In statistical contexts, it can be appropriate where analytical problems are free of carryover effects and in enumerative problems where (relatively) small simple random samples are being described.

Example 20
(Example 15 revisited)

Residence Hall Depot Counter Service Times and iid Variables

Returning once more to the service time example of Jenkins, Milbrath, and Worth, consider the next two excess service times encountered,

S_1 = the first/next excess (over a threshold of 7.5 sec) time required to complete a postage stamp sale at the residence hall service counter

S_2 = the second excess service time

To the extent that the service process is physically stable (i.e., excess service times can be thought of in terms of sampling with replacement from a single population), an iid model seems appropriate for S_1 and S_2. Treating excess service times as

Example 20 (*continued*)

marginally exponential with mean $\alpha = 16.5$ thus leads to the joint density for S_1 and S_2:

$$f(s_1, s_2) = \begin{cases} \dfrac{1}{(16.5)^2} e^{-(s_1+s_2)/16.5} & \text{if } s_1 > 0 \text{ and } s_2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

Section 4 Exercises

1. Explain in qualitative terms what it means for two random variables X and Y to be independent. What advantage is there when X and Y can be described as independent?

2. Quality audit records are kept on numbers of major and minor failures of circuit packs during burn-in of large electronic switching devices. They indicate that for a device of this type, the random variables

$$X = \text{the number of major failures}$$

and

$$Y = \text{the number of minor failures}$$

can be described at least approximately by the accompanying joint distribution.

$y \backslash x$	0	1	2
0	.15	.05	.01
1	.10	.08	.01
2	.10	.14	.02
3	.10	.08	.03
4	.05	.05	.03

(a) Find the marginal probability functions for both X and Y—$f_X(x)$ and $f_Y(y)$.
(b) Are X and Y independent? Explain.
(c) Find the mean and variance of X—EX and $\text{Var}\,X$.
(d) Find the mean and variance of Y—EY and $\text{Var}\,Y$.
(e) Find the conditional probability function for Y, given that $X = 0$—i.e., that there are no major circuit pack failures. (That is, find $f_{Y|X}(y \mid 0)$.) What is the mean of this conditional distribution?

3. A laboratory receives four specimens having identical appearances. However, it is possible that (a single unknown) one of the specimens is contaminated with a toxic material. The lab must test the specimens to find the toxic specimen (if in fact one is contaminated). The testing plan first put forth by the laboratory staff is to test the specimens one at a time, stopping when (and if) a contaminated specimen is found.

Define two random variables

$$X = \text{the number of contaminated specimens}$$

and

$$Y = \text{the number of specimens tested}$$

Let $p = P[X = 0]$ and therefore $P[X = 1] = 1 - p$.

(a) Give the conditional distributions of Y given $X = 0$ and $X = 1$ for the staff's initial testing plan. Then use them to determine the joint probability function of X and Y. (Your joint distribution will involve p, and you may simply fill out tables like the accompanying ones.)

y	$f_{Y\|X}(y \mid 0)$
1	
2	
3	
4	

y	$f_{Y\|X}(y \mid 1)$
1	
2	
3	
3	

$y \backslash x$	0	1
	$f(x, y)$	
1		
2		
3		
4		

(b) Based on your work in (a), find the marginal distribution of Y. What is EY, the mean number of specimens tested using the staff's original plan?

(c) A second testing method devised by the staff involves testing composite samples of material taken from possibly more than one of the original specimens. By initially testing a composite of all four specimens, then (if the first test reveals the presence of toxic material) following up with a composite of two, and then an appropriate single specimen, it is possible to do the lab's job in one test if $X = 0$, and in three tests if $X = 1$. Suppose that because testing is expensive, it is desirable to hold the number of tests to a minimum. For what values of p is this second method preferable to the first? (*Hint:* What is EY for this second method?)

4. A machine element is made up of a rod and a ring bearing. The rod must fit through the bearing. Model

$$X = \text{the diameter of the rod}$$

and

$$Y = \text{the inside diameter of the ring bearing}$$

as independent random variables, X uniform on (1.97, 2.02) and Y uniform on (2.00, 2.06). ($f_X(x) = 1/.05$ for $1.97 < x < 2.02$, while $f_X(x) = 0$ otherwise. Similarly, $f_Y(y) = 1/.06$ for $2.00 < y < 2.06$, while $f_Y(y) = 0$ otherwise.) With this model, do the following:

(a) Write out the joint probability density for X and Y. (It will be positive only when $1.97 < x < 2.02$ and $2.00 < y < 2.06$.)

(b) Evaluate $P[Y - X < 0]$, the probability of an interference in assembly.

5. Suppose that a pair of random variables have the joint probability density

$$f(x, y) = \begin{cases} 4x(1-y) & \text{if } 0 \le x \le 1 \\ & \text{and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal probability densities for X and Y. What is the mean of X?

(b) Are X and Y independent? Explain.

(c) Evaluate $P[X + 2Y \ge 1]$.

(d) Find the conditional probability density for X given that $Y = .5$. (Find $f_{X|Y}(x \mid .5)$.) What is the mean of this (conditional) distribution?

6. An engineering system consists of two subsystems operating independently of each other. Let

$$X = \text{the time till failure of the first subsystem}$$

and

$$Y = \text{the time till failure of the second subsystem}$$

Suppose that X and Y are independent exponential random variables each with mean $\alpha = 1$ (in appropriate time units).

(a) Write out the joint probability density for X and Y. Be sure to state carefully where the density is positive and where it is 0.

Suppose first that the system is a *series system* (i.e., one that fails when either of the subsystems fail).

(b) The probability that the system is still functioning at time $t > 0$ is then

$$P[X > t \text{ and } Y > t]$$

Find this probability using your answer to (a). (What region in the (x, y)-plane corresponds to the possibility that the system is still functioning at time t?)

(c) If one then defines the random variable

$$T = \text{the time until the system fails}$$

the cumulative probability function for T is

$$F(t) = 1 - P[X > t \text{ and } Y > t]$$

so that your answer to (b) can be used to find the distribution for T. Use your answer to (b) and some differentiation to find the probability density for T. What kind of distribution does T have? What is its mean?

Suppose now that the system is a *parallel system* (i.e., one that fails only when both subsystems fail).

(d) The probability that the system has failed by time t is

$$P[X \le t \text{ and } Y \le t]$$

Find this probability using your answer to part (a).

(e) Now, as before, let T be the time until the system fails. Use your answer to (d) and some differentiation to find the probability density for T. Then calculate the mean of T.

5.5 Functions of Several Random Variables

The last section introduced the mathematics used to simultaneously model several random variables. An important engineering use of that material is in the analysis of system outputs that are functions of random inputs.

This section studies how the variation seen in an output random variable depends upon that of the variables used to produce it. It begins with a few comments on what is possible using exact methods of mathematical analysis. Then the simple and general tool of simulation is introduced. Next, formulas for means and variances of linear combinations of random variables and the related propagation of error formulas are presented. Last is the pervasive central limit effect, which often causes variables to have approximately normal distributions.

5.5.1 The Distribution of a Function of Random Variables

The problem considered in this section is this. Given a joint distribution for the random variables $X, Y, \ldots, Z$ and a function $g(x, y, \ldots, z)$, the object is to predict the behavior of the random variable

$$U = g(X, Y, \ldots, Z) \tag{5.51}$$

In some special simple cases, it is possible to figure out exactly what distribution U inherits from $X, Y, \ldots, Z$.

Example 21 The Distribution of the Clearance Between Two Mating Parts with Randomly Determined Dimensions

Suppose that a steel plate with nominal thickness .15 in. is to rest in a groove of nominal width .155 in., machined on the surface of a steel block. A lot of plates has been made and thicknesses measured, producing the relative fre-

Table 5.19
Relative Frequency Distribution of Plate Thicknesses

Plate Thickness (in.)	Relative Frequency
.148	.4
.149	.3
.150	.3

Table 5.20
Relative Frequency Distribution of Slot Widths

Slot Width (in.)	Relative Frequency
.153	.2
.154	.2
.155	.4
.156	.2

quency distribution in Table 5.19; a relative frequency distribution for the slot widths measured on a lot of machined blocks is given in Table 5.20.

If a plate is randomly selected and a block is separately randomly selected, a natural joint distribution for the random variables

$$X = \text{the plate thickness}$$
$$Y = \text{the slot width}$$

is one of independence, where the marginal distribution of X is given in Table 5.19 and the marginal distribution of Y is given in Table 5.20. That is, Table 5.21 gives a plausible joint probability function for X and Y.

A variable derived from X and Y that is of substantial potential interest is the clearance involved in the plate/block assembly,

$$U = Y - X$$

Notice that taking the extremes represented in Tables 5.19 and 5.20, U is guaranteed to be at least $.153 - .150 = .003$ in. but no more than $.156 - .148 = .008$ in. In fact, much more than this can be said. Looking at Table 5.21, one can see that the diagonals of entries (lower left to upper right) all correspond to the same value of $Y - X$. Adding probabilities on those diagonals produces the distribution of U given in Table 5.22.

Example 21 *(continued)*

Table 5.21
Marginal and Joint Probabilities for X and Y

$y \setminus x$	.148	.149	.150	$f_Y(y)$
.156	.08	.06	.06	.2
.155	.16	.12	.12	.4
.154	.08	.06	.06	.2
.153	.08	.06	.06	.2
$f_X(x)$	.4	.3	.3	

Table 5.22
The Probability Function for the Clearance $U = Y - X$

u	$f(u)$
.003	.06
.004	$.12 = .06 + .06$
.005	$.26 = .08 + .06 + .12$
.006	$.26 = .08 + .12 + .06$
.007	$.22 = .16 + .06$
.008	.08

Example 21 involves a very simple discrete joint distribution and a very simple function g—namely, $g(x, y) = y - x$. In general, exact complete solution of the problem of finding the distribution of $U = g(X, Y, \ldots, Z)$ is not practically possible. Happily, for many engineering applications of probability, approximate and/or partial solutions suffice to answer the questions of practical interest. The balance of this section studies methods of producing these approximate and/or partial descriptions of the distribution of U, beginning with a brief look at simulation-based methods.

5.5.2 Simulations to Approximate the Distribution of $U = g(X, Y, \ldots, Z)$

Simulation for independent $X, Y, \ldots, Z$

Many computer programs and packages can be used to produce pseudorandom values, intended to behave as if they were realizations of independent random variables following user-chosen marginal distributions. If the model for $X, Y, \ldots, Z$ is one of independence, it is then a simple matter to generate a simulated value for each of $X, Y, \ldots, Z$ and plug those into g to produce a simulated value for U. If this process is repeated a number of times, a relative frequency distribution for these simulated values of U is developed. One might reasonably use this relative frequency distribution to approximate an exact distribution for U.

Example 22

Uncertainty in the Calculated Efficiency of an Air Solar Collector

The article "Thermal Performance Representation and Testing of Air Solar Collectors" by Bernier and Plett (*Journal of Solar Energy Engineering*, May 1988) considers the testing of air solar collectors. Its analysis of thermal performance based on enthalpy balance leads to the conclusion that under inward leakage conditions, the thermal efficiency of a collector can be expressed as

$$\begin{aligned}\text{Efficiency} &= \frac{M_o C(T_o - T_i) + (M_o - M_i)C(T_i - T_a)}{GA} \\ &= \frac{C}{GA}\left(M_o T_o - M_i T_i - (M_o - M_i)T_a\right) \end{aligned} \tag{5.52}$$

where

C = air specific heat (J/kg°C)

G = global irradiance incident on the plane of the collector (W/m^2)

A = collector gross area (m^2)

M_i = inlet mass flowrate (kg/s)

M_o = outlet mass flowrate (kg/s)

T_a = ambient temperature (°C)

T_i = collector inlet temperature (°C)

T_o = collector outlet temperature (°C)

The authors further give some uncertainty values associated with each of the terms appearing on the right side of equation (5.52) for an example set of measured values of the variables. These are given in Table 5.23.

Table 5.23
Reported Uncertainties in the Measured Inputs to Collector Efficiency

Variable	Measured Value	Uncertainty
C	1003.8	1.004 (i.e., ± .1%)
G	1121.4	33.6 (i.e., ± 3%)
A	1.58	.005
M_i	.0234	.00035 (i.e., ± 1.5%)
M_o	.0247	.00037 (i.e., ± 1.5%)
T_a	−13.22	.5
T_i	−6.08	.5
T_o	24.72	.5*

*This value is not given explicitly in the article.

Example 22 *(continued)*

Plugging the measured values from Table 5.23 into formula (5.52) produces a measured efficiency of about .44. But how good is the .44 value? That is, how do the uncertainties associated with the measured values affect the reliability of the .44 figure? Should you think of the calculated solar collector efficiency as .44 plus or minus .001, or plus or minus .1, or what?

One way of approaching this is to ask the related question, "What would be the standard deviation of *Efficiency* if all of C through T_o were independent random variables with means approximately equal to the measured values and standard deviations related to the uncertainties as, say, half of the uncertainty values?" (This "two sigma" interpretation of uncertainty appears to be at least close to the intention in the original article.)

Printout 1 is from a MINITAB session in which 100 normally distributed realizations of variables C through T_o were generated (using means equal to measured values and standard deviations equal to half of the corresponding uncertainties) and the resulting efficiencies calculated. (The routine under the "Calc/Random Data/Normal" menu was used to generate the realizations of C through T_o. The "Calc/Calculator" menu was used to combine these values according to equation (5.52). Then routines under the "Stat/Basic Statistics/Describe" and "Graph/Character Graphs/Stem-and-Leaf" menus were used to produce the summaries of the simulated efficiencies.) The simulation produced a roughly bell-shaped distribution of calculated efficiencies, possessing a mean value of approximately .437 and standard deviation of about .009. Evidently, if one continues with the understanding that uncertainty means something like "2 standard deviations," an uncertainty of about .02 is appropriate for the nominal efficiency figure of .44.

Printout 1 Simulation of Solar Collector Efficiency

WWW

```
Descriptive Statistics


Variable            N      Mean    Median    TrMean     StDev   SE Mean
Efficien          100   0.43729   0.43773   0.43730   0.00949   0.00095

Variable      Minimum   Maximum        Q1        Q3
Efficien      0.41546   0.46088   0.43050   0.44426



Character Stem-and-Leaf Display


Stem-and-leaf of Efficien   N  = 100
Leaf Unit = 0.0010


    5    41  58899
   10    42  22334
   24    42  66666777788999
   39    43  001112233333444
  (21)   43  555556666777889999999
   40    44  00000011122333444
```

```
23   44 555556667788889
 8   45 023344
 2   45 7
 1   46 0
```

The beauty of Example 22 is the ease with which a simulation can be employed to approximate the distribution of U. But the method is so powerful and easy to use that some cautions need to be given about the application of this whole topic before going any further.

Practical cautions

Be careful not to expect more than is sensible from a derived probability distribution ("exact" or approximate) for

$$U = g(X, Y, \ldots, Z)$$

The output distribution can be no more realistic than are the assumptions used to produce it (i.e., the form of the joint distribution and the form of the function $g(x, y, \ldots, z)$). It is all too common for people to apply the methods of this section using a g representing some approximate physical law and U some measurable physical quantity, only to be surprised that the variation in U observed in the real world is *substantially larger* than that predicted by methods of this section. The fault lies not with the methods, but with the naivete of the user. Approximate physical laws are just that, often involving so-called constants that aren't constant, using functional forms that are too simple, and ignoring the influence of variables that aren't obvious or easily measured. Further, although independence of $X, Y, \ldots, Z$ is a very convenient mathematical property, its use is not always justified. When it is inappropriately used as a model assumption, it can produce an inappropriate distribution for U. For these reasons, think of the methods of this section as useful but likely to provide only a *best-case* picture of the variation you should expect to see.

5.5.3 Means and Variances for Linear Combinations of Random Variables

For engineering purposes, it often suffices to know the mean and variance for U given in formula (5.51) (as opposed to knowing the whole distribution of U). When this is the case and g is linear, there are explicit formulas for these.

Proposition 1

If $X, Y, \ldots, Z$ are n independent random variables and $a_0, a_1, a_2, \ldots, a_n$ are $n + 1$ constants, then the random variable $U = a_0 + a_1X + a_2Y + \cdots + a_nZ$ has mean

$$EU = a_0 + a_1EX + a_2EY + \cdots + a_nEZ \quad \textbf{(5.53)}$$

and variance

$$\boxed{\operatorname{Var} U = a_1^2 \operatorname{Var} X + a_2^2 \operatorname{Var} Y + \cdots + a_n^2 \operatorname{Var} Z} \qquad \textbf{(5.54)}$$

Formula (5.53) actually holds regardless of whether or not the variables $X, Y, \ldots, Z$ are independent, and although formula (5.54) does depend upon independence, there is a generalization of it that can be used even if the variables are dependent. However, the form of Proposition 1 given here is adequate for present purposes.

One type of application in which Proposition 1 is immediately useful is that of geometrical tolerancing problems, where it is applied with $a_0 = 0$ and the other a_i's equal to plus and minus 1's.

Example 21 *(continued)*

Consider again the situation of the clearance involved in placing a steel plate in a machined slot on a steel block. With X, Y, and U being (respectively) the plate thickness, slot width, and clearance, means and variances for these variables can be calculated from Tables 5.19, 5.20, and 5.22, respectively. The reader is encouraged to verify that

$$EX = .1489 \quad \text{and} \quad \operatorname{Var} X = 6.9 \times 10^{-7}$$
$$EY = .1546 \quad \text{and} \quad \operatorname{Var} Y = 1.04 \times 10^{-6}$$

Now, since

$$U = Y - X = (-1)X + 1Y$$

Proposition 1 can be applied to conclude that

$$EU = -1EX + 1EY = -.1489 + .1546 = .0057 \text{ in.}$$
$$\operatorname{Var} U = (-1)^2 6.9 \times 10^{-7} + (1)^2 1.04 \times 10^{-6} = 1.73 \times 10^{-6}$$

so that

$$\sqrt{\operatorname{Var} U} = .0013 \text{ in.}$$

It is worth the effort to verify that the mean and standard deviation of the clearance produced using Proposition 1 agree with those obtained using the distribution of U given in Table 5.22 and the formulas for the mean and variance given in Section 5.1. The advantage of using Proposition 1 is that if all that is needed are EU and $\operatorname{Var} U$, there is no need to go through the intermediate step of deriving the distribution of U. The calculations via Proposition 1 use only characteristics of the marginal distributions.

Another particularly important use of Proposition 1 concerns n iid random variables where each a_i is $\frac{1}{n}$. That is, in cases where random variables $X_1, X_2, \ldots, X_n$ are conceptually equivalent to random selections (*with* replacement) from a single numerical population, Proposition 1 tells how the mean and variance of the random variable

$$\overline{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n$$

are related to the population parameters μ and σ^2. For independent variables $X_1, X_2, \ldots, X_n$ with common mean μ and variance σ^2, Proposition 1 shows that

The mean of an average of n iid random variables

$$E\overline{X} = \frac{1}{n}EX_1 + \frac{1}{n}EX_2 + \cdots + \frac{1}{n}EX_n = n\left(\frac{1}{n}\mu\right) = \mu \qquad \textbf{(5.55)}$$

and

The variance of an average of n iid random variables

$$\begin{aligned}\operatorname{Var}\overline{X} &= \left(\tfrac{1}{n}\right)^2 \operatorname{Var}X_1 + \left(\tfrac{1}{n}\right)^2 \operatorname{Var}X_2 + \cdots + \left(\tfrac{1}{n}\right)^2 \operatorname{Var}X_n \\ &= n\left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}\end{aligned} \qquad \textbf{(5.56)}$$

Since σ^2/n is decreasing in n, equations (5.55) and (5.56) give the reassuring picture of $\overline{X}$ having a probability distribution centered at the population mean μ, with spread that decreases as the sample size increases.

Example 23 (*Example 15 revisited*)

The Expected Value and Standard Deviation for a Sample Mean Service Time

To illustrate the application of formulas (5.55) and (5.56), consider again the stamp sale service time example. Suppose that the exponential model with $\alpha = 16.5$ that was derived in Example 15 for excess service times continues to be appropriate and that several more postage stamp sales are observed and excess service times noted. With

$S_i =$ the excess (over a 7.5 sec threshold) time required to complete the ith additional stamp sale

Example 23 (*continued*)

consider what means and standard deviations are associated with the probability distributions of the sample average, $\overline{S}$, of first the next 4 and then the next 100 excess service times.

$S_1, S_2, \ldots, S_{100}$ are, to the extent that the service process is physically stable, reasonably modeled as independent, identically distributed, exponential random variables with mean $\alpha = 16.5$. The exponential distribution with mean $\alpha = 16.5$ has variance equal to $\alpha^2 = (16.5)^2$. So, using formulas (5.55) and (5.56), for the first 4 additional service times,

$$E\overline{S} = \alpha = 16.5 \text{ sec}$$

$$\sqrt{\operatorname{Var}\overline{S}} = \sqrt{\frac{\alpha^2}{4}} = 8.25 \text{ sec}$$

Then, for the first 100 additional service times,

$$E\overline{S} = \alpha = 16.5 \text{ sec}$$

$$\sqrt{\operatorname{Var}\overline{S}} = \sqrt{\frac{\alpha^2}{100}} = 1.65 \text{ sec}$$

Notice that going from a sample size of 4 to a sample size of 100 decreases the standard deviation of $\overline{S}$ by a factor of 5 $(= \sqrt{\frac{100}{4}})$.

Relationships (5.55) and (5.56), which perfectly describe the random behavior of $\overline{X}$ under random sampling with replacement, are also approximate descriptions of the behavior of $\overline{X}$ under simple random sampling in enumerative contexts. (Recall Example 18 and the discussion about the approximate independence of observations resulting from simple random sampling of large populations.)

5.5.4 The Propagation of Error Formulas

Proposition 1 gives exact values for the mean and variance of $U = g(X, Y, \ldots, Z)$ only when g is linear. It doesn't seem to say anything about situations involving nonlinear functions like the one specified by the right-hand side of expression (5.52) in the solar collector example. But it is often possible to obtain useful approximations to the mean and variance of U by applying Proposition 1 to a first-order multivariate Taylor expansion of a not-too-nonlinear g. That is, if g is reasonably well-behaved, then for $x, y, \ldots, z$ (respectively) close to $EX, EY, \ldots, EZ$,

$$\left.\begin{aligned} g(x, y, \ldots, z) \approx g(EX, EY, \ldots, EZ) &+ \frac{\partial g}{\partial x} \cdot (x - EX) + \frac{\partial g}{\partial y} \cdot (y - EY) \\ &+ \cdots + \frac{\partial g}{\partial z} cdot (z - EZ) \end{aligned}\right\} \quad \textbf{(5.57)}$$

where the partial derivatives are evaluated at $(x, y, \ldots, z) = (EX, EY, \ldots, EZ)$. Now the right side of approximation (5.57) is linear in $x, y, \ldots, z$. Thus, if the variances of $X, Y, \ldots, Z$ are small enough so that with high probability, $X, Y, \ldots, Z$ are such that approximation (5.57) is effective, one might think of plugging $X, Y, \ldots, Z$ into expression (5.57) and applying Proposition 1, thus winding up with approximations for the mean and variance of $U = g(X, Y, \ldots, Z)$.

Proposition 2 (*The Propagation of Error Formulas*)

If $X, Y, \ldots, Z$ are independent random variables and g is well-behaved, for small enough variances $\text{Var}\, X, \text{Var}\, Y, \ldots, \text{Var}\, Z$, the random variable $U = g(X, Y, \ldots, Z)$ has approximate mean

$$\boxed{EU \approx g(EX, EY, \ldots, EZ)} \qquad \textbf{(5.58)}$$

and approximate variance

$$\boxed{\text{Var}\, U \approx \left(\frac{\partial g}{\partial x}\right)^2 \text{Var}\, X + \left(\frac{\partial g}{\partial y}\right)^2 \text{Var}\, Y + \cdots + \left(\frac{\partial g}{\partial z}\right)^2 \text{Var}\, Z} \qquad \textbf{(5.59)}$$

Formulas (5.58) and (5.59) are often called the **propagation of error** or **transmission of variance** formulas. They describe how variability or error is propagated or transmitted through an exact mathematical function.

Comparison of Propositions 1 and 2 shows that when g is exactly linear, expressions (5.58) and (5.59) reduce to expressions (5.53) and (5.54), respectively. (a_1 through a_n are the partial derivatives of g in the case where $g(x, y, \ldots, z) = a_0 + a_1 x + a_2 y + \cdots + a_n z$.) Proposition 2 is purposely vague about when the approximations (5.58) and (5.59) will be adequate for engineering purposes. Mathematically inclined readers will not have much trouble constructing examples where the approximations are quite poor. But often in engineering applications, expressions (5.58) and (5.59) are at least of the right order of magnitude and certainly better than not having any usable approximations.

Example 24

A Simple Electrical Circuit and the Propagation of Error

Figure 5.35 is a schematic of an assembly of three resistors. If R_1, R_2, and R_3 are the respective resistances of the three resistors making up the assembly, standard theory says that

$$R = \text{the assembly resistance}$$

Example 24 (*continued*)

is related to R_1, R_2, and R_3 by

$$R = R_1 + \frac{R_2 R_3}{R_2 + R_3} \tag{5.60}$$

A large lot of resistors is manufactured and has a mean resistance of 100 Ω with a standard deviation of resistance of 2 Ω. If three resistors are taken at random from this lot and assembled as in Figure 5.35, consider what formulas (5.58) and (5.59) suggest for an approximate mean and an approximate standard deviation for the resulting assembly resistance.

The g involved here is $g(r_1, r_2, r_3) = r_1 + \dfrac{r_2 r_3}{r_2 + r_3}$, so

$$\frac{\partial g}{\partial r_1} = 1$$

$$\frac{\partial g}{\partial r_2} = \frac{(r_2 + r_3)r_3 - r_2 r_3}{(r_2 + r_3)^2} = \frac{r_3^2}{(r_2 + r_3)^2}$$

$$\frac{\partial g}{\partial r_3} = \frac{(r_2 + r_3)r_2 - r_2 r_3}{(r_2 + r_3)^2} = \frac{r_2^2}{(r_2 + r_3)^2}$$

Also, R_1, R_2, and R_3 are approximately independent with means 100 and standard deviations 2. Then formulas (5.58) and (5.59) suggest that the probability distribution inherited by R has mean

$$ER \approx g(100, 100, 100) = 100 + \frac{(100)(100)}{100 + 100} = 150\ \Omega$$

and variance

$$\text{Var}\, R \approx (1)^2(2)^2 + \left(\frac{(100)^2}{(100 + 100)^2}\right)^2 (2)^2 + \left(\frac{(100)^2}{(100 + 100)^2}\right)^2 (2)^2 = 4.5$$

so that the standard deviation inherited by R is

$$\sqrt{\text{Var}\, R} \approx \sqrt{4.5} = 2.12\ \Omega$$

As something of a check on how good the 150 Ω and 2.12 Ω values are, 1,000 sets of normally distributed R_1, R_2, and R_3 values with the specified population mean and standard deviation were simulated and resulting values of R calculated via formula (5.60). These simulated assembly resistances had $\overline{R} = 149.80\ \Omega$ and a sample standard deviation of 2.14 Ω. A histogram of these values is given in Figure 5.36.

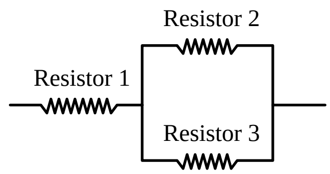


Figure 5.35 Schematic of a simple assembly of three resistors

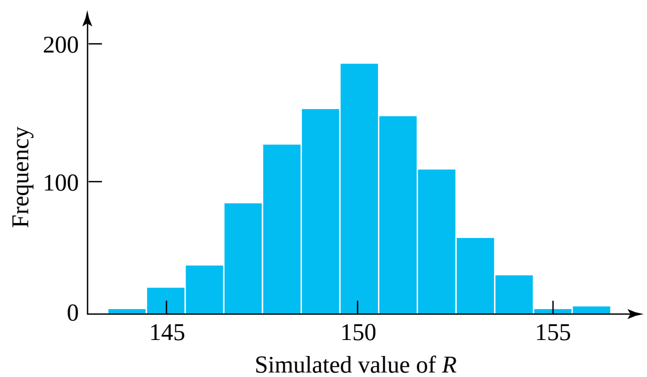


Figure 5.36 Histogram of 1,000 simulated values of R

Example 24 is one to which the cautions following Example 22 (page 307) apply. Suppose you were to actually take a large batch of resistors possessing a mean resistance of 100 Ω and a standard deviation of resistances of 2 Ω, make up a number of assemblies of the type represented in Figure 5.35, and measure the assembly resistances. The standard deviation figures in Example 24 will likely underpredict the variation observed in the assembly resistances.

The propagation of error and simulation methods may do a good job of approximating the (exact) theoretical mean and standard deviation of assembly resistances. But the extent to which the probability model used for assembly resistances can be expected to represent the physical situation is another matter. Equation (5.60) is highly useful, but of necessity only an approximate description of real assemblies. For example, it ignores small but real temperature, inductance, and other second-order physical effects on measured resistance. In addition, although the probability model allows for variation in the resistances of individual components, it does not account for instrument variation or such vagaries of real-world assembly as the quality of contacts achieved when several parts are connected.

In Example 24, the simulation and propagation of error methods produce comparable results. Since the simulation method is so easy to use, why bother to do the calculus and arithmetic necessary to use the propagation of error formulas? One important answer to this question concerns intuition that formula (5.59) provides.

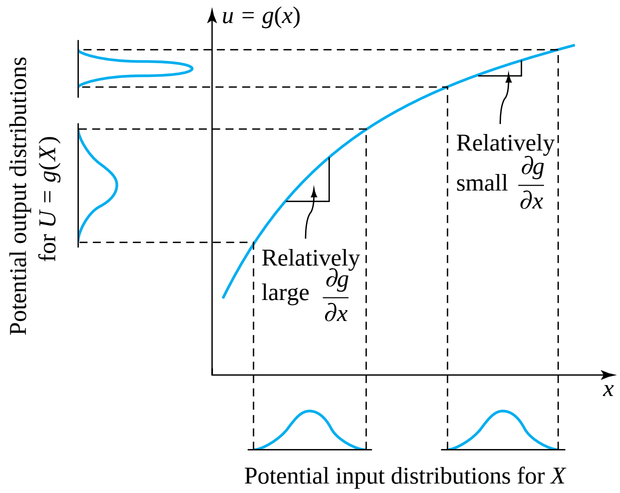


Figure 5.37 Illustration of the Effect of $\frac{\partial g}{\partial x}$ on Var U

The effects of the partial derivatives of g on Var U

Consider first the effect that g's partial derivatives have on Var U. Formula (5.59) implies that depending on the size of $\frac{\partial g}{\partial x}$, the variance of X is either inflated or deflated before becoming an ingredient of Var U. And even though formula (5.59) may not be an exact expression, it provides correct intuition. If a given change in x produces a big change in $g(x, y, \ldots, z)$, the impact Var X has on Var U will be greater than if the change in x produces a small change in $g(x, y, \ldots, z)$. Figure 5.37 is a rough illustration of this point. In the case that $U = g(X)$, two different approximately normal distributions for X with different means but a common variance produce radically different spreads in the distribution of U, due to differing rates of change of g (different derivatives).

Partitioning the variance of U

Then, consider the possibility of partitioning the variance of U into interpretable pieces. Formula (5.59) suggests thinking of (for example)

$$\left(\frac{\partial g}{\partial x}\right)^2 \text{Var}\, X$$

as the contribution the variation in X makes to the variation inherent in U. Comparison of such individual contributions makes it possible to analyze how various potential reductions in input variation can be expected to affect the output variability in U.

Example 22 *(continued)*

Return to the solar collector example. For means of C through T_o taken to be the measured values in Table 5.23 (page 305), and standard deviations of C through T_o equal to half of the uncertainties listed in the same table, formula

(5.59) might well be applied to the calculated efficiency given in formula (5.52). The squared partial derivatives of *Efficiency* with respect to each of the inputs, times the variances of those inputs, are as given in Table 5.24. Thus, the approximate standard deviation for the efficiency variable provided by formula (5.59) is

$$\sqrt{8.28 \times 10^{-5}} \approx .009$$

which agrees quite well with the value obtained earlier via simulation.

What's given in Table 5.24 that doesn't come out of a simulation is some understanding of the *biggest contributors* to the uncertainty. The largest contribution listed in Table 5.24 corresponds to variable G, followed in order by those corresponding to variables M_o, T_o, and T_i. At least for the values of the means used in this example, it is the uncertainties in those variables that principally produce the uncertainty in *Efficiency*. Knowing this gives direction to efforts to improve measurement methods. Subject to considerations of feasibility and cost, measurement of the variable G deserves first attention, followed by measurement of the variables M_o, T_o, and T_i.

Notice, however, that reduction of the uncertainty in G alone to essentially 0 would still leave a total in Table 5.24 of about 4.01×10^{-5} and thus an approximate standard deviation for *Efficiency* of about $\sqrt{4.01 \times 10^{-5}} \approx .006$. Calculations of this kind emphasize the need for reductions in the uncertainties of M_o, T_o, and T_i as well, if dramatic (order of magnitude) improvements in overall uncertainty are to be realized.

Table 5.24
Contributions to the Output Variation in Collector Efficiency

Variable	Contributions to Var Efficiency
C	4.73×10^{-8}
G	4.27×10^{-5}
A	4.76×10^{-7}
M_i	5.01×10^{-7}
M_o	1.58×10^{-5}
T_a	3.39×10^{-8}
T_i	1.10×10^{-5}
T_o	1.22×10^{-5}
Total	8.28×10^{-5}

5.5.5 The Central Limit Effect

One of the most frequently used statistics in engineering applications is the sample mean. Formulas (5.55) and (5.56) relate the mean and variance of the probability distribution of the sample mean to those of a single observation when an iid model is appropriate. One of the most useful facts of applied probability is that if the sample size is reasonably large, it is also possible to approximate the *shape* of the probability distribution of $\overline{X}$, independent of the shape of the underlying distribution of individual observations. That is, there is the following fact:

Proposition 3 (*The Central Limit Theorem*)

If $X_1, X_2, \ldots, X_n$ are iid random variables (with mean μ and variance σ^2), then for large n, the variable $\overline{X}$ is approximately normally distributed. (That is, approximate probabilities for $\overline{X}$ can be calculated using the normal distribution with mean μ and variance σ^2/n.)

A proof of Proposition 3 is outside the purposes of this text. But intuition about the effect is fairly easy to develop through an example.

Example 25 (*Example 2 revisited*)

The Central Limit Effect and the Sample Mean of Tool Serial Numbers

Consider again the example from Section 5.1 involving the last digit of essentially randomly selected serial numbers of pneumatic tools. Suppose now that

W_1 = the last digit of the serial number observed next Monday at 9 A.M.

W_2 = the last digit of the serial number observed the following Monday at 9 A.M.

A plausible model for the pair of random variables W_1, W_2 is that they are independent, each with the marginal probability function

$$f(w) = \begin{cases} .1 & \text{if } w = 0, 1, 2, \ldots, 9 \\ 0 & \text{otherwise} \end{cases} \tag{5.61}$$

that is pictured in Figure 5.38.

Using such a model, it is a straightforward exercise (along the lines of Example 21, page 303) to reason that $\overline{W} = \frac{1}{2}(W_1 + W_2)$ has the probability function given in Table 5.25 and pictured in Figure 5.39.

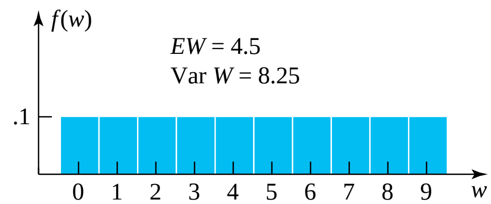


Figure 5.38 Probability histogram for W

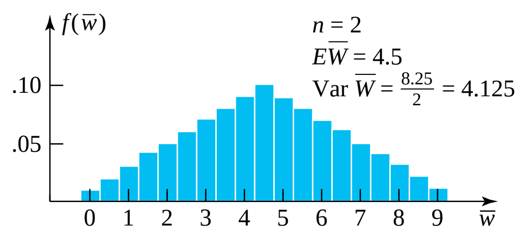


Figure 5.39 Probability histogram for $\overline{W}$ based on $n = 2$

Table 5.25
The Probability Function for $\overline{W}$ for $n = 2$

$\bar{w}$	$f(\bar{w})$	$\bar{w}$	$f(\bar{w})$	$\bar{w}$	$f(\bar{w})$	$\bar{w}$	$f(\bar{w})$	$\bar{w}$	$f(\bar{w})$
0.0	.01	2.0	.05	4.0	.09	6.0	.07	8.0	.03
0.5	.02	2.5	.06	4.5	.10	6.5	.06	8.5	.02
1.0	.03	3.0	.07	5.0	.09	7.0	.05	9.0	.01
1.5	.04	3.5	.08	5.5	.08	7.5	.04		

Comparing Figures 5.38 and 5.39, it is clear that even for a completely flat/uniform underlying distribution of W and the small sample size of $n = 2$, the probability distribution of $\overline{W}$ looks far more bell-shaped than the underlying distribution. It is clear why this is so. As you move away from the mean or central value of $\overline{W}$, there are relatively fewer and fewer combinations of w_1 and w_2 that can produce a given value of $\bar{w}$. For example, to observe $\overline{W} = 0$, you must have $W_1 = 0$ and $W_2 = 0$—that is, you must observe not one but two extreme values. On the other hand, there are ten different combinations of w_1 and w_2 that lead to $\overline{W} = 4.5$.

It is possible to use the same kind of logic leading to Table 5.25 to produce exact probability distributions for $\overline{W}$ based on larger sample sizes n. But such

Example 25 (*continued*)

work is tedious, and for the purpose of indicating roughly how the central limit effect takes over as n gets larger, it is sufficient to approximate the distribution of $\overline{W}$ via simulation for a larger sample size. To this end, 1,000 sets of values for iid variables $W_1, W_2, \ldots, W_8$ (with marginal distribution (5.61)) were simulated and each set averaged to produce 1,000 simulated values of $\overline{W}$ based on $n = 8$. Figure 5.40 is a histogram of these 1,000 values. Notice the bell-shaped character of the plot. (The simulated mean of $\overline{W}$ was $4.508 \approx 4.5 = E\overline{W} = EW$, while the variance of $\overline{W}$ was $1.025 \approx 1.013 = \text{Var}\,\overline{W} = 8.25/8$, in close agreement with formulas (5.55) and (5.56).)

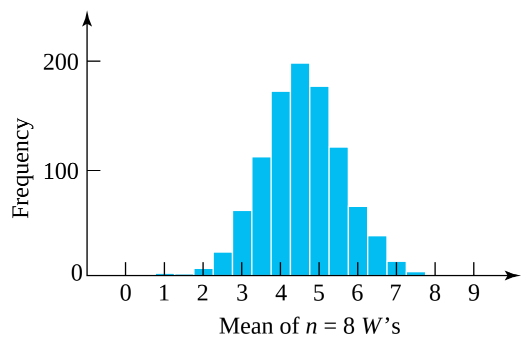


Figure 5.40 Histogram of 1,000 simulated values of $\overline{W}$ based on $n = 8$

Sample size and the central limit effect

What constitutes "large n" in Proposition 3 isn't obvious. The truth of the matter is that what sample size is required before $\overline{X}$ can be treated as essentially normal depends on the shape of the underlying distribution of a single observation. Underlying distributions with decidedly nonnormal shapes require somewhat bigger values of n. But for most engineering purposes, $n \geq 25$ or so is adequate to make $\overline{X}$ essentially normal for the majority of data-generating mechanisms met in practice. (The exceptions are those subject to the occasional production of wildly outlying values.) Indeed, as Example 25 suggests, in many cases $\overline{X}$ is essentially normal for sample sizes much smaller than 25.

The practical usefulness of Proposition 3 is that in many circumstances, only a normal table is needed to evaluate probabilities for sample averages.

Example 23 (*continued*)

Return one more time to the stamp sale time requirements problem and consider observing and averaging the next $n = 100$ excess service times, to produce

$\overline{S}$ = the sample mean time (over a 7.5 sec threshold) required to complete the next 100 stamp sales

And consider approximating $P[\overline{S} > 17]$.

As discussed before, an iid model with marginal exponential $\alpha = 16.5$ distribution is plausible for the individual excess service times, S. Then

$$E\overline{S} = \alpha = 16.5 \text{ sec}$$

and

$$\sqrt{\operatorname{Var} \overline{S}} = \sqrt{\frac{\alpha^2}{100}} = 1.65 \text{ sec}$$

are appropriate for $\overline{S}$, via formulas (5.55) and (5.56). Further, in view of the fact that $n = 100$ is large, the normal probability table may be used to find approximate probabilities for $\overline{S}$. Figure 5.41 shows an approximate distribution for $\overline{S}$ and the area corresponding to $P[\overline{S} > 17]$.

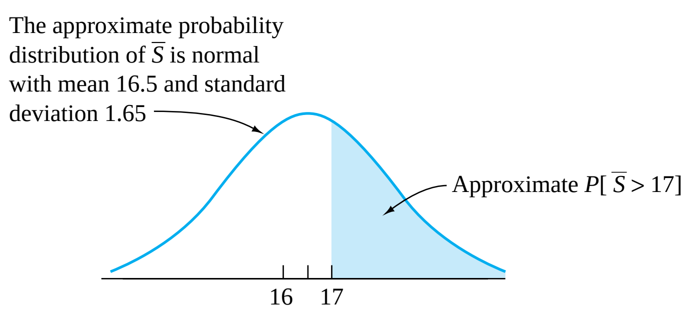


Figure 5.41 Approximate probability distribution for $\overline{S}$ and $P[\overline{S} > 17]$

As always, one must convert to z-values before consulting the standard normal table. In this case, the mean and standard deviation to be used are (respectively) 16.5 sec and 1.65 sec. That is, a z-value is calculated as

$$z = \frac{17 - 16.5}{1.65} = .30$$

so

$$P[\overline{S} > 17] \approx P[Z > .30] = 1 - \Phi(.30) = .38$$

The z-value calculated in the example is an application of the general form

z-value for a sample mean

$$z = \frac{\bar{x} - E\overline{X}}{\sqrt{\operatorname{Var} \overline{X}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \tag{5.62}$$

appropriate when using the central limit theorem to find approximate probabilities for a sample mean. Formula (5.62) is relevant because by Proposition 3, $\overline{X}$ is approximately normal for large n and formulas (5.55) and (5.56) give its mean and standard deviation.

The final example in this section illustrates how the central limit theorem and some idea of a process or population standard deviation can help guide the choice of sample size in statistical applications.

Example 26
(*Example 10 revisited*)

Sampling a Jar-Filling Process

The process of filling food containers, discussed by J. Fisher in his 1983 "Quality Progress" article, appears (from a histogram in the paper) to have an inherent standard deviation of measured fill weights on the order of 1.6 g. Suppose that in order to calibrate a fill-level adjustment knob on such a process, you set the knob and fill a run of n jars. Their sample mean net contents will then serve as an indication of the process mean fill level corresponding to that knob setting. Suppose further that you would like to choose a sample size, n, large enough that a priori there is an 80% chance the sample mean is within .3 g of the actual process mean.

If the filling process can be thought of as physically stable, it makes sense to model the n observed net weights as iid random variables with (unknown) marginal mean μ and standard deviation $\sigma = 1.6$ g. For large n,

$$\overline{V} = \text{the observed sample average net weight}$$

can be thought of as approximately normal with mean μ and standard deviation $\sigma/\sqrt{n} = 1.6/\sqrt{n}$ (by Proposition 3 and formulas (5.55) and (5.56)).

Now the requirement that $\overline{V}$ be within .3 g of μ can be written as

$$\mu - .3 < \overline{V} < \mu + .3$$

so the problem at hand is to choose n such that

$$P[\mu - .3 < \overline{V} < \mu + .3] = .80$$

Figure 5.42 pictures the situation. The .90 quantile of the standard normal distribution is roughly 1.28—that is, $P[-1.28 < Z < 1.28] = .8$. So evidently Figure 5.42 indicates that $\mu + .3$ should have z-value 1.28. That is, you want

$$1.28 = \frac{(\mu + .3) - \mu}{\frac{1.6}{\sqrt{n}}}$$

or

$$.3 = 1.28\frac{1.6}{\sqrt{n}}$$

So, solving for n, a sample size of $n \approx 47$ would be required to provide the kind of precision of measurement desired.

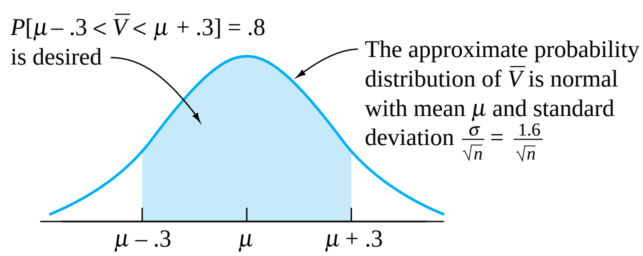


Figure 5.42 Approximate probability distribution for $\overline{V}$

Section 5 Exercises

1. A type of nominal $\frac{3}{4}$ inch plywood is made of five layers. These layers can be thought of as having thicknesses roughly describable as independent random variables with means and standard deviations as follows:

Layer	Mean (in.)	Standard Deviation (in.)
1	.094	.001
2	.156	.002
3	.234	.002
4	.172	.002
5	.094	.001

Find the mean and standard deviation of total thickness associated with the combination of these individual values.

2. The coefficient of linear expansion of brass is to be obtained as a laboratory exercise. For a brass bar that is L_1 meters long at T_1°C and L_2 meters long at T_2°C, this coefficient is

$$\alpha = \frac{L_2 - L_1}{L_1(T_2 - T_1)}$$

Suppose that the equipment to be used in the laboratory is thought to have a standard deviation for repeated length measurements of about .00005 m and a standard deviation for repeated temperature measurements of about .1°C.

(a) If using $T_1 \approx 50°$C and $T_2 \approx 100°$C, $L_1 \approx 1.00000$ m and $L_2 \approx 1.00095$ m are obtained, and it is desired to attach an approximate standard deviation to the derived value of α, find such an approximate standard deviation two different ways. First, use simulation as was done in Printout 1. Then use the propagation of error formula. How well do your two values agree?

(b) In this particular lab exercise, the precision of which measurements (the lengths or the temperatures) is the primary limiting factor in the precision of the derived coefficient of linear expansion? Explain.

(c) Within limits, the larger $T_2 - T_1$, the better the value for α. What (in qualitative terms) is the physical origin of those limits?

3. Consider again the random number generator discussed in Exercise 1 of Section 5.2. Suppose that it is used to generate 25 random numbers and that these may reasonably be thought of as independent random variables with common individual (marginal) distribution as given in Exercise 1 of Section 5.2. Let $\overline{X}$ be the sample mean of these 25 values.

(a) What are the mean and standard deviation of the random variable $\overline{X}$?

(b) What is the approximate probability distribution of $\overline{X}$?

(c) Approximate the probability that $\overline{X}$ exceeds .5.

(d) Approximate the probability that $\overline{X}$ takes a value within .02 of its mean.

(e) Redo parts (a) through (d) using a sample size of 100 instead of 25.

4. Passing a large production run of piston rings through a grinding operation produces edge widths possessing a standard deviation of .0004 in. A simple random sample of rings is to be taken and their edge widths measured, with the intention of using $\overline{X}$ as an estimate of the population mean thickness μ. Approximate the probabilities that $\overline{X}$ is within .0001 in. of μ for samples of size $n = 25$, 100, and 400.

5. A pendulum swinging through small angles approximates simple harmonic motion. The period of the pendulum, τ, is (approximately) given by

$$\tau = 2\pi\sqrt{\frac{L}{g}}$$

where L is the length of the pendulum and g is the acceleration due to gravity. This fact can be used to derive an experimental value for g. Suppose that the length L of about 5 ft can be measured with a standard deviation of about .25 in. (about .0208 foot), and the period τ of about 2.48 sec can be measured with standard deviation of about .1 sec. What is a reasonable standard deviation to attach to a value of g derived using this equipment? Is the precision of the length measurement or the precision of the period measurement the principal limitation on the precision of the derived g?

Chapter 5 Exercises

1. Suppose 90% of all students taking a beginning programming course fail to get their first program to run on first submission. Use a binomial distribution and assign probabilities to the possibilities that among a group of six such students,

(a) all fail on their first submissions

(b) at least four fail on their first submissions

(c) less than four fail on their first submissions

Continuing to use this binomial model,

(d) what is the mean number who will fail?

(e) what are the variance and standard deviation of the number who will fail?

2. Suppose that for single launches of a space shuttle, there is a constant probability of O-ring failure (say,

.15). Consider ten future launches, and let X be the number of those involving an O-ring failure. Use an appropriate probability model and evaluate all of the following:

(a) $P[X = 2]$ (b) $P[X \geq 1]$
(c) EX (d) $\text{Var}\, X$
(e) the standard deviation of X

3. An injection molding process for making auto bumpers leaves an average of 1.3 visual defects per bumper prior to painting. Let Y and Z be the numbers of visual defects on (respectively) the next two bumpers produced. Use an appropriate probability distribution and evaluate the following:

(a) $P[Y = 2]$ (b) $P[Y \geq 1]$
(c) $\sqrt{\text{Var}\, Y}$
(d) $P[Y + Z \geq 2]$ (*Hint:* What is a sensible distribution for $Y + Z$, the number of blemishes on two bumpers?)

4. Suppose that the random number generator supplied in a pocket calculator actually generates values in such a way that if X is the next value generated, X can be adequately described using a probability density of the form

$$f(x) = \begin{cases} k((x - .5)^2 + 1) & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Evaluate k and sketch a graph of $f(x)$.
(b) Evaluate $P[X \geq .5]$, $P[X > .5]$, $P[.75 > X \geq .5]$, and $P[|X - .5| \geq .2]$.
(c) Compute EX and $\text{Var}\, X$.
(d) Compute and graph $F(x)$, the cumulative probability function for X. Read from your graph the .8 quantile of the distribution of X.

5. Suppose that Z is a standard normal random variable. Evaluate the following probabilities involving Z:

(a) $P[Z \leq 1.13]$ (b) $P[Z > -.54]$
(c) $P[-1.02 < Z < .06]$ (d) $P[|Z| \leq .25]$
(e) $P[|Z| > 1.51]$ (f) $P[-3.0 < Z < 3.0]$

Find numbers # such that the following statements about Z are true:

(g) $P[|Z| < \#] = .80$ (h) $P[Z < \#] = .80$
(i) $P[|Z| > \#] = .04$

6. Suppose that X is a normal random variable with mean $\mu = 10.2$ and standard deviation $\sigma = .7$. Evaluate the following probabilities involving X:

(a) $P[X \leq 10.1]$ (b) $P[X > 10.5]$
(c) $P[9.0 < X < 10.3]$ (d) $P[|X - 10.2| \leq .25]$
(e) $P[|X - 10.2| > 1.51]$

Find numbers # such that the following statements about X are true:

(f) $P[|X - 10.2| < \#] = .80$
(g) $P[X < \#] = .80$
(h) $P[|X - 10.2| > \#] = .04$

7. In a grinding operation, there is an upper specification of 3.150 in. on a dimension of a certain part after grinding. Suppose that the standard deviation of this normally distributed dimension for parts of this type ground to any particular mean dimension μ is $\sigma = .002$ in. Suppose further that you desire to have no more than 3% of the parts fail to meet specifications. What is the maximum (minimum machining cost) μ that can be used if this 3% requirement is to be met?

8. A 10 ft cable is made of 50 strands. Suppose that, individually, 10 ft strands have breaking strengths with mean 45 lb and standard deviation 4 lb. Suppose further that the breaking strength of a cable is roughly the sum of the strengths of the strands that make it up.

(a) Find a plausible mean and standard deviation for the breaking strengths of such 10 ft cables.
(b) Evaluate the probability that a 10 ft cable of this type will support a load of 2230 lb. (*Hint:* If $\overline{X}$ is the mean breaking strength of the strands, $\sum(\text{Strengths}) \geq 2230$ is the same as $\overline{X} \geq (\frac{2230}{50})$. Now use the central limit theorem.)

9. The electrical resistivity, ρ, of a piece of wire is a property of the material involved and the temperature at which it is measured. At a given temperature, if a cylindrical piece of wire of length L and cross-sectional area A has resistance R, the material's resistivity is calculated using the formula $\rho = \frac{RA}{L}$. Thus, if a wire's cross section is assumed to be circular with diameter D, the resistivity at a given temperature is

$$\rho = \frac{R\pi D^2}{4L}$$

In a lab exercise to determine the resistivity of copper at 20°C, students measure lengths, diameters, and resistances of wire nominally 1.0 m in length (L), 2.0×10^{-3} m in diameter (D), and of resistance (R) $.54 \times 10^{-2}$ Ω. Suppose that it is sensible to describe the measurement precisions in this laboratory with the standard deviations $\sigma_L \approx 10^{-3}$ m, $\sigma_D \approx 10^{-4}$ m, and $\sigma_R \approx 5 \times 10^{-4}$ Ω.

(a) Find an approximate standard deviation that might be used to describe the expected precision for an experimentally derived value of ρ.
(b) Imprecision in which of the measurements is likely to be the largest contributor to imprecision in measured resistivity? Explain.
(c) You should expect that the value derived in (a) underpredict the kind of variation that would be observed in such laboratory exercises over a period of years. Explain why this is so.

10. Suppose that the thickness of sheets of a certain weight of book paper have mean .1 mm and a standard deviation of .003 mm. A particular textbook will be printed on 370 sheets of this paper. Find sensible values for the mean and standard deviation of the thicknesses of copies of this text (excluding, of course, the book's cover).

11. Pairs of resistors are to be connected in parallel and a difference in electrical potential applied across the resistor assembly. Ohm's law predicts that in such a situation, the current flowing in the circuit will be

$$I = V\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

where R_1 and R_2 are the two resistances and V the potential applied. Suppose that R_1 and R_2 have means $\mu_R = 10$ Ω and standard deviations $\sigma_R = .1$ Ω and that V has mean $\mu_V = 9$ volt and $\sigma_V = .2$ volt.

(a) Find an approximate mean and standard deviation for I, treating R_1, R_2, and V as independent random variables.
(b) Based on your work in (a), would you say that the variation in voltage or the combined variations in R_1 and R_2 are the biggest contributors to variation in current? Explain.

12. Students in a materials lab are required to experimentally determine the heat conductivity of aluminum.

(a) If student-derived values are normally distributed about a mean of .5 (cal/(cm)(sec)(°C)) with standard deviation of .03, evaluate the probability that an individual student will obtain a conductivity from .48 to .52.
(b) If student values have the mean and standard deviation given in (a), evaluate the probability that a class of 25 students will produce a sample mean conductivity from .48 to .52.
(c) If student values have the mean and standard deviation given in (a), evaluate the probability that at least 2 of the next 5 values produced by students will be in the range from .48 to .52.

13. Suppose that 10 ft lengths of a certain type of cable have breaking strengths with mean $\mu = 450$ lb and standard deviation $\sigma = 50$ lb.

(a) If five of these cables are used to support a single load L, suppose that the cables are loaded in such a way that support fails if any one of the cables has strength below $\frac{L}{5}$. With $L = 2{,}000$ lb, assess the probability that the support fails, if individual cable strength is normally distributed. Do this in two steps. First find the probability that a particular individual cable fails, then use that to evaluate the desired probability.
(b) Approximate the probability that the sample mean strength of 100 of these cables is below 457 lb.

14. Find EX and Var X for a continuous distribution with probability density

$$f(x) = \begin{cases} .3 & \text{if } 0 < x < 1 \\ .7 & \text{if } 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

15. Suppose that it is adequate to describe the 14-day compressive strengths of test specimens of a certain concrete mixture as normally distributed with mean $\mu = 2{,}930$ psi and standard deviation $\sigma = 20$ psi.

(a) Assess the probability that the next specimen of this type tested for compressive strength will have strength above 2,945 psi.

(b) Use your answer to part (a) and assess the probability that in the next four specimens tested, at least one has compressive strength above 2,945 psi.

(c) Assess the probability that the next 25 specimens tested have a sample mean compressive strength within 5 psi of $\mu = 2{,}930$ psi.

(d) Suppose that although the particular concrete formula under consideration in this problem is relatively strong, it is difficult to pour in large quantities without serious air pockets developing (which can have important implications for structural integrity). In fact, suppose that using standard methods of pouring, serious air pockets form at an average rate of 1 per 50 cubic yards of poured concrete. Use an appropriate probability distribution and assess the probability that two or more serious air pockets will appear in a 150 cubic yard pour to be made tomorrow.

16. For X with a continuous distribution specified by the probability density

$$f(x) = \begin{cases} .5x & \text{for } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

find $P[X < 1.0]$ and find the mean, EX.

17. The viscosity of a liquid may be measured by placing it in a cylindrical container and determining the force needed to turn a cylindrical rotor (of nearly the same diameter as the container) at a given velocity in the liquid. The relationship between the viscosity η, force F, area A of the side of the rotor in contact with the liquid, the size L of the gap between the rotor and the inside of the container, and the velocity v at which the rotor surface moves is

$$\eta = \frac{FL}{vA}$$

Suppose that students are to determine an experimental viscosity for SAE no. 10 oil as a laboratory exercise and that appropriate means and standard deviations for the measured variables F, L, v, and A in this laboratory are as follows:

$$\begin{array}{ll} \mu_F = 151 \text{ N} & \sigma_F = .05 \text{ N} \\ \mu_A = 1257 \text{ cm}^2 & \sigma_A = .2 \text{ cm}^2 \\ \mu_L = .5 \text{ cm} & \sigma_L = .05 \text{ cm} \\ \mu_v = 30 \text{ cm/sec} & \sigma_v = 1 \text{ cm/sec} \end{array}$$

(a) Use the propagation of error formulas and find an approximate standard deviation that might serve as a measure of precision for an experimentally derived value of η from this laboratory.

(b) Explain why, if experimental values of η obtained for SAE no. 10 oil in similar laboratory exercises conducted over a number of years at a number of different universities were compared, the approximate standard deviation derived in (a) would be likely to understate the variability actually observed in those values.

18. The heat conductivity, λ, of a cylindrical bar of diameter D and length L, connected between two constant temperature devices of temperatures T_1 and T_2 (respectively), that conducts Q calories in t seconds is

$$\lambda = \frac{4QL}{\pi(T_1 - T_2)tD^2}$$

In a materials laboratory exercise to determine λ for brass, the following means and standard deviations for the variables D, L, T_1, T_2, Q, and t are appropriate, as are the partial derivatives of λ with respect to the various variables (evaluated at the means of the variables):

	D	L	T_1
μ	1.6 cm	100 cm	100°C
σ	.1 cm	.1 cm	1°C
partial	−.249	.199	−.00199

	T_2	Q	t
μ	0°C	240 cal	600 sec
σ	1°C	10 cal	1 sec
partial	.00199	.000825	.000332

(The units of the partial derivatives are the units of λ(cal/(cm)(sec)(°C)) divided by the units of the variable in question.)

(a) Find an approximate standard deviation to associate with an experimentally derived value of λ.

(b) Which of the variables appears to be the biggest contributor to variation in experimentally determined values of λ? Explain your choice.

19. Suppose that 15% of all daily oxygen purities delivered by an air-products supplier are below 99.5% purity and that it is plausible to think of daily purities as independent random variables. Evaluate the probability that in the next five-day workweek, 1 or less delivered purities will fall below 99.5%.

20. Suppose that the raw daily oxygen purities delivered by an air-products supplier have a standard deviation $\sigma \approx .1$ (percent), and it is plausible to think of daily purities as independent random variables. Approximate the probability that the sample mean $\overline{X}$ of $n = 25$ delivered purities falls within .03 (percent) of the raw daily purity mean, μ.

21. Students are going to measure Young's Modulus for copper by measuring the elongation of a piece of copper wire under a tensile force. For a cylindrical wire of diameter D subjected to a tensile force F, if the initial length (length before applying the force) is L_0 and final length is L_1, Young's Modulus for the material in question is

$$Y = \frac{4FL_0}{\pi D^2(L_1 - L_0)}$$

The test and measuring equipment used in a particular lab are characterized by the standard deviations

$$\sigma_F \approx 10 \text{ lb} \qquad \sigma_D \approx .001 \text{ in.}$$
$$\sigma_{L_0} = \sigma_{L_1} = .01 \text{ in.}$$

and in the setup employed, $F \approx 300$ lb, $D \approx .050$ in., $L_0 \approx 10.00$ in., and $L_1 \approx 10.10$ in.

(a) Treating the measured force, diameter, and lengths as independent variables with the preceding means and standard deviations, find an approximate standard deviation to attach to an experimentally derived value of Y. (Partial derivatives of Y at the nominal values of F, D, L_0, and L_1 are approximately $\frac{\partial Y}{\partial F} \approx 5.09 \times 10^4$, $\frac{\partial Y}{\partial D} \approx -6.11 \times 10^8$, $\frac{\partial Y}{\partial L_0} \approx 1.54 \times 10^8$, and $\frac{\partial Y}{\partial L_1} \approx -1.53 \times 10^8$ in the appropriate units.)

(b) Uncertainty in which of the variables is the biggest contributor to uncertainty in Y?

(c) Notice that the equation for Y says that for a particular material (and thus supposedly constant Y), circular wires of constant initial lengths L_0, but of different diameters and subjected to different tensile forces, will undergo elongations $\Delta L = L_1 - L_0$ of approximately

$$\Delta L \approx \kappa \frac{F}{D^2}$$

for κ a constant depending on the material and the initial length. Suppose that you decide to measure ΔL for a factorial arrangement of levels of F and D. Does the equation predict that F and D *will* or *will not* have important interactions? Explain.

22. Exercise 6 of Chapter 3 concerns the lifetimes (in numbers of 24 mm deep holes drilled in 1045 steel before failure) of 12 D952-II (8 mm) drills.

(a) Make a normal plot of the data given in Exercise 6 of Chapter 3. In what specific way does the shape of the data distribution appear to depart from a Gaussian shape?

(b) The 12 lifetimes have mean $\bar{y} = 117.75$ and standard deviation $s \approx 51.1$. Simply using these in place of μ and σ for the underlying drill life distribution, use the normal table to find an approximate fraction of drill lives below 40 holes.

(c) Based on your answer to (a), if your answer to (b) is seriously different from the real fraction of drill lives below 40, is it most likely high or low? Explain.

23. Metal fatigue causes cracks to appear on the skin of older aircraft. Assume that it is reasonable to model the number of cracks appearing on a 1 m^2 surface of planes of a certain model and vintage as Poisson with mean $\lambda = .03$.

(a) If 1 m^2 is inspected, assess the probability that at least one crack is present on that surface.

(b) If 10 m^2 are inspected, assess the probability that at least one crack (total) is present.

(c) If ten areas, each of size 1 m^2, are inspected, assess the probability that exactly one of these has cracks.

24. If a dimension on a mechanical part is normally distributed, how small must the standard deviation be if 95% of such parts are to be within specifications of 2 cm $\pm$ .002 cm when the mean dimension is ideal ($\mu = 2$ cm)?

25. The fact that the "exact" calculation of normal probabilities requires either numerical integration or the use of tables (ultimately generated using numerical integration) has inspired many people to develop approximations to the standard normal cumulative distribution function. Several of the simpler of these approximations are discussed in the articles "A Simpler Approximation for Areas Under the Standard Normal Curve," by A. Shah (*The American Statistician*, 1985), "Pocket-Calculator Approximation for Areas under the Standard Normal Curve," by R. Norton (*The American Statistician*, 1989), and "Approximations for Hand Calculators Using Small Integer Coefficients," by S. Derenzo (*Mathematics of Computation*, 1977). For $z > 0$, consider the approximations offered in these articles:

$$\Phi(z) \approx g_{\mathrm{S}}(z) = \begin{cases} .5 + \dfrac{z(4.4 - z)}{10} & 0 \le z \le 2.2 \\ .99 & 2.2 < z < 2.6 \\ 1.00 & 2.6 \le z \end{cases}$$

$$\Phi(z) \approx g_{\mathrm{N}}(z) = 1 - \frac{1}{2}\exp\left(-\frac{z^2 + 1.2z^{.8}}{2}\right)$$

$$\Phi(z) \approx g_{\mathrm{D}}(z)$$

$$= 1 - \frac{1}{2}\exp\left(-\frac{(83z + 351)z + 562}{703/z + 165}\right)$$

Evaluate $g_{\mathrm{S}}(z)$, $g_{\mathrm{N}}(z)$, and $g_{\mathrm{D}}(z)$ for $z = .5$, 1.0, 1.5, 2.0, and 2.5. How do these values compare to the corresponding entries in Table B.3?

26. Exercise 25 concerned approximations for normal probabilities. People have also invested a fair amount of effort in finding useful formulas approximating standard normal *quantiles*. One such approximation was given in formula (3.3). A more complicated one, again taken from the article by S. Derenzo mentioned in Exercise 25, is as follows. For $p > .50$, let $y = -\ln(2(1-p))$ and

$$Q_z(p) \approx \sqrt{\frac{((4y + 100)y + 205)\,y^2}{((2y + 56)y + 192)\,y + 131}}$$

For $p < .50$, let $y = -\ln(2p)$ and

$$Q_z(p) \approx -\sqrt{\frac{((4y + 100)y + 205)\,y^2}{((2y + 56)y + 192)\,y + 131}}$$

Use these formulas to approximate $Q_z(p)$ for $p = .01, .05, .1, .3, .7, .9, .95$, and .99. How do the values you obtain compare with the corresponding entries in Table 3.10 and the results of using formula (3.3)?

27. The article "Statistical Strength Evaluation of Hot-pressed Si_3N_4" by R. Govila (*Ceramic Bulletin*, 1983) contains summary statistics from an extensive study of the flexural strengths of two high-strength hot-pressed silicon nitrides in $\frac{1}{4}$ point, 4 point bending. The values below are fracture strengths of 30 specimens of one of the materials tested at 20°C. (The units are MPa, and the data were read from a graph in the paper and may therefore individually differ by perhaps as much as 10 MPa from the actual measured values.)

514, 533, 543, 547, 584, 619, 653, 684,
689, 695, 700, 705, 709, 729, 729, 753,
763, 800, 805, 805, 814, 819, 819, 839,
839, 849, 879, 900, 919, 979

(a) The materials researcher who collected the original data believed the Weibull distribution to be an adequate model for flexural strength of this material. Make a Weibull probability plot using the method of display (5.35) of Section 5.3 and investigate this possibility. Does a Weibull model fit these data?

(b) Eye-fit a line through your plot from part (a). Use it to help you determine an appropriate shape parameter, β, and an appropriate scale parameter, α, for a Weibull distribution used to describe flexural strength of this material at 20°C. For a Weibull distribution with your fitted values of α and β, what is the median strength? What is a strength exceeded by 80% of such Si_3N_4 specimens? By 90% of such specimens? By 99% of such specimens?

(c) Make normal plots of the raw data and of the logarithms of the raw data. Comparing the three probability plots made in this exercise, is there strong reason to prefer a Weibull model, a normal model, or a lognormal model over the other two possibilities as a description of the flexural strength?

(d) Eye-fit lines to your plots from part (c). Use them to help you determine appropriate means and standard deviations for normal distributions used to describe flexural strength and the logarithm of flexural strength. Compare the .01, .10, .20, and .50 quantiles of the fitted normal and lognormal distributions for strength to the quantiles you computed in part (b).

28. The article "Using Statistical Thinking to Solve Maintenance Problems" by Brick, Michael, and Morganstein (*Quality Progress*, 1989) contains the following data on lifetimes of sinker rollers. Given are the numbers of 8-hour shifts that 17 sinker rollers (at the bottom of a galvanizing pot and used to direct steel sheet through a coating operation) lasted before failing and requiring replacement.

10, 12, 15, 17, 18, 18, 20, 20,
21, 21, 23, 25, 27, 29, 29, 30, 35

(a) The authors of the article considered a Weibull distribution to be a likely model for the lifetimes of such rollers. Make a zero-threshold Weibull probability plot for use in assessing the reasonableness of such a description of roller life.

(b) Eye-fit a line to your plot in (a) and use it to estimate parameters for a Weibull distribution for describing roller life.

(c) Use your estimated parameters from (a) and the form of the Weibull cumulative distribution function given in Section 5.2 to estimate the .10 quantile of the roller life distribution.

29. The article "Elementary Probability Plotting for Statistical Data Analysis" by J. King (*Quality Progress*, 1988) contains 24 measurements of deviations from nominal of a distance between two holes drilled in a steel plate. These are reproduced here. The units are mm.

$$-2, -2, 7, -10, 4, -3, 0, 8, -5, 5, -6, 0,$$
$$2, -2, 1, 3, 3, -4, -6, -13, -7, -2, 2, 2$$

(a) Make a dot diagram for these data and compute $\bar{x}$ and s.
(b) Make a normal plot for these data. Eye-fit a line on the plot and use it to find graphical estimates of a process mean and standard deviation for this deviation from nominal. Compare these graphical estimates with the values you calculated in (a).
(c) Engineering specifications on this deviation from nominal were ± 10 mm. Suppose that $\bar{x}$ and s from (a) are adequate approximations of the process mean and standard deviation for this variable. Use the normal distribution with those parameters and compute a fraction of deviations that fall outside specifications. Does it appear from this exercise that the drilling operation is *capable* (i.e., precise) enough to produce essentially all measured deviations in specifications, at least if properly aimed? Explain.

30. An engineer is responsible for setting up a monitoring system for a critical diameter on a turned metal part produced in his plant. Engineering specifications for the diameter are 1.180 in. $\pm$ .004 in. For ease of communication, the engineer sets up the following nomenclature for measured diameters on these parts:

Green Zone Diameters: 1.178 in. $\le$ Diameter $\le$ 1.182 in.
Red Zone Diameters: Diameter $\le$ 1.176 in. or Diameter $\ge$ 1.184 in.
Yellow Zone Diameters: any other Diameter

Suppose that in fact the diameters of parts coming off the lathe in question can be thought of as independent normal random variables with mean $\mu = 1.181$ in. and standard deviation $\sigma = .002$ in.

(a) Find the probabilities that a given diameter falls into each of the three zones.
(b) Suppose that a technician simply begins measuring diameters on consecutive parts and continues until a Red Zone measurement is found. Assess the probability that more than ten parts must be measured. Also, give the expected number of measurements that must be made.

The engineer decides to use the Green/Yellow/Red gauging system in the following way. Every hour, parts coming off the lathe will be checked. First, a single part will be measured. If it is in the Green Zone, no further action is needed that hour. If the initial part is in the Red Zone, the lathe will be stopped and a supervisor alerted. If the first part is in the Yellow Zone, a second part is measured. If this second measurement is in the Green Zone, no further action is required, but if it is in the Yellow or the Red Zone, the lathe is stopped and a supervisor alerted. It is possible to argue that under this scheme (continuing to suppose that measurements are independent normal variables with mean 1.181 in. and standard deviation .002 in.), the probability that the lathe is stopped in any given hour is .1865.

(c) Use the preceding fact and evaluate the probability that the lathe is stopped exactly twice in 8 consecutive hours. Also, what is the expected number of times the lathe will be stopped in 8 time periods?

31. A random variable X has a cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \sin(x) & \text{for } 0 < x \le \pi/2 \\ 1 & \text{for } \pi/2 < x \end{cases}$$

(a) Find $P[X \le .32]$.
(b) Give the probability density for X, $f(x)$.
(c) Evaluate EX and $\text{Var}\,X$.

32. Return to the situation of Exercise 2 of Section 5.4.

Suppose that demerits are assigned to devices of the type considered there according to the formula $D = 5X + Y$.

(a) Find the mean value of D, ED. (Use your answers to (c) and (d) Exercise 2 of Section 5.4 and formula (5.53) of Section 5.5. Formula (5.53) holds whether or not X and Y are independent.)

(b) Find the probability a device of this type scores 7 or less demerits. That is, find $P[D \le 7]$.

(c) On average, how many of these devices will have to be inspected in order to find one that scores 7 or less demerits? (Use your answer to (b).)

33. Consider jointly continuous random variables X and Y with density

$$f(x,y) = \begin{cases} x+y & \text{for } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the probability that the product of X and Y is at least $\frac{1}{4}$.

(b) Find the marginal probability density for X. (Notice that Y's is similar.) Use this to find the expected value and standard deviation of X.

(c) Are X and Y independent? Explain.

(d) Compute the mean of $X + Y$. Why can't formula (5.54) of Section 5.5 be used to find the variance of $X + Y$?

34. Return to the situation of Exercise 4 of Section 5.4.

(a) Find EX, $\text{Var}\,X$, EY, and $\text{Var}\,Y$ using the marginal densities for X and Y.

(b) Use your answer to (a) and Proposition 1 to find the mean and variance of $Y - X$.

35. Visual inspection of integrated circuit chips, even under high magnification, is often less than perfect. Suppose that an inspector has an 80% chance of detecting any given flaw. We will suppose that the inspector never "cries wolf"—that is, sees a flaw where none exists. Then consider the random variables

X = the true number of flaws on a chip

Y = the number of flaws identified by the inspector

(a) What is a sensible conditional distribution for Y given that $X = 5$? Given that $X = 5$, find the (conditional) probability that $Y = 3$.

In general, a sensible conditional probability function for Y given $X = x$ is the binomial probability function with number of trials x and success probability .8. That is, one could use

$$f_{Y|X}(y \mid x) = \begin{cases} \binom{x}{y} .8^y .2^{x-y} & \text{for } y = 0, 1, 2, \ldots, x \\ 0 & \text{otherwise} \end{cases}$$

Now suppose that X is modeled as Poisson with mean $\lambda = 3$—i.e.,

$$f_X(x) = \begin{cases} \dfrac{e^{-3}3^x}{x!} & \text{for } x = 0, 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Multiplication of the two formulas gives a joint probability function for X and Y.

(b) Find the (marginal) probability that $Y = 0$. (Note that this is obtained by summing $f(x, 0)$ over all possible values of x.)

(c) Find $f_Y(y)$ in general. What (marginal) distribution does Y have?

36. Suppose that cans to be filled with a liquid are circular cylinders. The radii of these cans have mean $\mu_r = 1.00$ in. and standard deviation $\sigma_r = .02$ in. The volumes of liquid dispensed into these cans have mean $\mu_v = 15.10$ in.^{3} and standard deviation $\sigma_v = .05$ in.^{3}.

(a) If the volumes dispensed into the cans are approximately normally distributed, about what fraction will exceed 15.07 in.^{3}?

(b) Approximate the probability that the total volume dispensed into the next 100 cans exceeds 1510.5 in.^{3} (if the total exceeds 1510.5, $\overline{X}$ exceeds 15.105).

(c) Approximate the mean μ_h and standard deviation σ_h of the heights of the liquid in the filled cans. (Recall that the volume of a circular cylinder is $v = \pi r^2 h$, where h is the height of the cylinder.)

(d) Does the variation in bottle radius or the variation in volume of liquid dispensed into the bottles have the biggest impact on the variation in liquid height? Explain.

37. Suppose that a pair of random variables have the joint probability density

$$f(x, y) = \begin{cases} \exp(x - y) & \text{if } 0 \le x \le 1 \text{ and } x \le y \\ 0 & \text{otherwise} \end{cases}$$

(a) Evaluate $P[Y \le 1.5]$.
(b) Find the marginal probability densities for X and Y.
(c) Are X and Y independent? Explain.
(d) Find the conditional probability density for Y given $X = .25$, $f_{Y|X}(y \mid .25)$. Given that $X = .25$, what is the mean of Y? (*Hint:* Use $f_{Y|X}(y \mid .25)$.)

38. (**Defects per Unit Acceptance Sampling**) Suppose that in the inspection of an incoming product, nonconformities on an inspection unit are counted. If too many are seen, the incoming lot is rejected and returned to the manufacturer. (For concreteness, you might think of blemishes on rolled paper or wire, where an inspection unit consists of a certain length of material from the roll.) Suppose further that the number of nonconformities on a piece of product of any particular size can be modeled as Poisson with an appropriate mean.

(a) Suppose that this rule is followed: "Accept the lot if on a standard size inspection unit, 1 or fewer nonconformities are seen." The *operating characteristic curve* of this acceptance sampling plan is a plot of the probability that the lot is accepted as a function of λ = the mean defects per inspection unit. (For X = the number of nonconformities seen, X has Poisson distribution with mean λ and $\text{OC}(\lambda) = P[X \le 1]$.) Make a plot of the operating characteristic curve. List values of the operating characteristic for $\lambda = .25$, .5, and 1.0.

(b) Suppose that instead of the rule in (a), this rule is followed: "Accept the lot if on 2 standard size inspection units, 2 or fewer total nonconformities are seen." Make a plot of the operating characteristic curve for this second plan and compare it with the plot from part (a). (Note that here, for X = the total number of nonconformities seen, X has a Poisson distribution with mean 2λ and $\text{OC}(\lambda) = P[X \le 2]$.) List values of the operating characteristic for $\lambda = .25$, .5, and 1.0.

39. A discrete random variable X can be described using the following probability function:

x	1	2	3	4	5
$f(x)$	.61	.24	.10	.04	.01

(a) Make a probability histogram for X. Also plot $F(x)$, the cumulative probability function for X.
(b) Find the mean and standard deviation for the random variable X.
(c) Evaluate $P[X \ge 3]$ and then find $P[X < 3]$.

40. A classical data set of Rutherford and Geiger (referred to in Example 6) suggests that for a particular experimental setup involving a small bar of polonium, the number of collisions of α particles with a small screen placed near the bar during an 8-minute period can be modeled as a Poisson variable with mean $\lambda = 3.87$. Consider an experimental setup of this type, and let X and Y be (respectively) the numbers of collisions in the next two 8-minute periods. Evaluate the following:

(a) $P[X \ge 2]$ (b) $\sqrt{\text{Var}\, X}$
(c) $P[X + Y = 6]$ (d) $P[X + Y \ge 3]$

(*Hint for parts (c) and (d):* What is a sensible probability distribution for $X + Y$, the number of collisions in a 16-minute period?)

41. Suppose that X is a continuous random variable with probability density of the form

$$f(x) = \begin{cases} k\left(x^2(1-x)\right) & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Evaluate k and sketch a graph of $f(x)$.
(b) Evaluate $P[X \leq .25]$, $P[X \leq .75]$, $P[.25 < X \leq .75]$, and $P[|X - .5| > .1]$.
(c) Compute EX and $\sqrt{\text{Var}\,X}$.
(d) Compute and graph $F(x)$, the cumulative distribution function for X. Read from your graph the .6 quantile of the distribution of X.

42. Suppose that engineering specifications on the shelf depth of a certain slug to be turned on a CNC lathe are from .0275 in. to .0278 in. and that values of this dimension produced on the lathe can be described using a normal distribution with mean μ and standard deviation σ.
(a) If $\mu = .0276$ and $\sigma = .0001$, about what fraction of shelf depths are in specifications?
(b) What machine precision (as measured by σ) would be required in order to produce about 98% of shelf depths within engineering specifications (assuming that μ is at the midpoint of the specifications)?

43. The resistance of an assembly of several resistors connected in series is the sum of the resistances of the individual resistors. Suppose that a large lot of nominal 10 Ω resistors has mean resistance $\mu = 9.91\ \Omega$ and standard deviation of resistances $\sigma = .08\ \Omega$. Suppose that 30 resistors are randomly selected from this lot and connected in series.
(a) Find a plausible mean and variance for the resistance of the assembly.
(b) Evaluate the probability that the resistance of the assembly exceeds 298.2 Ω. (*Hint:* If $\overline{X}$ is the mean resistance of the 30 resistors involved, the resistance of the assembly exceeding 298.2 Ω is the same as $\overline{X}$ exceeding 9.94 Ω. Now apply the central limit theorem.)

44. At a small metal fabrication company, steel rods of a particular type cut to length have lengths with standard deviation .005 in.
(a) If lengths are normally distributed about a mean μ (which can be changed by altering the setup of a jig) and specifications on this length are 33.69 in. $\pm$ .01 in., what appears to be the best possible fraction of the lengths in specifications? What does μ need to be in order to achieve this fraction?
(b) Suppose now that in an effort to determine the mean length produced using the current setup of the jig, a sample of rods is to be taken and their lengths measured, with the intention of using the value of $\overline{X}$ as an estimate of μ. Approximate the probabilities that $\overline{X}$ is within .0005 in. of μ for samples of size $n = 25, 100$, and 400. Do your calculations for this part of the question depend for their validity on the length distribution being normal? Explain.

45. Suppose that the measurement of the diameters of #10 machine screws produced on a particular machine yields values that are normally distributed with mean μ and standard deviation $\sigma = .03$ mm.
(a) If $\mu = 4.68$ mm, about what fraction of all measured diameters will fall in the range from 4.65 mm to 4.70 mm?
(b) Use your value from (a) and an appropriate discrete probability distribution to evaluate the probability (assuming $\mu = 4.68$) that among the next five measurements made, exactly four will fall in the range from 4.65 mm to 4.70 mm.
(c) Use your value from (a) and an appropriate discrete probability distribution to evaluate the probability (assuming that $\mu = 4.68$) that if one begins sampling and measuring these screws, the first diameter in the range from 4.65 mm to 4.70 mm will be found on the second, third, or fourth screw measured.
(d) Now suppose that μ is unknown but is to be estimated by $\overline{X}$ obtained from measuring a sample of $n = 25$ screws. Evaluate the probability that the sample mean, $\overline{X}$, takes a value within .01 mm of the long-run (population) mean μ.

(e) What sample size, n, would be required in order to a priori be 90% sure that $\overline{X}$ from n measurements will fall within .005 mm of μ?

46. The random variable $X =$ the number of hours till failure of a disk drive is described using an exponential distribution with mean 15,000 hours.

(a) Evaluate the probability that a given drive lasts at least 20,000 hours.

(b) A new computer network has ten of these drives installed on computers in the network. Use your answer to (a) and an assumption of independence of the ten drive lifetimes and evaluate the probability that at least nine of these drives are failure-free through 20,000 hours.

47. Miles, Baumhover, and Miller worked with a company on a packaging problem. Cardboard boxes, nominally 9.5 in. in length were supposed to hold four units of product stacked side by side. They did some measuring and found that in fact the individual product units had widths with mean approximately 2.577 in. and standard deviation approximately .061 in. Further, the boxes had (inside) lengths with mean approximately 9.566 in. and standard deviation approximately .053 in.

(a) If X_1, X_2, X_3, and X_4 are the actual widths of four of the product units and Y is the actual inside length of a box into which they are to be packed, then the "head space" in the box is $U = Y - (X_1 + X_2 + X_3 + X_4)$. What are a sensible mean and standard deviation for U?

(b) If X_1, X_2, X_3, X_4, and Y are normally distributed and independent, it turns out that U is also normal. Suppose this is the case. About what fraction of the time should the company expect to experience difficulty packing a box? (What is the probability that the head space as calculated in (a) is negative?)

(c) If it is your job to recommend a new mean inside length of the boxes and the company wishes to have packing problems in only .5% of the attempts to load four units of product into a box, what is the minimum mean inside length you would recommend? (Assume that standard deviations will remain unchanged.)