APPENDIX

A

More on Probability and Model Fitting

The introduction to probability theory in Chapter 5 was relatively brief. There are, of course, important engineering applications of probability that require more background in the subject. So this appendix gives a few more details and discusses some additional uses of the theory that are reasonably elementary, particularly in the contexts of reliability analysis and life data analysis.

The appendix begins by discussing the formal/axiomatic basis for the mathematics of probability and several of the most useful simple consequences of the basic axioms. It then applies those simple theorems of probability to the prediction of reliability for series, parallel, and combination series-parallel systems. A brief section treats principles of counting (permutations and combinations) that are sometimes useful in engineering applications of probability. There follows a section on special probability concepts used with life-length (or time-to-failure) variables. The appendix concludes with a discussion of maximum likelihood methods for model fitting and drawing inferences.

A.1 More Elementary Probability

Like any other mathematical theory or system, probability theory is built on a few basic definitions and some "rules of the game" called *axioms*. Logic is applied to determine what consequences (or theorems) follow from the definitions and axioms. These, in turn, can be helpful guides as an engineer seeks to understand and predict the behavior of physical systems that involve chance.

For the sake of logical completeness, this section gives the formal axiomatic basis for probability theory and states and then illustrates the use of some simple theorems that follow from this base. Conditional probability and the independence of events are then defined, and a simple theorem related to these concepts is stated and its use illustrated.

A.1.1 Basic Definitions and Axioms

As was illustrated informally in Chapter 5, the practical usefulness of probability theory is in assigning sensible likelihoods of occurrence to possible happenings in chance situations. The basic, irreducible, potential results in such a chance situation are called **outcomes** belonging to a **sample space**.

Definition 1

A single potential result of a chance situation is called an **outcome**. All outcomes of a chance situation taken together make up a **sample space** for the situation. A script capital $\mathcal{S}$ is often used to stand for a sample space.

Mathematically, outcomes are points in a universal set that is the sample space. And notions of simple set theory become relevant. For one thing, subsets of $\mathcal{S}$ containing more than one outcome can be of interest.

Definition 2

A collection of outcomes (a subset of $\mathcal{S}$) is called an **event**. Capital letters near the beginning of the alphabet are sometimes used as symbols for events, as are English phrases describing the events.

Once one has defined events, the standard set-theoretic operations of complementation, union, and intersection can be applied to them. However, rather than using the typical "^{c}," "$\cup$," and "$\cap$" mathematical notation for these operations, it is common in probability theory to substitute the use of the words *not*, *or*, and *and*, respectively.

Definition 3

For event A and event B, subsets of some sample space $\mathcal{S}$,

1. ***notA*** is an event consisting of all outcomes not belonging to A;
2. ***AorB*** is an event consisting of all outcomes belonging to one, the other, or both of the two events; and
3. ***AandB*** is an event consisting of all outcomes belonging simultaneously to the two events.

Example 1

A Redundant Inspection System for Detecting Metal Fatigue Cracks

Consider a redundant inspection system for the detection of fatigue cracks in metal specimens. Suppose the system involves the making of a fluorescent penetrant inspection (FPI) and also a (magnetic) eddy current inspection (ECI). When a

Example 1 (*continued*)

metal specimen is to be tested using this two-detector system, a potential sample space consists of four outcomes corresponding to the possible combinations of what can happen at each detector. That is, a possible sample space is specified in a kind of set notation as

$$\mathcal{S} = \{(\text{FPI signal and ECI signal}), (\text{no FPI signal and ECI signal}), (\text{FPI signal and no ECI signal}), (\text{no FPI signal and no ECI signal})\} \quad \textbf{(A.1)}$$

and in tabular and pictorial forms as in Table A.1 and Figure A.1. Notice that Figure A.1 can be treated as a kind of Venn diagram—the big square standing for $\mathcal{S}$ and the four smaller squares making up $\mathcal{S}$ standing for events that each consist of one of the four different possible outcomes.

Using this four-outcome sample space to describe experience with a metal specimen, one can define several events of potential interest and illustrate the use of the notation described in Definition 3. That is, let

$$A = \{(\text{FPI signal and ECI signal}), (\text{FPI signal and no ECI signal})\} \quad \textbf{(A.2)}$$

$$B = \{(\text{FPI signal and ECI signal}), (\text{no FPI signal and ECI signal})\} \quad \textbf{(A.3)}$$

Table A.1
A List of the Possible Outcomes for Two Inspections

Possible Outcome	FPI Detection Signal?	ECI Detection Signal?
1	yes	yes
2	no	yes
3	yes	no
4	no	no

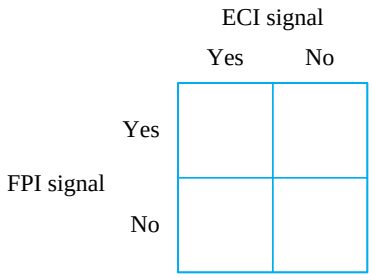

Figure A.1 Graphical representation of four outcomes of two inspections

Then in words,

$$A = \text{the FPI detector signals}$$
$$B = \text{the ECI detector signals}$$

Part 1 of Definition 3 means, for example, that using notations () and (A.2),

$$\begin{aligned} notA &= \{(\text{no FPI signal and ECI signal}), (\text{no FPI signal and no ECI signal})\} \\ &= \text{the FPI detector doesn't signal} \end{aligned}$$

Part 2 of Definition 3 means, for example, that using notations (A.2) and (A.3),

$$\begin{aligned} AorB &= \{(\text{FPI signal and ECI signal}), (\text{FPI signal and no ECI signal}), \\ &\qquad (\text{no FPI signal and ECI signal})\} \\ &= \text{at least one of the two detectors' signals} \end{aligned}$$

And Part 3 of Definition 3 means that again using (A.2) and (A.3), one has

$$\begin{aligned} AandB &= \{(\text{FPI signal and ECI signal})\} \\ &= \text{both of the two detectors' signals} \end{aligned}$$

$notA$, $AorB$, and $AandB$ are shown in Venn diagram fashion in Figure A.2.

Elementary set theory allows the possibility that a set can be empty—that is, have no elements. Such a concept is also needed in probability theory.

Definition 4

The **empty event** is an event containing no outcomes. The symbol $\emptyset$ is typically used to stand for the empty event.

$\emptyset$ has the interpretation that none of the possible outcomes of a chance situation occur. The way in which $\emptyset$ is most useful in probability is in describing the relationship between two events that have no outcomes in common, and thus cannot both occur. There is special terminology for this eventuality (that $AandB = \emptyset$).

Definition 5

If event A and event B have no outcomes in common (i.e., $AandB = \emptyset$), then the two events are called **disjoint** or **mutually exclusive**.

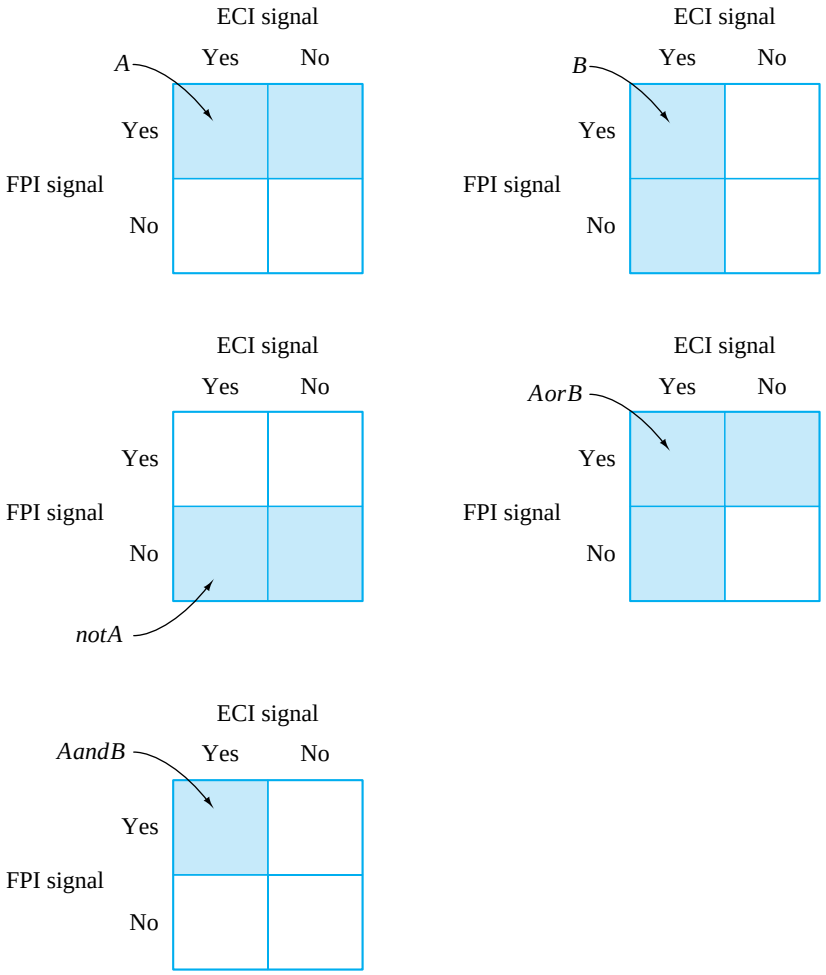

Figure A.2 Graphical representations of *A*, *B*, *notA*, *AorB*, and *AandB*

Example 1 (*continued*)

From Figure A.2 it is quite clear that, for example, the event A and the event $notA$ are disjoint. And the event $AandB$ and the event $not(AorB)$, for example, are also mutually exclusive events.

Manipulation of events using complementation, union, intersection, etc. is necessary background, but it is hardly the ultimate goal of probability theory. The goal is assignment of likelihoods to events. In order to guarantee that such assignments are internally coherent, probabilists have devised what seem to be intuitively sensible axioms (or rules of operation) for probability models. Assignment of likelihoods in conformance to those rules guarantees that (at a minimum) the assignment is

logically consistent. (Whether it is realistic or useful is a separate question.) The axioms of probability are laid out next.

Definition 6

A **system of probabilities** is an assignment of numbers (probabilities) $P[A]$ to events A in such a way that

1. for each event A, $0 \leq P[A] \leq 1$,
2. $P[\mathcal{S}] = 1$ and $P[\emptyset] = 0$, and
3. for mutually exclusive events $A_1, A_2, A_3, \ldots,$

$$P[A_1 or A_2 or A_3 or \ldots] = P[A_1] + P[A_2] + P[A_3] + \cdots$$

The relationships (1), (2), and (3) are the **axioms of probability theory**.

Definition 6 is meant to be in agreement with the ways that empirical relative frequencies behave. Axiom (1) says that, as in the case of relative frequencies, only probabilities between 0 and 1 make sense. Axiom (2) says that if one interprets a probability of 1 as sure occurrence and a probability of 0 as no chance of occurrence, it is certain that one of the outcomes in $\mathcal{S}$ will occur. Axiom (3) says that if an event can be made up of smaller nonoverlapping pieces, the probability assigned to that event must be equal to the sum of the probabilities assigned to the pieces.

Although it was not introduced in any formal way, the third axiom of probability was put to good use in Chapter 5. For example, when concluding that for a Poisson random variable X

$$\begin{aligned} P[2 \leq X \leq 5] &= P[X = 2] + P[X = 3] + P[X = 4] + P[X = 5] \\ &= f(2) + f(3) + f(4) + f(5) \end{aligned}$$

one is really using the third axiom with

$$\begin{aligned} A_1 &= \{X = 2\} \\ A_2 &= \{X = 3\} \\ A_3 &= \{X = 4\} \\ A_4 &= \{X = 5\} \end{aligned}$$

It is only in very simple situations that one would ever try to make use of Definition 6 by checking that an entire candidate set of probabilities satisfies the axioms of probability. It is more common to assign probabilities (totaling to 1) to individual outcomes and then simply declare that the third axiom of Definition 6

will be followed in making up any other probabilities. (This strategy guarantees that subsequent probability assignments will be logically consistent.)

Example 2

A System of Probabilities for Describing a Single Inspection of a Metal Part

As an extremely simple illustration, consider the result of a single inspection of a metal part for fatigue cracks using fluoride penetrant technology. With a sample space

$$\mathcal{S} = \{\text{crack signaled, crack not signaled}\}$$

there are only four events:

$\mathcal{S}$

{crack signaled}

{no crack signaled}

$\emptyset$

An assignment of probabilities that can be seen to conform to Definition 6 is

$$P[\mathcal{S}] = 1$$
$$P[\text{crack signaled}] = .3$$
$$P[\text{no crack signaled}] = .7$$
$$P[\emptyset] = 0$$

Since they conform to Definition 6, these values make up a mathematically valid system of probabilities. Whether or not they constitute a *realistic* or *useful* model is a separate question that can really be answered only on the basis of empirical evidence.

Example 1 (*continued*)

Returning to the situation of redundant inspection of metal parts using both fluoride penetrant and eddy current technologies, suppose that via extensive testing it is possible to verify that for cracks of depth .005 in., the following four values are sensible:

$$P[\text{FPI signal and ECI signal}] = .48 \quad \textbf{(A.4)}$$
$$P[\text{FPI signal and no ECI signal}] = .02 \quad \textbf{(A.5)}$$

$$P[\text{no FPI signal and ECI signal}] = .32 \quad \textbf{(A.6)}$$

$$P[\text{no FPI signal and no ECI signal}] = .18 \quad \textbf{(A.7)}$$

This assignment of probabilities to the basic outcomes in $\mathcal{S}$ is illustrated in Figure A.3. Since these four potential probabilities do total to 1, one can adopt them together with provision (3) of Definition 6 and have a mathematically consistent assignment. Then simple addition gives appropriate probabilities for all other events. For example, with event A and event B as defined earlier ($A =$ the FPI detector signals and $B =$ the ECI detector signals),

$$\begin{aligned} P[A] &= P[\text{the FPI detector signals}] \\ &= P[\text{FPI signal and ECI signal}] + P[\text{FPI signal and no ECI signal}] \\ &= .48 + .02 \\ &= .50 \end{aligned}$$

And further,

$$\begin{aligned} P[A \text{or} B] &= P[\text{at least one of the two detectors signals}] \\ &= P[\text{FPI signal and ECI signal}] + P[\text{FPI signal and no ECI signal}] \\ &\quad + P[\text{no FPI signal and ECI signal}] \\ &= .48 + .02 + .32 \\ &= .82 \end{aligned}$$

It is clear that to find the two values, one simply adds the numbers that appear in Figure A.3 in the regions that are shaded in Figure A.2 delimiting the events in question.

FPI signal \ ECI signal	Yes	No
Yes	.48	.02
No	.32	.18

Figure A.3 An assignment of probabilities to four possible outcomes of two inspections

A.1.2 Simple Theorems of Probability Theory

The preceding discussion is typical of probability analyses, in that the probabilities for all possible events are not explicitly written down. Rather, probabilities for *some* events, together with logic and the basic rules of the game (the probability axioms), are used to deduce appropriate values for probabilities of *other* events that are of particular interest. This enterprise is often facilitated by the existence of a number of simple theorems. These are general statements that are logical consequences of the axioms in Definition 6 and thus govern the assigning of probabilities for all probability models.

One such simple theorem concerns the relationship between $P[A]$ and $P[notA]$.

Proposition 1

For any event A,

$$P[notA] = 1 - P[A]$$

This fact is again one that was used freely in Chapter 5 without explicit reference. For example, in the context of independent, identical success-failure trials, the fact that the probability of at least one success (i.e., $P[X \geq 1]$ for a binomial random variable X) is 1 minus the probability of 0 successes (i.e., $1 - P[X = 0] = 1 - f(0)$) is really a consequence of Proposition 1.

Example 1 (*continued*)

Upon learning, via the addition of probabilities for individual outcomes given in displays (A.4) through (A.7), that the assignment

$$\begin{aligned} P[A] &= P[\text{the FPI detector signals}] \\ &= .50 \end{aligned}$$

is appropriate, Proposition 1 immediately implies that

$$\begin{aligned} P[notA] &= P[\text{the FPI detector doesn't signal}] \\ &= 1 - P[A] \\ &= 1 - .50 \\ &= .50 \end{aligned}$$

is also appropriate. (Of course, if the point here weren't to illustrate the use of Proposition 1, this value could just as well have been gotten by adding .32 and .18.)

A second simple theorem of probability theory is a variation on axiom (3) of Definition 6 for two events that are not necessarily disjoint. It is sometimes called the **addition rule of probability**.

Proposition 2 (*The Addition Rule of Probability*)

For any two events, event A and event B,

$$P[A \text{or} B] = P[A] + P[B] - P[A \text{and} B] \qquad \textbf{(A.8)}$$

Note that when dealing with mutually exclusive events, the last term in equation (A.8) is $P[\emptyset] = 0$. Therefore, equation (A.8) simplifies to a two-event version of part (3) of Definition 6. When the event A and the event B are not mutually exclusive, the simple addition $P[A] + P[B]$ (so to speak) counts $P[A \text{and} B]$ twice, and the subtraction in equation (A.8) corrects for this in the computing of $P[A \text{or} B]$.

The practical usefulness of an equation like (A.8) is that when furnished with any three of the four terms appearing in it, the fourth can be gotten by using simple arithmetic.

Example 3

Describing the Dual Inspection of a Single Cracked Part

Suppose that two different inspectors, both using a fluoride penetrant inspection technique, are to inspect a metal part actually possessing a crack .007 in. deep. Suppose further that some relevant probabilities in this context are

$$P[\text{inspector 1 detects the crack}] = .50$$

$$P[\text{inspector 2 detects the crack}] = .45$$

$$P[\text{at least one inspector detects the crack}] = .55$$

Then using equation (A.8),

$$P[\text{at least one inspector detects the crack}] = P[\text{inspector 1 detects the crack}] + P[\text{inspector 2 detects the crack}] - P[\text{both inspectors detect the crack}]$$

Thus,

$$.55 = .50 + .45 - P[\text{both inspectors detect the crack}]$$

so

$$P[\text{both inspectors detect the crack}] = .40$$

Example 3 (*continued*)

Of course, the .40 value is only as good as the three others used to produce it. But it is at least logically consistent with the given probabilities, and if they have practical relevance, so does the .40 value.

A third simple theorem of probability concerns cases where the basic outcomes in a sample space are judged to be equally likely.

Proposition 3

If the outcomes in a finite sample space $\mathcal{S}$ all have the same probability, then for any event A,

$$P[A] = \frac{\text{the number of outcomes in } A}{\text{the number of outcomes in } \mathcal{S}}$$

Proposition 3 shows that if one is clever or fortunate enough to be able to conceive of a sample space where an **equally likely outcomes** assignment of probabilities is sensible, the assessment of probabilities can be reduced to a simple counting problem. This fact is particularly useful in enumerative contexts (see again Definition 4 in Chapter 1 for this terminology) where one is drawing random samples from a finite population.

Example 4

Equally Likely Outcomes in a Random Sampling Scenario

Suppose that a storeroom holds, among other things, four integrated circuit chips of a particular type and that two of these are needed in the fabrication of a prototype of an advanced electronic instrument. Suppose further that one of these chips is defective. Consider assigning a probability that both of two chips selected on the first trip to the storeroom are good chips. One way to find such a value (there are others) is to use Proposition 3. Naming the three good chips G1, G2, and G3 and the single defective chip D, one can invent a sample space made up of ordered pairs, the first entry naming the first chip selected and the second entry naming the second chip selected. This is given in set notation as follows:

$$\mathcal{S} = \{(\text{G1, G2}), (\text{G1, G3}), (\text{G1, D}), (\text{G2, G1}), (\text{G2, G3}), (\text{G2, D}), (\text{G3, G1}), (\text{G3, G2}), (\text{G3, D}), (\text{D, G1}), (\text{D, G2}), (\text{D, G3})\}$$

A pictorial representation of $\mathcal{S}$ is given in Figure A.4.

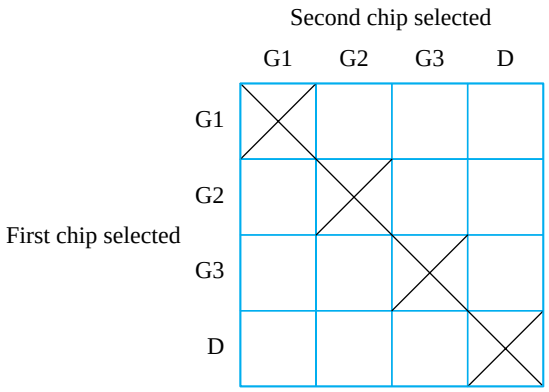

Figure A.4 Graphical representation of 12 possible outcomes when selecting two of four IC chips

Then, noting that the 12 outcomes in this sample space are reasonably thought of as equally likely and that 6 of them do not have D listed either first or second, Proposition 3 suggests the assessment

$$P[\text{two good chips}] = \frac{6}{12} = .50$$

A.1.3 Conditional Probability and the Independence of Events

Chapter 5 discussed the notion of independence for random variables. In that discussion, the idea of assigning probabilities for one variable conditional on the value of another was essential. The concept of conditional assignment of probabilities of *events* is spelled out next.

Definition 7

For event A and event B, provided event B has nonzero probability, the **conditional probability of A given B** is

$$P[A \mid B] = \frac{P[A \text{and} B]}{P[B]} \qquad \textbf{(A.9)}$$

The ratio (A.9) ought to make reasonable intuitive sense. If, for example, $P[A \text{and} B] = .3$ and $P[B] = .5$, one might reason that "B occurs only 50% of the time, but of those times B occurs, A also occurs $\frac{.3}{.5} = 60\%$ of the time. So .6 is a sensible assessment of the likelihood of A knowing that indeed B occurs."

Example 4 (*continued*)

Return to the situation of selecting two integrated circuit chips at random from four residing in a storeroom, one of which is defective. Consider using expression (A.9) and evaluating

$$P[\text{the second chip selected is defective} \mid \text{the first chip selected is good}]$$

Simple counting in the 12-outcome sample space leads to the assignments

$$P[\text{the first chip selected is good}] = \frac{9}{12} = .75$$

$$P[\text{first chip selected is good and second is defective}] = \frac{3}{12} = .25$$

So using Definition 7,

$$P[\text{the second chip selected is defective} \mid \text{the first selected is good}] = \frac{\frac{3}{12}}{\frac{9}{12}} = \frac{1}{3}$$

Of the 9 equally likely outcomes in $\mathcal{S}$ for which the first chip selected is good, there are 3 for which the second chip selected is defective. If one thinks of the 9 outcomes for which the first chip selected is good as a kind of reduced sample space (brought about by the partial restriction that the first chip selected is good), then the $\frac{3}{9}$ figure above is a perfectly plausible value for the likelihood that the second chip is defective.

There are sometimes circumstances that make it obvious how a conditional probability ought to be assigned. For example, in the context of Example 4, one might argue that it is obvious that

$$P[\text{the second chip selected is defective} \mid \text{the first selected is good}] = \frac{1}{3}$$

because if the first is good, when the second is to be selected, the storeroom will contain three chips, one of which is defective.

When one does have a natural value for $P[A \mid B]$, the relationship between this and the probabilities $P[A \text{ and } B]$ and $P[B]$ can sometimes be exploited to evaluate one or the other of them. This notion is important enough that the relationship (A.9) is often rewritten by multiplying both sides by the quantity $P[B]$ and calling the result the **multiplication rule of probability**.

Proposition 4 (*The Multiplication Rule of Probability*)

Provided $P[B] > 0$, so that $P[A \mid B]$ is defined,

$$P[A \text{and} B] = P[A \mid B] \cdot P[B] \qquad \textbf{(A.10)}$$

Example 5

The Multiplication Rule of Probability and a Probabilistic Risk Assessment

A probabilistic risk assessment of the solid rocket motor field joints used in space shuttles prior to the *Challenger* disaster was made in "Risk Analysis of the Space Shuttle: Pre-*Challenger* Prediction of Failure" (*Journal of the American Statistical Association*, 1989) by Dalal, Fowlkes, and Hoadley. They estimated that for each field joint (at 31° and 200 psi),

$$P[\text{primary O-ring erosion}] = .95$$
$$P[\text{primary O-ring blow-by} \mid \text{primary O-ring erosion}] = .29$$

Combining these two values according to rule (A.10), one then sees that the authors' assessment of the failure probability for each primary O-ring was

$$P[\text{primary O-ring erosion and blow-by}] = (.29)(.95) = .28$$

Typically, the numerical values of $P[A \mid B]$ and $P[A]$ are different. The difference can be thought of as reflecting the change in one's assessed likelihood of occurrence of A brought about by knowing that B's occurrence is certain. In cases where there is no difference, the terminology of **independence** is used.

Definition 8

If event A and event B are such that

$$P[A \mid B] = P[A]$$

they are said to be **independent**. Otherwise, they are called **dependent**.

Example 1 (*continued*)

Consider again the example of redundant fatigue crack inspection with probabilities given in Figure A.3. Since

$$P[\text{ECI signal}] = .80$$
$$P[\text{ECI signal} \mid \text{FPI signal}] = \frac{.48}{.50} = .96$$

the events {ECI signal} and {FPI signal} are dependent events.

Example 1 (*continued*)

FPI signal \ ECI signal	Yes	No
Yes	.4	.1
No	.4	.1

Figure A.5 A second assignment of probabilities to four possible outcomes of two inspections

Of course, different probabilities assigned to individual outcomes in this example can lead to the conclusion that the two events are independent. For example, the probabilities in Figure A.5 give

$$P[\text{ECI signal}] = .4 + .4 = .8$$

$$P[\text{ECI signal} \mid \text{FPI signal}] = \frac{.4}{.4 + .1} = .8$$

so with these probabilities, the two events would be independent.

The multiplication rule when A and B are independent

Independence is the mathematical formalization of the qualitative notion of *unrelatedness*. One way in which it is used in engineering applications is in conjunction with the multiplication rule. If one has values for $P[A]$ and $P[B]$ and judges the event A and the event B to be unrelated, then independence allows one to replace $P[A \mid B]$ with $P[A]$ in formula (A.10) and evaluate $P[A \text{and} B]$ as $P[A] \cdot P[B]$. (This fact was behind the scenes in Section 5.1 when sequences of independent identical success-failure trials and the binomial and geometric distributions were discussed.)

Example 5 (*continued*)

In their probabilistic risk assessment of the pre-*Challenger* space shuttle solid rocket motor field joints, Dalal, Fowlkes, and Hoadley arrived at the figure

$$P[\text{failure}] = .023$$

for a single field joint in a shuttle launch at 31°F. A shuttle's two solid rocket motors have a total of six such field joints, and it is perhaps plausible to think of their failures as independent events.

If a model of independence *is* adopted, it is possible to calculate as follows:

$$\begin{aligned} P[\text{ joint 1 and joint 2 are both effective}] &= P[\text{ joint 1 is effective}] \times \\ &\quad P[\text{ joint 2 is effective}] \\ &= (1 - .023)(1 - .023) \\ &= .95 \end{aligned}$$

And in fact, considering all six joints,

$$\begin{aligned} P[\text{at least one joint fails}] &= 1 - P[\text{all 6 joints are effective}] \\ &= 1 - (1 - .023)^6 \\ &= .13 \end{aligned}$$

Section 1 Exercises

1. Return to the situation of Chapter Exercise 30 of Chapter 5, where measured diameters of a turned metal part were coded as Green, Yellow, or Red, depending upon how close they were to a mid-specification. Suppose that the probabilities that a given diameter falls into the various zones are .6247 for the Green Zone, .3023 for the Yellow Zone, and .0730 for the Red Zone. Suppose further (as in the problem in Chapter 5) that the lathe turning the parts is checked once per hour according to the following rules: One diameter is measured, and if it is in the Green Zone, no further action is needed that hour. If it is in the Red Zone, the process is immediately stopped. If it is in the Yellow Zone, a second diameter is measured. If the second diameter is in the Green Zone, no further action is necessary, but if it is not, the process is stopped immediately. Suppose further that the lathe is physically stable, so that it makes sense to think of successive color codes as independent.

(a) Show that the probability that the process is stopped in a given hour is .1865.

(b) Given that the process is stopped, what is the conditional probability that the first diameter was in the Yellow Zone?

2. A bin of nuts is mixed, containing 30% $\frac{1}{2}$ in. nuts and 70% $\frac{9}{16}$ in. nuts. A bin of bolts has 40% $\frac{1}{2}$ in. bolts and 60% $\frac{9}{16}$ in. bolts. Suppose that one bolt and one nut are selected (independently and at random) from the two bins.

(a) What is the probability that the nut and bolt match?

(b) What is the conditional probability that the nut is a $\frac{9}{16}$ in. nut, given that the nut and bolt match?

3. A physics student is presented with six unmarked specimens of radioactive material. She knows that two are of substance A and four are of substance B. Further, she knows that when tested with a Geiger counter, substance A will produce an average of three counts per second, while substance B will produce an average of four counts per second. (Use Poisson models for the counts per time period.)

(a) Suppose the student selects a sample at random and makes a one-second check of radioactivity. If one count is observed, how should the student assess the (conditional) probability that the specimen is of substance A?

(b) Suppose the student selects a sample at random and makes a ten-second check of radioactivity.

If ten counts are observed, how should the student assess the (conditional) probability that the specimen is of substance A?

(c) Are your answers to (a) and (b) the same? How should this be understood?

4. At final inspection of certain integrated circuit chips, 20% of the chips are in fact defective. An automatic testing device does the final inspection. Its characteristics are such that 95% of good chips test as good. Also, 10% of the defective chips test as good.

(a) What is the probability that the next chip is good and tests as good?

(b) What is the probability that the next chip tests as good?

(c) What is the (conditional) probability that the next chip that tests as good is in fact good?

5. In the process of producing piston rings, the rings are subjected to a first grind. Those rings whose thicknesses remain above an upper specification are reground. The history of the grinding process has been that on the first grind,

60% of the rings meet specifications (and are done processing)

25% of the rings are above the upper specification (and are reground)

15% of the rings are below the lower specification (and are scrapped)

The history has been that after the second grind,

80% of the reground rings meet specifications

20% of the reground rings are below the lower specification

A ring enters the grinding process today.

(a) Evaluate P[the ring is ground only once].

(b) Evaluate P[the ring meets specifications].

(c) Evaluate P[the ring is ground only once | the ring meets specifications].

(d) Are the events {the ring is ground only once} and {the ring meets specifications} independent events? Explain.

(e) Describe any two mutually exclusive events in this situation.

6. A lot of machine parts is checked piece by piece for Brinell hardness and diameter, with the resulting counts as shown in the accompanying table. A single part is selected at random from this lot.

(a) What is the probability that it is more than 1.005 in. in diameter?

(b) What is the probability that it is more than 1.005 in. in diameter and has Brinell hardness of more than 210?

		Diameter		
		$<$ 1.000 in.	1.000 to 1.005 in.	$>$ 1.005 in.
Brinell Hardness	$<$ 190	154	98	48
	190–210	94	307	99
	$>$ 210	33	72	95

(c) What is the probability that it is more than 1.005 in. in diameter or has Brinell hardness of more than 210?

(d) What is the conditional probability that it has a diameter over 1.005 in., given that its Brinell hardness is over 210?

(e) Are the events {Brinell hardness over 210} and {diameter over 1.005 in.} independent? Explain.

(f) Name any two mutually exclusive events in this situation.

7. Widgets produced in a factory can be classified as defective, marginal, or good. At present, a machine is producing about 5% defective, 15% marginal, and 80% good widgets. An engineer plans the following method of checking on the machine's adjustment: Two widgets will be sampled initially, and if either is defective, the machine will be immediately adjusted. If both are good, testing will cease without adjustment. If neither of these first two possibilities occurs, an additional three widgets will be sampled. If all three of these are good, or two are good and one is marginal, testing will

cease without machine adjustment. Otherwise, the machine will be adjusted.

(a) Evaluate P[only two widgets are sampled and no adjustment is made].
(b) Evaluate P[only two widgets are sampled].
(c) Evaluate P[no adjustment is made].
(d) Evaluate P[no adjustment is made | only two widgets are sampled].
(e) Are the events {only two widgets are sampled} and {no adjustment is made} independent events? Explain.
(f) Describe any two mutually exclusive events in this situation.

8. Glass vials of a certain type are conforming, blemished (but usable), or defective. Two large lots of these vials have the following compositions.

Lot 1: 70% conforming, 20% blemished, and 10% defective

Lot 2: 80% conforming, 10% blemished, and 10% defective

Lot 1 is three times the size of Lot 2 and these two lots have been mixed in a storeroom. Suppose that a vial from the storeroom is selected at random to use in a chemical analysis.

(a) What is the probability that the vial is from Lot 1 and not defective?
(b) What is the probability that the vial is blemished?
(c) What is the conditional probability that the vial is from Lot 1 given that it is blemished?

9. A digital communications system transmits information encoded as strings of 0's and 1's. As a means of reducing transmission errors, each digit in a message string is repeated twice. Hence the message string {0 1 1 0} would (ideally) be transmitted as {00 11 11 00} and if digits received in a given pair don't match, one can be sure that the pair has been corrupted in transmission. When each individual digit in a "doubled string" like {00 11 11 00} is transmitted, there is a probability p of transmission error. Further, whether or not a particular digit is correctly transferred is independent of whether any other one is correctly transferred.

Suppose first that the single pair {00} is transmitted.

(a) Find the probability that the pair is correctly received.
(b) Find the probability that what is received has obviously been corrupted.
(c) Find the conditional probability that the pair is correctly received given that it is not obviously corrupted.

Suppose now that the "doubled string" {00 00 11 11} is transmitted and that the string received is not obviously corrupted.

(d) What is then a reasonable assignment of the "chance" that the correct message string (namely {0 0 1 1}) is received? (*Hint:* Use your answer to part c).)

10. Figure A.6 is a Venn diagram with some probabilities of events marked on it. In addition to the values marked on the diagram, it is the case that, $P[B] = .4$ and $P[C \mid A] = .8$.

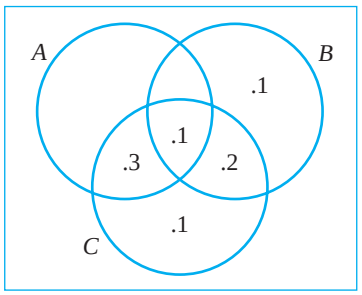

Figure A.6 Figure for Exercise 10

(a) Finish filling in the probabilities on the diagram. That is, evaluate the three probabilities $P[A \text{ and } B \text{ and not } C]$, $P[A \text{ and not } B \text{ and not } C]$ and $P[\text{not}(A \text{ or } B \text{ or } C)] = P[\text{not } A \text{ and not } B \text{ and not } C]$.
(b) Use the probabilities on the diagram (and your answers to (a)) and evaluate $P[A \text{ and } B]$.
(c) Use the probabilities on the diagram and evaluate $P[B \mid C]$.
(d) Based on the information provided here, are the events B, C independent events? Explain.

A.2 Applications of Simple Probability to System Reliability Prediction

Sometimes engineering systems are made up of identifiable components or subsystems that operate reasonably autonomously and for which fairly accurate reliability information is available. In such cases, it is sometimes of interest to try to predict overall system reliability from the available component reliabilities. This section considers how the simple probability material from Section A.1 can be used to help do this for series, parallel, and combination series-parallel systems.

A.2.1 Series Systems

Definition 9

A system consisting of components $C_1, C_2, C_3, \ldots, C_k$ is called a **series system** if its proper functioning requires the functioning of all k components.

Figure A.7 is a representation of a **series system** made up of $k = 3$ components. The interpretation to be attached to a diagram like Figure A.7 is that the system will function provided there is a path from point 1 to point 2 that crosses no failed component. (It is tempting, but *not* a good idea, to interpret a system diagram as a flow diagram or like an electrical circuit schematic. The flow diagram interpretation is often inappropriate because there need be no sequencing, time progression, communication, or other such relationship between components in a real series system. The circuit schematic notion often fails to be relevant, and even when it might seem to be, the independence assumptions typically used in arriving at a system reliability figure are of questionable practical appropriateness for electrical circuits.)

If it is sensible to model the functioning of the individual system components as *independent*, then the overall system reliability is easily deduced from component reliabilities via simple multiplication. For example, for a two-component series system,

$$\begin{aligned} P[\text{the system functions}] &= P[C_1 \text{ functions and } C_2 \text{ functions}] \\ &= P[C_2 \text{ functions} \mid C_1 \text{ functions}] \cdot P[C_1 \text{ functions}] \\ &= P[C_2 \text{ functions}] \cdot P[C_1 \text{ functions}] \end{aligned}$$

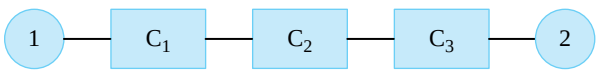

Figure A.7 Three-component series system

where the last step depends on the independence assumption. And in general, if the reliability of component C_i (i.e., $P[C_i \text{ functions}]$) is r_i, then assuming that the k components in a series system behave independently, the **(series) system reliability** (say, R_S), becomes

Series system reliability for independent components

$$R_S = r_1 \cdot r_2 \cdot r_3 \cdot \cdots \cdot r_k \tag{A.11}$$

Example 6 (*Example 5 revisited*)

Space Shuttle Solid Rocket Motor Field Joints as a Series System

The probabilistic risk assessment of Dalal, Fowlkes, and Hoadley put the reliability (at 31°F) of pre-*Challenger* solid rocket motor field joints at .977 apiece. Since the proper functioning of six such joints is necessary for the safe operation of the solid rocket motors, assuming independence of the joints, the reliability of the system of joints is then

$$R_S = (.977)(.977)(.977)(.977)(.977)(.977) = .87$$

as in Example 5. (The .87 figure might well be considered optimistic with regard to the entire solid rocket motor system, as it doesn't take into account any potential problems other than those involving field joints.)

Since typically each r_i is less than 1.0, formula (A.11) shows (as intuitively it should) that system reliability decreases as components are added to a series system. And system reliability is no better (larger) than the worst (smallest) component reliability.

A.2.2 Parallel Systems

In contrast to series system structure is **parallel** system structure.

Definition 10

A system consisting of components $C_1, C_2, C_3, \ldots, C_k$ is called a **parallel system** if its proper functioning requires only the functioning of at least one component.

Figure A.8 is a representation of a parallel system made up of $k = 3$ components. This diagram is interpreted in a manner similar to Figure A.7 (i.e., the system will function provided there is a path from point 1 to point 2 that crosses no failed component).

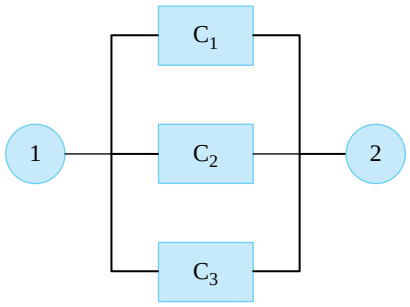

Figure A.8 Three-component parallel system

The fact that made it easy to develop formula (A.11) for the reliability of a series system is that for a series system to function, all components must function. The corresponding fact for a parallel system is that for a parallel system to *fail*, all components must *fail*. So if it is sensible to model the functioning of the individual components in a parallel system as independent, if r_i is the reliability of component i, and if R_p is the **(parallel) system reliability**,

$$\begin{aligned} 1 - R_p &= P[\text{the system fails}] \\ &= P[\text{all components fail}] \\ &= (1 - r_1)(1 - r_2)(1 - r_3)\cdots(1 - r_k) \end{aligned}$$

or

Parallel system reliability for independent components

$$R_p = 1 - (1 - r_1)(1 - r_2)(1 - r_3)\cdots(1 - r_k) \qquad \textbf{(A.12)}$$

Example 7

Parallel Redundancy and Critical Safety Systems

The principle of parallel redundancy is often employed to improve the reliability of critical safety systems. For example, two physically separate automatic shutdown subsystems might be called for in the design of a nuclear power plant. The hope would be that in a rare overheating emergency, at least one of the two would successfully trigger reactor shutdown.

In such a case, if the shutdown subsystems are truly physically separate (so that independence could reasonably be used in a model of their emergency operation), relationship (A.12) might well describe the reliability of the overall safety system. And if, for example, each subsystem is 90% reliable, the overall reliability becomes

$$R_p = 1 - (.10)(.10) = 1 - .01 = .99$$

Expression (A.12) is perhaps a bit harder to absorb than expression (A.11). But the formula functions the way one would intuitively expect. System reliability increases as components are added to a parallel system and is no worse (smaller) than the best (largest) component reliability.

One useful type of calculation that is sometimes done using expression (A.12) is to determine how many equally reliable components of a given reliability r are needed in order to obtain a desired system reliability, R_p. Substitution of r for each r_i in formula (A.12) gives

$$R_p = 1 - (1 - r)^k$$

and this can be solved for an approximate number of components required, giving

Approximate number of components with individual reliability r needed to produce parallel system reliability R_p

$$k \approx \frac{\ln(1 - R_p)}{\ln(1 - r)} \qquad \textbf{(A.13)}$$

Using (for the sake of example) the values $r = .80$ and $R_p = .98$, expression (A.13) gives $k \approx 2.4$, so rounding *up* to an integer, 3 components of individual 80% reliability will be required to give a parallel system reliability of at least 98%.

A.2.3 Combination Series-Parallel Systems

Real engineering systems rarely have purely series or purely parallel structure. However, it is sometimes possible to conceive of system structure as a **combination** of these two basic types. When this is the case, formulas (A.11) and (A.12) can be used to analyze successively larger subsystems until finally an overall reliability prediction is obtained.

Example 8

Predicting Reliability for a System with a Combination of Series and Parallel Structure

In order for an electronic mail message from individual A to reach individual B, the main computers at both A's site and B's site must be functioning, and at least one of three separate switching devices at a communications hub must be working. If the reliabilities for A's computer, B's computer, and each switching device are (respectively) .95, .99, and .90, a plausible figure for the reliability of the A-to-B electronic mail system can be determined as follows.

An appropriate system diagram is given in Figure A.9, with C_A, C_B, C_1, C_2, and C_3 standing (respectively) for the A site computer, the B site computer, and the three switching devices. Although this system is neither purely series nor purely parallel, mentally replacing components C_1, C_2, and C_3 with a single

Example 8
(*continued*)

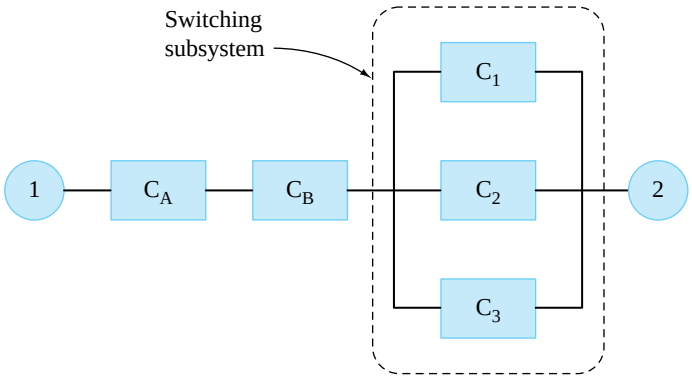

Figure A.9 System diagram for an electronic mail system

"switching subsystem" block, there would be a purely series structure. So,

$$C_1, C_2, \text{ and } C_3 \text{ parallel subsystem reliability} = 1 - (1 - .90)^3 = .999$$

via formula (A.12). Then using formula (A.11),

$$\text{System reliability} = (.95)(.99)(.999) = .94$$

It is clear that the weak link in this communications system is at site A, rather than at B or at the communications hub.

Section 2 Exercises

1. A series system is to consist of $k = 5$ independent components with comparable individual reliabilities. How reliable must each be if the system reliability is to be at least .999? Suppose that it is your job to guarantee components have this kind of individual reliability. Do you see any difficulty in empirically demonstrating this level of component performance? Explain.
2. A parallel system is to consist of k identical independent components. Design requirements are that system reliability be at least .99. Individual component reliability is thought to be at least .90. How large must k be?
3. A combination series-parallel system is to consist of $k = 3$ parallel subsystems, themselves in series. Engineering design requirements are that the entire system have overall reliability at least .99. Two kinds of components are available. Type A components cost \$8 apiece and have reliability .98. Type B components cost \$5 apiece and have reliability .90.
   (a) If only type A components are used, what will be the minimum system cost? If only type B components are used, what will be the minimum system cost?
   (b) Find a system design meeting engineering requirements that uses some components of each type and is cheaper than the best option in part (a).

A.3 Counting

Proposition 3 and Example 4 illustrate that using a model for a chance situation that consists of a finite sample space $\mathcal{S}$ with outcomes judged to be equally likely, the computation of probabilities for events of interest is conceptually a very simple matter. The number of outcomes in the event are simply counted up and divided by the total number of outcomes in the whole sample space. However, in most realistic applications of this simple idea, the process of writing down all outcomes in $\mathcal{S}$ and doing the counting involved would be most tedious indeed, and often completely impractical. Fortunately, there are some simple principles of counting that can often be applied to shortcut the process, allowing outcomes to be counted mentally. The purpose of this section is to present those counting techniques.

This section presents a multiplication principle of counting, the notion of permutations and how to count them, and the idea of combinations and how to count them, along with a few examples. This material is on the very fringe of what is appropriate for inclusion in this book. It is not statistics, nor even really probability, but rather a piece of discrete mathematics that has some engineering implications. It is included here for two reasons. First is the matter of tradition. Counting has traditionally been part of most elementary expositions of probability, because games of chance (cards, coins, and dice) are often assumed to be *fair* and thus describable in terms of sample spaces with equally likely outcomes. And for better or worse, games of chance have been a principal source of examples in elementary probability. A second and perhaps more appealing reason for including the material is that it does have engineering applications (regardless of whether they are central to the particular mission of this text). Ultimately, the reader should take this short section for what it is: a digression from the book's main story that can on occasion be quite helpful in engineering problems.

A.3.1 A Multiplication Principle, Permutations, and Combinations

The fundamental insight of this section is a multiplication principle that is simply stated but wide-ranging in its implications. To emphasize the principle, it will be stated in the form of a proposition.

Proposition 5 (*A Multiplication Principle*)

Suppose a complex action can be thought of as composed of r component actions, the first of which can be performed in n_1 different ways, the second of which can subsequently be performed in n_2 different ways, the third of which can subsequently be performed in n_3 different ways, etc. Then the total number of ways in which the complex action can be performed is

$$n = n_1 \cdot n_2 \cdot \cdots \cdot n_r$$

In graphical terms, this proposition is just a statement that a tree diagram that has n_1 first-level nodes, each of which leads to n_2 second-level nodes, and so on, must end up having a total of $n_1 \cdot n_2 \cdot \cdots \cdot n_r$ rth-level nodes.

Example 9

The Multiplication Principle and Counting the Number of Treatment Combinations in a Full Factorial

The familiar calculation of the number of different possible treatment combinations in a full factorial statistical study is an example of the use of Proposition 5. Consider a $3 \times 4 \times 2$ study in the factors A, B, and C. One may think of the process of writing down a combination of levels for A, B, and C as consisting of $r = 3$ component actions. There are $n_1 = 3$ different ways of choosing a level for A, subsequently $n_2 = 4$ different ways of choosing a level for B, and then subsequently $n_3 = 2$ different ways of choosing a level for C. There are thus

$$n_1 \cdot n_2 \cdot n_3 = 3 \cdot 4 \cdot 2 = 24$$

different ABC combinations.

Example 10

The Multiplication Principle and Counting the Number of Ways of Assigning 4 Out of 100 Pistons to Four Cylinders

Suppose that 4 out of a production run of 100 pistons are to be installed in a particular engine block. Consider the number of possible placements of (distinguishable) pistons into the four (distinguishable) cylinders. One may think of the installation process as composed of $r = 4$ component actions. There are $n_1 = 100$ different ways of choosing a piston for installation into cylinder 1, subsequently $n_2 = 99$ different ways of choosing a piston for installation into cylinder 2, subsequently $n_3 = 98$ different ways of choosing a piston for installation into cylinder 3, and finally, subsequently $n_4 = 97$ different ways of choosing a piston for installation into cylinder 4. There are thus

$$100 \cdot 99 \cdot 98 \cdot 97 = 94{,}109{,}400$$

different ways of placing 4 of 100 (distinguishable) pistons into four (distinguishable) cylinders. (Notice that the job of actually making a list of the different possibilities is not one that is practically doable.)

Example 10 is a generic type of enough importance that there is some special terminology and notation associated with it. The general problem is that of choosing an ordering for r out of n distinguishable objects, or equivalently, placing r out of n distinguishable objects into r distinguishable positions. The application of

Proposition 5 shows that the number of different ways in which this placement can be accomplished is

$$n(n-1)(n-2)\cdots(n-r+1) \tag{A.14}$$

since at each stage of sequentially placing objects into positions, there is one less object available for placement. The special terminology and notation for this are next.

Definition 11

If r out of n distinguishable objects are to be placed in an order 1 to r (or equivalently, placed into r distinguishable positions), each such potential arrangement is called a **permutation**. The number of such permutations possible will be symbolized as $P_{n,r}$, read **"the number of permutations of n things r at a time."**

In the notation of Definition 11, one has (from expression (A.14) that

$$P_{n,r} = n(n-1)(n-2)\cdots(n-r+1)$$

that is,

Formula for the number of permutations of n things r at a time

$$P_{n,r} = \frac{n!}{(n-r)!} \tag{A.15}$$

Example 10 (*continued*)

In the special permutation notation, the number of different ways of installing the four pistons is

$$P_{100,4} = \frac{100!}{96!}$$

Example 11

Permutations and Counting the Number of Possible Circular Arrangements of 12 Turbine Blades

The permutation idea of Definition 11 can be used not only in straightforward ways, as in the previous example, but in slightly more subtle ways as well. To illustrate, consider a situation where 12 distinguishable turbine blades are to be placed into a central disk or hub at successive 30° angles, as sketched in Figure A.10. Notice that if one of the slots into which these blades fit is marked on the

Example 11 (*continued*)

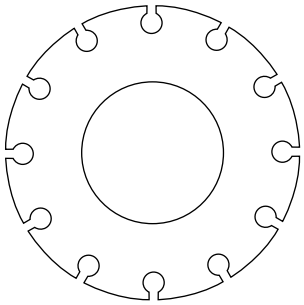

Figure A.10 Hub with 12 slots for blade installation

front face of the hub (and one therefore thinks of the blade positions as completely distinguishable), there are

$$P_{12,12} = 12 \cdot 11 \cdot 10 \cdot \cdots \cdot 2 \cdot 1$$

different possible arrangements of the blades.

But now also consider the problem of counting all possible (circular) arrangements of the 12 blades if relative position is taken into account but absolute position is not. (Moving each blade 30° counterclockwise after first installing them would not create an arrangement different from the first, with this understanding.) The permutation idea can be used here as well, thinking as follows. Placing blade 1 anywhere in the hub essentially establishes a point of reference and makes the remaining 11 positions distinguishable (relative to the point of reference). One then has 11 distinguishable blades to place in 11 distinguishable positions. Thus, there must be

$$P_{11,11} = 11 \cdot 10 \cdot 9 \cdot \cdots \cdot 2 \cdot 1$$

such circular arrangements of the 12 blades, where only relative position is considered.

A second generic counting problem is related to the permutation idea and is particularly relevant in describing simple random sampling. That is the problem of choosing an unordered collection of r out of n distinguishable objects. The special terminology and notation associated with this generic problem are as follows.

Definition 12

If an unordered collection of r out of n distinguishable objects is to be made, each such potential collection is called a **combination**. The number of such

combinations possible will be symbolized as $\binom{n}{r}$, read "the number of combinations of n things r at a time."

There is in Definition 12 a slight conflict in terminology with other usage in this text. The "combination" in Definition 12 is not the same as the "treatment combination" terminology used in connection with multifactor statistical studies to describe a set of conditions under which a sample is taken. (The "treatment combination" terminology has been used in this very section in Example 9.) But this conflict rarely causes problems, since the intended meaning of the word *combination* is essentially always clear from context.

Appropriate use of Proposition 5 and formula (A.15) makes it possible to develop a formula for $\binom{n}{r}$ as follows. A permutation of r out of n distinguishable objects can be created through a two-step process. First a combination of r out of the n objects is selected and then those selected objects are placed in an order. This thinking suggests that $P_{n,r}$ can be written as

$$P_{n,r} = \binom{n}{r} \cdot P_{r,r}$$

But this means that

$$\frac{n!}{(n-r)!} = \binom{n}{r}\frac{r!}{0!}$$

that is,

Formula for the number of combinations of n things r at a time

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} \qquad \textbf{(A.16)}$$

The ratio in equation (A.16) ought to look familiar to readers who have studied Section 5.1. The multiplier of $p^x(1-p)^{n-x}$ in the binomial probability function is of the form $\binom{n}{x}$, counting up the number of ways of placing x successes in a series of n trials.

Example 12
(*Example 10, Chapter 3, revisited—page 105*)

Combinations and Counting the Numbers of Possible Samples of Cable Connectors with Prescribed Defect Class Compositions

In the cable connector inspection scenario of Delva, Lynch, and Stephany, 3,000 inspections of connectors produced 2,985 connectors classified as having no defects, 1 connector classified as having only minor defects, and 14 others classified as having moderately serious, serious, or very serious defects. Suppose that in an effort to audit the work of the inspectors, a sample of 100 of the 3,000 previously inspected connectors is to be reinspected.

Example 12 (*continued*)

Then notice that directly from expression (A.16), there are in fact

$$\binom{3000}{100} = \frac{3000!}{100!\,2900!}$$

different (unordered) possible samples for reinspection. Further, there are

$$\binom{2985}{100} = \frac{2985!}{100!\,2885!}$$

different possible samples of size 100 made up of only connectors judged to be defect-free. If (for some reason) the connectors to be reinspected were to be chosen as a simple random sample of the 3,000, the ratio

$$\frac{\binom{2985}{100}}{\binom{3000}{100}}$$

would then be a sensible figure to use for the probability that the sample is composed entirely of connectors initially judged to be defect-free.

It is instructive to take this example one step further and combine the use of Definition 12 and Proposition 5. So consider the problem of counting up the number of different samples containing 96 connectors initially judged defect-free, 1 judged to have only minor defects, and 3 judged to have moderately serious, serious, or very serious defects. To solve this problem, the creation of such a sample can be considered as a three-step process. In the first, 96 nominally defect-free connectors are chosen from 2,985. In the second, 1 connector nominally having minor defects only is chosen from 1. And finally, 3 connectors are chosen from the remaining 14. There are thus

$$\binom{2985}{96} \cdot \binom{1}{1} \cdot \binom{14}{3}$$

different possible samples of this rather specialized type.

Example 13

An Application of Counting Principles to the Calculation of a Probability in a Scenario of Equally Likely Outcomes

As a final example in this section, most of the ideas discussed here can be applied to the computation of a probability in another situation of equally likely outcomes where writing out a list of the possible outcomes is impractical.

Consider a hypothetical situation where 15 manufactured devices of a particular kind are to be sent 5 apiece to three different testing labs. Suppose further

that 3 of the seemingly identical devices are defective. Consider the probability that each lab receives 1 defective device, if the assignment of devices to labs is done at random.

The total number of possible assignments of devices to labs can be computed by thinking first of choosing 5 of 15 to send to Lab A, then 5 of the remaining 10 to send to Lab B, then sending the remaining 5 to Lab C. There are thus

$$\binom{15}{5} \cdot \binom{10}{5} \cdot \binom{5}{5}$$

such possible assignments of devices to labs.

On the other hand, if each lab is to receive 1 defective device, there are $P_{3,3}$ ways to assign defective devices to labs and then subsequently $\binom{12}{4} \cdot \binom{8}{4} \cdot \binom{4}{4}$ possible ways of completing the three shipments. So ultimately, an appropriate probability assignment for the event that each lab receives 1 defective device is

$$\frac{P_{3,3} \cdot \binom{12}{4} \cdot \binom{8}{4} \cdot \binom{4}{4}}{\binom{15}{5} \cdot \binom{10}{5} \cdot \binom{5}{5}} = \frac{3 \cdot 2 \cdot 1 \cdot 12! \cdot 8! \cdot 5! \cdot 10! \cdot 5! \cdot 5!}{15! \cdot 10! \cdot 4! \cdot 8! \cdot 4! \cdot 4!}$$

$$= \frac{3 \cdot 2 \cdot 1 \cdot 5 \cdot 5 \cdot 5}{15 \cdot 14 \cdot 13}$$

$$= .27$$

Section 3 Exercises

1. A lot of 100 machine parts contains 10 with diameters that are out of specifications on the low side, 20 with diameters that are out of specifications on the high side, and 70 that are in specifications.

(a) How many different possible samples of $n = 10$ of these parts are there?

(b) How many different possible samples of size $n = 10$ are there that each contain exactly 1 part with diameter out of specifications on the low side, 2 parts with diameters out of specifications on the high side, and 7 parts with diameters that are in specifications?

(c) Based on your answers to (a) and (b), what is the probability that a simple random sample of $n = 10$ of these contains exactly 1 part with diameter out of specifications on the low side, 2 parts with diameters out of specifications on the high side, and 7 parts with diameters that are in specifications?

2. The lengths of bolts produced in a factory are checked with two "go–no go" gauges and the bolts sorted into piles of short, OK, and long bolts. Suppose that of the bolts produced, about 20% are short, 30% are long, and 50% are OK.

(a) Find the probability that among the next ten bolts checked, the first three are too short, the next three are OK, and the last four are too long.

(b) Find the probability that among the next ten bolts checked, there are three that are too short, three that are OK, and four that are too long. (*Hint:* In how many ways it is possible to choose three of the group to be short, three

to be OK, and four to be long? Then use your answer to (a).)

3. User names on a computer system consist of three letters A through Z, followed by two digits 0 through 9. (Letters and digits may appear more than once in a name.)
   (a) How many user names of this type are there?
   (b) Suppose that Joe has user name TPK66, but unfortunately he's forgotten it. Joe remembers only the format of the user names and that the letters K, P, and T appear in his name. If he picks a name at random from those consistent with his memory, what's the probability that he selects his own?
   (c) If Joe in part (b) also remembers that his digits match, what's the probability that he selects his own user name?
4. A lot contains ten pH meters, three of which are miscalibrated. A technician selects these meters one at a time, at random without replacement, and checks their calibration.
   (a) What is the probability that among the first four meters selected, exactly one is miscalibrated?
   (b) What is the probability that the technician discovers his second miscalibrated meter when checking his fifth one?
5. A student decides to use the random digit function on her calculator to select a three-digit PIN number for use with her new ATM card. (Assume that all numbers 000 through 999 are then equally likely to be chosen.)
   (a) What is the probability that her number uses only odd digits?
   (b) What is the probability that all three digits in her number are different?
   (c) What is the probability that her number uses three different digits and lists them in either ascending or descending order?
6. When ready to configure a PC order, a consumer must choose a Processor Chip, a MotherBoard, a Drive Controller and a Hard Drive. The choices are:

Processor Chip	Mother-Board	Drive Controller	Hard Drive
Fast New Generation	Premium	Premium	Premium
Slow New Generation	Standard	Standard	Standard
Fast Old Generation		Economy	Economy
Slow Old Generation			

   (a) Suppose initially that all components are compatible with all components. How many different configurations are possible?

   Suppose henceforth that:
   (i) a Premium MotherBoard is needed to run a New Generation Processor,
   (ii) a Premium MotherBoard is needed to run a Premium Drive Controller, and
   (iii) a Premium Drive Controller is needed to run a Premium Hard Drive.
   (b) How many permissible configurations are there with a Standard MotherBoard?
   (c) How many permissible configurations are there total? Explain carefully.

A.4 Probabilistic Concepts Useful in Survival Analysis

Section A.2 is meant to provide only the most elementary insights into how probability might prove useful in the context of reliability modeling and prediction. The ideas discussed in that section are of an essentially "static" nature. They are most appropriate when considering the likelihood of a system performing adequately at a single point in time—for example, at its installation, or at the end of its warranty period.

Reliability engineers also concern themselves with matters possessing a more dynamic flavor, having to do with the modeling and prediction of life-length variables associated with engineering systems and their components. It is outside the intended scope of this text to provide anything like a serious introduction to the large body of methods available for probability modeling and subsequent formal inference for such variables. But what will be done here is to provide some material (supplementary to that found in Section 5.2) that is part of the everyday jargon and intellectual framework of reliability engineering. This section will consider several descriptions and constructs related to continuous random variables that, like system or component life lengths, take only positive values.

A.4.1 Survivorship and Force-of-Mortality Functions

In this section, T will stand for a continuous random variable taking only nonnegative values. The reader may think of T as the time till failure of an engineering component. In Section 5.2, continuous random variables X (or more properly, their distributions) were described through their probability densities $f(x)$ and cumulative probability functions $F(x)$. In the present context of lifetime random variables, there are several other, more popular ways of conveying the information carried by $f(t)$ or $F(t)$. Two of these devices are introduced next.

Definition 13

The **survivorship function** for a nonnegative random variable T is the function

$$S(t) = P[T > t] = 1 - F(t)$$

The survivorship function is also sometimes known as the **reliability function**. It specifies the probability that the component being described survives beyond time t.

Example 14

The Survivorship Function and Diesel Engine Fan Blades

Data on 70 diesel engines of a single type (given in Table 1.1 of Nelson's *Applied Life Data Analysis*) indicate that lifetimes in hours of a certain fan on such engines could be modeled using an exponential distribution with mean $\alpha \approx 27{,}800$. So from Definition 17 in Chapter 5, to describe a fan lifetime T, one could use the density

$$f(t) = \begin{cases} 0 & \text{if } t < 0 \\ \dfrac{1}{27{,}800} e^{-t/27{,}800} & \text{if } t > 0 \end{cases}$$

Example 14 (*continued*)

or the cumulative probability function

$$F(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ 1 - e^{-t/27{,}800} & \text{if } t > 0 \end{cases}$$

or from Definition 13, the survivorship function

$$S(t) = \begin{cases} 1 & \text{if } t \leq 0 \\ e^{-t/27{,}800} & \text{if } t > 0 \end{cases}$$

The probability of a fan surviving at least 10,000 hours is then

$$S(10{,}000) = e^{-10{,}000/27{,}800} = .70$$

A second way of specifying the distribution of a life-length variable (unlike anything discussed in Section 5.2) is through a function giving a kind of "instantaneous rate of death of survivors."

Definition 14

The **force-of-mortality function** for a nonnegative continuous random variable T is, for $t > 0$, the function

$$h(t) = \frac{f(t)}{S(t)}$$

$h(t)$ is sometimes called the *hazard function* for T, but such usage tends to perpetuate unfortunate confusion with the *entirely different* concept of "hazard rate" for repairable systems. (The important difference between the two concepts is admirably explained in the paper "On the Foundations of Reliability" by W. A. Thompson (*Technometrics*, 1981) and in the book *Repairable Systems Reliability* by Ascher and Feingold.) This book will thus stick to the term *force of mortality*.

The force-of-mortality function can be thought of heuristically as

$$h(t) = \frac{f(t)}{S(t)} = \lim_{\Delta \to 0} \frac{P[t < T < t + \Delta]/\Delta}{P[t < T]} = \lim_{\Delta \to 0} \frac{P[t < T < t + \Delta \mid t < T]}{\Delta}$$

which is indeed a sort of "death rate of survivors at time t."

Example 14 (*continued*)

The force-of-mortality function for the diesel engine fan example is, for $t > 0$,

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{27{,}800} e^{-t/27{,}800}}{e^{-t/27{,}800}} = \frac{1}{27{,}800}$$

The exponential (mean $\alpha = 27{,}800$) model for fan life implies a constant $\left(\frac{1}{27{,}800}\right)$ force of mortality.

Constant force of mortality is equivalent to exponential distribution

The property of the fan-life model shown in the previous example is characteristic of exponential distributions. That is, a distribution has constant force of mortality exactly when that distribution is exponential. So having a constant force of mortality is equivalent to possessing the *memoryless* property of the exponential distributions discussed in Section 5.2. If the lifetime of an engineering component is described using a constant force of mortality, there is no (mathematical) reason to replace such a component before it fails. The distribution of its remaining life from any point in time is the same as the distribution of the time till failure of a new component of the same type.

Potential probability models for lifetime random variables are often classified according to the nature of their force-of-mortality functions, and these classifications are taken into account when selecting models for reliability engineering applications. If $h(t)$ is increasing in t, the corresponding distribution is called an **increasing force-of-mortality (IFM) distribution**, and if $h(t)$ is decreasing in t, the corresponding distribution is called a **decreasing force-of-mortality (DFM) distribution**. The reliability engineering implications of an IFM distribution being appropriate for modeling the lifetimes of a particular type of component are often that (as a form of preventative maintenance) such components are retired from service once they reach a particular age, even if they have not failed.

Example 15

The Weibull Distributions and Their Force-of-Mortality Functions

The Weibull family of distributions discussed in Section 5.2 is commonly used in reliability engineering contexts. Using formulas (5.26) and (5.27) of Section 5.2 for the Weibull cumulative probability function and probability density, the Weibull force-of-mortality function for shape parameter β and scale parameter α is, for $t > 0$

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = \frac{\frac{\beta}{\alpha^{\beta}} t^{\beta-1} e^{-(t/\alpha)^{\beta}}}{e^{-(t/\alpha)^{\beta}}} = \frac{\beta t^{\beta-1}}{\alpha^{\beta}}$$

For $\beta = 1$ (the exponential distribution case) this is constant. For $\beta < 1$, this is decreasing in t, and the Weibull distributions with $\beta < 1$ are DFM distributions. For $\beta > 1$, this is increasing in t, and the Weibull distributions with $\beta > 1$ are IFM distributions.

Example 16

Force-of-Mortality Function for a Uniform Distribution

As an artificial but instructive example, consider the use of a uniform distribution on the interval (0, 1) as a life-length model. With

$$f(t) = \begin{cases} 1 & \text{if } 0 < t < 1 \\ 0 & \text{otherwise} \end{cases}$$

the survivorship function is

$$S(t) = \begin{cases} 1 & \text{if } t < 0 \\ 1 - t & \text{if } 0 \le t < 1 \\ 0 & \text{if } 1 \le t \end{cases}$$

so

$$h(t) = \frac{1}{1-t} \qquad \text{if } 0 < t < 1$$

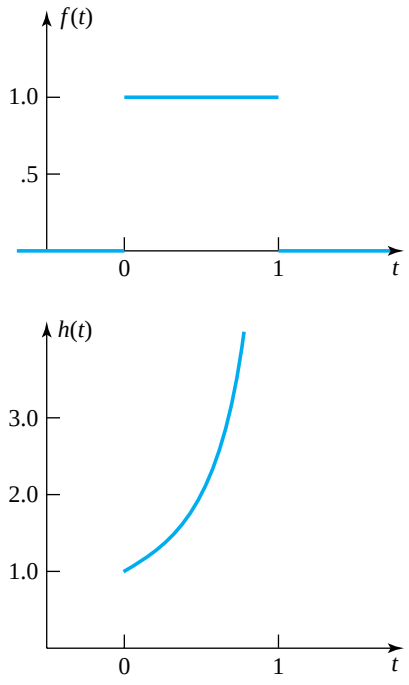

Figure A.11 Probability density and force-of-mortality function for a uniform distribution

Figure A.11 shows plots of both $f(t)$ and $h(t)$ for the uniform model. $h(t)$ is clearly increasing for $0 < t < 1$ (quite drastically so, in fact, as one approaches $t = 1$). And well it should be. Knowing that (according to the uniform model) life will certainly end by $t = 1$, nervousness about impending death should skyrocket as one nears $t = 1$.

Conventional wisdom in reliability engineering is that many kinds of manufactured devices have life distributions that ought to be described by force-of-mortality functions qualitatively similar to the hypothetical one sketched in Figure A.12.

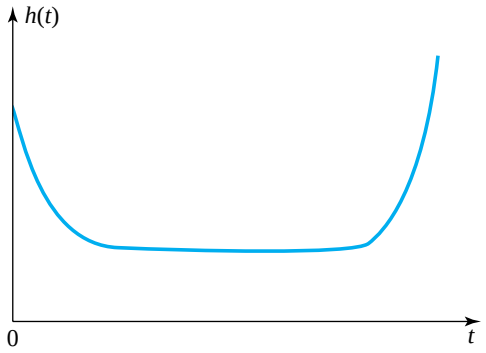

Figure A.12 A "bathtub curve" force-of-mortality function

The shape in Figure A.12 is often referred to as the **bathtub curve** shape. It includes an early region of decreasing force of mortality, a long central period of relatively constant force of mortality, and a late period of rapidly increasing force of mortality. Devices with lifetimes describable as in Figure A.12 are sometimes subjected to a **burn-in period** to eliminate the devices that will fail in the early period of decreasing force of mortality, and then sold with the recommendation that they be replaced before the onset of the late period of increasing force of mortality or **wear-out**. Although this story is intuitively appealing, the most tractable models for life length do not, in fact, have force-of-mortality functions with shapes like that in Figure A.12. For a further discussion of this matter and references to papers presenting models with bathtub-shaped force-of-mortality functions, refer to Chapter 2 of Nelson's *Applied Life Data Analysis*.

The functions $f(t)$, $F(t)$, $S(t)$, and $h(t)$ all carry the same information about a life distribution. They simply express it in different terms. Given one of them, the derivation of the others is (at least in theory) straightforward. Some of the

relationships that exist among the four different characterizations are collected here for the reader's convenience. For $t > 0$,

Relationships between $F(t)$, $f(t)$, $S(t)$, and $h(t)$

$$F(t) = \int_0^t f(x)\,dx$$

$$f(t) = \frac{d}{dt}F(t)$$

$$S(t) = 1 - F(t)$$

$$h(t) = \frac{f(t)}{S(t)}$$

$$S(t) = \exp\left(-\int_0^t h(x)\,dx\right)$$

$$f(t) = h(t)\exp\left(-\int_0^t h(x)\,dx\right)$$

Section 4 Exercises

1. An engineer begins a series of presentations to his corporate management with a working bulb in his slide projector and (an inferior-quality) Brand W replacement bulb in his briefcase. Suppose that the random variables

X = the number of hours of service given by the bulb in the projector

Y = the number of hours of service given by the spare bulb

may be modeled as independent exponential random variables with respective means 15 and 5. The number of hours that the engineer may operate without disaster is $X + Y$.

(a) Find the mean and standard deviation of $X + Y$ using Proposition 1 in Chapter 5.
(b) Find, for $t > 0$, $P[X + Y \le t]$.
(c) Use your answer to (b) and find the probability density for $T = X + Y$.
(d) Find the survivorship and force-of-mortality functions for T. What is the nature of the force-of-mortality function? Is it constant like those of X and Y?

2. A common modeling device in reliability applications is to assume that the (natural) logarithm of a lifetime variable, T, has a normal distribution. That is, one might suppose that for some parameters μ and σ, if $t > 0$

$$F(t) = P[T \le t] = \Phi\left(\frac{\ln t - \mu}{\sigma}\right)$$

Consider the $\mu = 0$ and $\sigma = 1$ version of this.

(a) Plot $F(t)$ versus t.
(b) Plot $S(t)$ versus t.
(c) Compute and plot $f(t)$ versus t.
(d) Compute and plot $h(t)$ versus t.
(e) Is this distribution for T an IFM distribution, a DFM distribution, or neither? What implication does your answer have for in-service replacement of devices possessing this lifetime distribution?

A.5 Maximum Likelihood Fitting of Probability Models and Related Inference Methods

The model-fitting and inference methods discussed in this text are, for the most part, methods for independent, *normally distributed* observations. This is in spite of the fact that there are strong hints in Chapter 5 and this appendix that other kinds of probability models often prove useful in engineering problem solving. (For example, binomial, geometric, Poisson, exponential, and Weibull distributions have been discussed, and parts of Sections A.1 and A.2 should suggest that probability models not even necessarily involving these standard distributions will often be helpful.) It thus seems wise to present at least a brief introduction to some principles of probability-model fitting and inference that can be applied more generally than to only scenarios involving independent, normal observations. This will be done to give at least the flavor of what is possible, as well as an idea of some kinds of things likely to be found in the engineering statistics literature.

This section considers the use of likelihood functions in the fitting of parametric probability models and in large-sample inference for the model parameters. It begins by discussing the idea of a likelihood function and maximum likelihood model fitting for discrete data. Similar discussions are then conducted for continuous and mixed data. Finally, there is a discussion of how for large samples, approximate confidence regions and tests can often be developed using likelihood functions.

A.5.1 Likelihood Functions for Discrete Data and Maximum Likelihood Model Fitting

To begin, consider scenarios where the outcome of a chance situation can be described in terms of a data vector of jointly discrete random variables (or a single discrete random variable) $\boldsymbol{Y}$, whose probability function f depends on some (unknown) vector of parameters (or single parameter) $\boldsymbol{\Theta}$. To make the dependence of f on $\boldsymbol{\Theta}$ explicit, this section will use the notation

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y})$$

for the (joint) probability function of $\boldsymbol{Y}$.

Chapter 5 made heavy use of particular parametric probability functions, primarily thinking of them as functions of $\boldsymbol{y}$. In this section, it will be very helpful to shift perspective. With data $\boldsymbol{Y} = \boldsymbol{y}$ in hand, think of

A discrete data likelihood function

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y}) \tag{A.17}$$

or (often more conveniently) its natural logarithm

A discrete data log likelihood function

$$L(\boldsymbol{\Theta}) = \ln\left(f_{\boldsymbol{\Theta}}(\boldsymbol{y})\right) \tag{A.18}$$

as functions of $\boldsymbol{\Theta}$, specifying for various possible vectors of parameters "how likely" it would be to observe the particular data in hand. With this perspective, the function (A.17) is often called the **likelihood function** and function (A.18) the **log likelihood function** for the problem under consideration.

Example 17
(*Example 4, Chapter 5, revisited—page 235*)

The Log Likelihood Function for the Number Conforming in a Sample of Hexamine Pellets

In the pelletizing machine example used in Chapter 5 and earlier, it is possible to argue that under stable conditions,

$$X = \text{the number of conforming pellets produced in a batch of 100 pellets}$$

might well be modeled using a binomial distribution with $n = 100$ and p some unknown parameter. The corresponding probability function is thus

$$f(x) = \begin{cases} \dfrac{100!}{x!\,(100-x)!}\, p^x (1-p)^{100-x} & x = 0, 1, \ldots, 100 \\ 0 & \text{otherwise} \end{cases}$$

Should one observe $X = 66$ conforming pellets be observed in a batch, the material just introduced says that the function of p

$$L(p) = \ln(f(66)) = \ln(100!) - \ln(66!) - \ln(34!) + 66\ln(p) + 34\ln(1-p) \tag{A.19}$$

is an appropriate log likelihood function. Figure A.13 is a sketch of $L(p)$ for this problem. Notice that (in an intuitively appealing fashion) the value of p maximizing $L(p)$ is

$$\hat{p} = \frac{66}{100} = .66$$

That is, $p = .66$ makes the chance of observing the particular data in hand ($X = 66$) as large as possible.

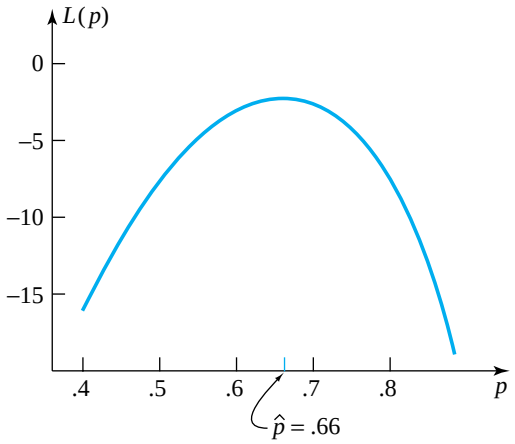

Figure A.13 Plot of the log likelihood function based on 66 conforming tablets out of 100

Example 18

The Log Likelihood Function for *n* Independent Poisson Observations

As a second, somewhat more abstract, example of the idea of a likelihood function, suppose that $X_1, X_2, \ldots, X_n$ are independent Poisson random variables, X_i with mean $k_i\lambda$ for $k_1, k_2, \ldots, k_n$ known constants, and λ an unknown parameter. Such a model might, for example, be appropriate in a quality monitoring context, where at time i, k_i standard-size units of product are inspected, X_i defects are observed, and λ is a constant mean defects per unit.

The joint probability function for $X_1, X_2, \ldots, X_n$ is

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} \displaystyle\prod_{i=1}^{n} \frac{e^{-k_i\lambda}(k_i\lambda)^{x_i}}{x_i!} & \text{for each } x_i \text{ a nonnegative integer} \\ 0 & \text{otherwise} \end{cases}$$

The log likelihood function in the present context is thus

$$L(\lambda) = -\lambda \sum_{i=1}^{n} k_i + \sum_{i=1}^{n} x_i \ln(k_i) + \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \ln(x_i!) \qquad \textbf{(A.20)}$$

The likelihood functions in Examples 17 and 18 are for individual (univariate) parameters. The next example involves two parameters.

Example 19

A Log Likelihood Function Based on Pre-*Challenger* Space Shuttle O-Ring Failure Data

Table A.2 contains pre-*Challenger* data on field joint primary O-ring failures on 23 (out of 24) space shuttle flights. (On one flight, the rocket motors were lost at sea, so no data are available.) The failure counts $x_1, x_2, \ldots, x_{23}$ are the numbers (out of 6 possible) of primary O-rings showing evidence of erosion or blow-by in postflight inspections of the solid rocket motors, and $t_1, t_2, \ldots, t_{23}$ are the corresponding *temperatures at the times of launch*.

Table A.2
Pre-*Challenger* Field Joint Primary O-Ring Failure Data

Flight Date	x, Number of Field Joint Primary O-Ring Incidents	t, Temperature at Launch (°F)
4/12/81	0	66
11/12/81	1	70
3/22/82	0	69
11/11/82	0	68
4/4/83	0	67
6/18/83	0	72
8/30/83	0	73
11/28/83	0	70
2/3/84	1	57
4/6/84	1	63
8/30/84	1	70
10/5/84	0	78
11/8/84	0	67
1/24/85	2	53
4/12/85	0	67
4/29/85	0	75
6/17/85	0	70
7/29/85	0	81
8/27/85	0	76
10/3/85	0	79
10/30/85	2	75
11/26/85	0	76
1/12/86	1	58

In "Risk Analysis of the Space Shuttle: Pre-*Challenger* Prediction of Failure" (*Journal of the American Statistical Association*, 1989), Dalal, Fowlkes, and Hoadley considered several analyses of the data in Table A.2 (and other pre-*Challenger* shuttle failure data). In one of their analyses of the data given here, Dalal, Fowlkes, and Hoadley used a likelihood approach based on the observations

$$y_i = \begin{cases} 1 & \text{if } x_i \geq 1 \\ 0 & \text{if } x_i = 0 \end{cases}$$

that indicate which flights experienced primary O-ring incidents. (They also considered a likelihood approach based on the counts x_i themselves. But here only the slightly simpler analysis based on the y_i's will be discussed.) The authors modeled $Y_1, Y_2, \ldots, Y_{23}$ as a priori independent variables and treated the probability of at least one O-ring incident on flight i,

$$p_i = P[Y_i = 1] = P[X_i \geq 1]$$

as a function of (temperature) t_i. The particular form of dependence of p_i on t_i used by the authors was a "linear (in t) log odds" form

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta t \tag{A.21}$$

for α and β some unknown parameters. Equation (A.21) can be solved for p to produce the function of t

$$p(t) = \frac{1}{1 + e^{-(\alpha+\beta t)}} \tag{A.22}$$

From either equation (A.21) or (A.22), it is possible to see that if $\beta > 0$, the probability of at least one O-ring incident is increasing in t (low-temperature launches are best). On the other hand, if $\beta < 0$, p is decreasing in t (high-temperature launches are best).

The joint probability function for $Y_1, Y_2, \ldots, Y_{23}$ employed by Dalal, Fowlkes, and Hoadley was then

$$f(y_1, y_2, \ldots, y_{23}) = \begin{cases} \displaystyle\prod_{i=1}^{23} p(t_i)^{y_i} \left(1 - p(t_i)\right)^{1-y_i} & \text{for each } y_i = 0 \text{ or } 1 \\ 0 & \text{otherwise} \end{cases}$$

Example 19
(*continued*)

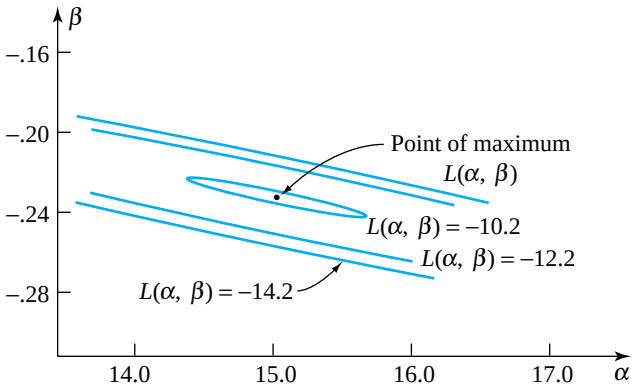

Figure A.14 Contour plot of the Dalal, Fowlkes, and Hoadley log likelihood function

The log likelihood function is then (using equations (A.21) and (A.22))

$$\begin{aligned} L(\alpha,\beta) &= \sum_{i=1}^{23} y_i \ln\left(\frac{p(t_i)}{1-p(t_i)}\right) + \sum_{i=1}^{23} \ln\left(1-p(t_i)\right) \\ &= \sum_{i=1}^{23} y_i(\alpha+\beta t_i) + \sum_{i=1}^{23} \ln\left(\frac{e^{-(\alpha+\beta t_i)}}{1+e^{-(\alpha+\beta t_i)}}\right) \\ &= 7\alpha + \beta(70+57+63+70+53+75+58) \\ &\quad + \ln\left(\frac{e^{-(\alpha+66\beta)}}{1+e^{-(\alpha+66\beta)}}\right) + \ln\left(\frac{e^{-(\alpha+70\beta)}}{1+e^{-(\alpha+70\beta)}}\right) \\ &\quad + \cdots + \ln\left(\frac{e^{-(\alpha+58\beta)}}{1+e^{-(\alpha+58\beta)}}\right) \end{aligned} \qquad \textbf{(A.23)}$$

where the sum abbreviated in equation (A.23) is over all 23 t_i's. Figure A.14 is a contour plot of $L(\alpha, \beta)$ given in equation (A.23).

It is interesting (and sadly, of great engineering importance) that the region of (α, β) pairs making the data of Table A.2 most likely is in the $\beta < 0$ part of the (α, β)-plane—that is, where $p(t)$ is decreasing in t (i.e., increases as t falls). (Remember that the tragic *Challenger* launch was made at $t = 31^\circ$.)

The binomial and Poisson examples of discrete-data likelihoods given thus far have arisen from situations that are most naturally thought of as intrinsically discrete. However, the details of how engineering data are collected sometimes lead to the production of essentially discrete data from intrinsically continuous

variables. For example, consider a life test of some electrical components, where a technician begins a test by connecting 50 devices to a power source, goes away, and then returns every ten hours to note which devices are still functioning. The details of data collection produce only discrete data (which ten-hour period produces failure) from the intrinsically continuous life lengths of the 50 devices. The next example shows how the likelihood idea might be used in another situation where the underlying phenomenon is continuous.

Example 20

A Log Likelihood Function for a Crudely Gauged Normally Distributed Dimension of Five Machined Metal Parts

In many contexts where industrial process monitoring involves relatively stable processes and relatively crude gauging, intrinsically continuous product characteristics are measured and recorded as essentially discrete data. For example, Table A.3 gives values (in units of .0001 in. over nominal) of a critical dimension measured on a sample of $n = 5$ consecutive metal parts produced by a CNC lathe.

It might make sense to model underlying values of this critical dimension as normal, with some (unknown) mean μ and some (unknown) standard deviation σ, but nonetheless to want to explicitly recognize the discreteness of the recorded data. One way of doing so in this context is to think of the observed values as arising (after coding) from rounding normally distributed dimensions to the nearest integer. For a single metal part, this would mean that for any integer y,

$$\begin{aligned} P[\text{the value recorded is } y] &= P[\text{the actual dimension is between } y-.5 \text{ and } y+.5] \\ &= \Phi\left(\frac{y+.5-\mu}{\sigma}\right) - \Phi\left(\frac{y-.5-\mu}{\sigma}\right) \end{aligned} \quad \textbf{(A.24)}$$

Table A.3
Measurements of a Critical Dimension on Five Metal Parts Produced on a CNC Lathe

Part	Measured Dimension, y
1	4
2	3
3	3
4	2
5	3

Example 20
(*continued*)

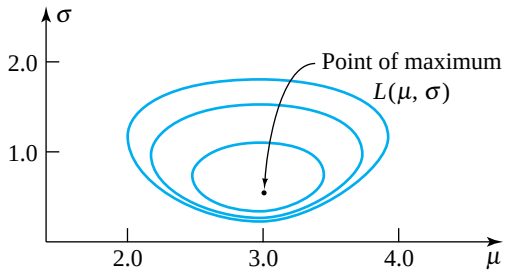

Figure A.15 Contour plot of the "rounded normal data" log likelihood for the data of Table A.3

So treating $n = 5$ consecutive recorded dimensions as independent, equation (A.24) leads to the joint probability function

$$f(y_1, y_2, \ldots, y_5) = \prod_{i=1}^{5} \left\{ \Phi\left(\frac{y_i + .5 - \mu}{\sigma}\right) - \Phi\left(\frac{y_i - .5 - \mu}{\sigma}\right) \right\}$$

and log likelihood function for the data in Table A.3

$$\left.\begin{aligned} L(\mu, \sigma) = \ln &\left(\Phi\left(\frac{2 + .5 - \mu}{\sigma}\right) - \Phi\left(\frac{2 - .5 - \mu}{\sigma}\right) \right) \\ &+ 3 \ln \left(\Phi\left(\frac{3 + .5 - \mu}{\sigma}\right) - \Phi\left(\frac{3 - .5 - \mu}{\sigma}\right) \right) \\ &+ \ln \left(\Phi\left(\frac{4 + .5 - \mu}{\sigma}\right) - \Phi\left(\frac{4 - .5 - \mu}{\sigma}\right) \right) \end{aligned}\right\} \quad \textbf{(A.25)}$$

Figure A.15 is a contour plot of $L(\mu, \sigma)$.

Consideration of a likelihood function $f_{\boldsymbol{\Theta}}(\boldsymbol{y})$ or its log version $L(\boldsymbol{\Theta})$ can be thought of as a way of assessing how compatible various probability models indexed by $\boldsymbol{\Theta}$ are with the data in hand, $\boldsymbol{Y} = \boldsymbol{y}$. Different parameter vectors $\boldsymbol{\Theta}$ having the same value of $L(\boldsymbol{\Theta})$ can be viewed as equally compatible with data in hand. A value of $\boldsymbol{\Theta}$ maximizing $L(\boldsymbol{\Theta})$ might then be considered to be as compatible with the observed data as is possible. This value is often termed the **maximum likelihood estimate** of the parameter vector $\boldsymbol{\Theta}$. Finding maximum likelihood estimates of parameters is a very common method of fitting probability models to data. In simple situations, calculus can sometimes be employed to see how to maximize $L(\boldsymbol{\Theta})$, but in most nonstandard situations, numerical or graphical methods are required.

Example 17 (*continued*)

In the pelletizing example, simple investigation of Figure A.13 shows

$$\hat{p} = \frac{66}{100}$$

to maximize $L(p)$ given in display (A.19) and thus to be the maximum likelihood estimate of p. The reader is encouraged to verify that by differentiating $L(p)$ with respect to p, setting the result equal to 0, and solving for p, this maximizing value can also be found analytically.

Example 18 (*continued*)

Differentiating the log likelihood (A.20) with respect to λ, one obtains

$$\frac{d}{d\lambda} L(\lambda) = -\sum_{i=1}^{n} k_i + \frac{1}{\lambda} \sum_{i=1}^{n} x_i$$

Setting this derivative equal to 0 and solving for λ produces

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} k_i} = \hat{u}$$

which is the total number of defects observed divided by the total number of units inspected. Since the second derivative of $L(\lambda)$ is easily seen to be negative for all λ, $\hat{u}$ is the unique maximizer of $L(\lambda)$—that is, the maximum likelihood estimate of λ.

Example 19 (*continued*)

Careful examination of contour plots like Figure A.14, or use of a numerical search method for the (α, β) pair maximizing $L(\alpha, \beta)$, produces maximum likelihood estimates

$$\hat{\alpha} = 15.043$$
$$\hat{\beta} = -.2322$$

based on the pre-*Challenger* data. Figure A.16 is a plot of $p(t)$ given in display (A.22) for these values of α and β. Notice the disconcerting fact that the corresponding estimate of $p(31)$ (the probability of at least one O-ring failure in a 31° launch) exceeds .99. ($t = 31$ is clearly a huge extrapolation away from any t values in Table A.2, but even so, this kind of analysis conducted before the *Challenger* launch could well have helped cast legitimate doubt on the advisability of a low-temperature launch.)

Example 19 (*continued*)

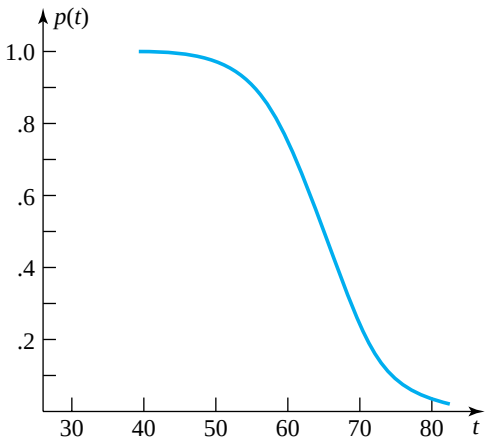

Figure A.16 Plot of fitted probability of at least one O-ring failure as a function of shuttle launch temperature

Example 20 (*continued*)

Examination of the contour plot in Figure A.15 shows maximum likelihood estimates of μ and σ based on the rounded normal data model and the data in Table A.3 to be approximately

$$\hat{\mu} = 3.0$$
$$\hat{\sigma} = .55$$

It is worth noting that for these data, $s = .71$, which is noticeably larger than $\hat{\sigma}$. This illustrates a well-established piece of statistical folklore. It is fairly well known that to ignore rounding of intrinsically continuous data will typically have the effect of inappropriately inflating the apparent spread of the underlying distribution.

A.5.2 Likelihood Functions for Continuous and Mixed Data and Maximum Likelihood Model Fitting

The likelihood function ideas discussed thus far depend on treating the $\boldsymbol{\Theta}$ *probability* of discrete data in hand, $\boldsymbol{Y} = \boldsymbol{y}$, as a function of $\boldsymbol{\Theta}$. When analyzing data using continuous distributions, a slight logical snag is therefore encountered: If a continuous model is employed, the probability associated with observing any particular exact realization $\boldsymbol{y}$ is always 0, for every $\boldsymbol{\Theta}$.

To understand how to employ likelihood methods in continuous models, it is then useful to consider the probability of observing a value of $\boldsymbol{Y}$ "near" $\boldsymbol{y}$ as a function of $\boldsymbol{\Theta}$. That is, suppose that

$$f_{\boldsymbol{\Theta}}(\boldsymbol{y})$$

is a joint probability density for $\boldsymbol{Y}$ depending on an unknown parameter vector $\boldsymbol{\Theta}$. Then in rough terms, if Δ is a small positive number and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$,

$$P[\text{each } Y_i \text{ is within } \tfrac{\Delta}{2} \text{ of } y_i] \approx f_{\boldsymbol{\Theta}}(\boldsymbol{y})\Delta^n \qquad \textbf{(A.26)}$$

But in expression (A.26), Δ^n doesn't depend on $\boldsymbol{\Theta}$—that is, the approximate probability is proportional to the function of $\boldsymbol{\Theta}$, $f_{\boldsymbol{\Theta}}(\boldsymbol{y})$. It is therefore plausible to use the joint density with data plugged in,

A continuous data likelihood function

$$\boxed{f_{\boldsymbol{\Theta}}(\boldsymbol{y})} \qquad \textbf{(A.27)}$$

as a likelihood function and to use its logarithm,

A continuous data log likelihood function

$$\boxed{L(\boldsymbol{\Theta}) = \ln(f_{\boldsymbol{\Theta}}(\boldsymbol{y}))} \qquad \textbf{(A.28)}$$

as a log likelihood for data modeled as jointly continuous. (Formulas (A.27) and (A.28) are formally identical to formulas (A.17) and (A.18), but they involve a different type of data.) Contemplation of formula (A.27) or (A.28) can be thought of as a way of assessing the consonance of different parameter vectors, $\boldsymbol{\Theta}$, with continuous data, $\boldsymbol{y}$. And as for the discrete case, a vector $\boldsymbol{\Theta}$ maximizing $L(\boldsymbol{\Theta})$ is often termed a *maximum likelihood estimate* of the parameter vector.

Example 21

Maximum Likelihood Estimation Based on iid Exponential Data

The exponential distribution is a popular model for life-length variables. The following are hypothetical life lengths (in hours) for $n = 4$ nominally identical electrical components, which will be assumed to have been a priori adequately described as iid exponential variables with mean α,

$$75.4, \quad 39.4, \quad 3.7, \quad 4.5 \qquad \textbf{(A.29)}$$

Example 21 (*continued*)

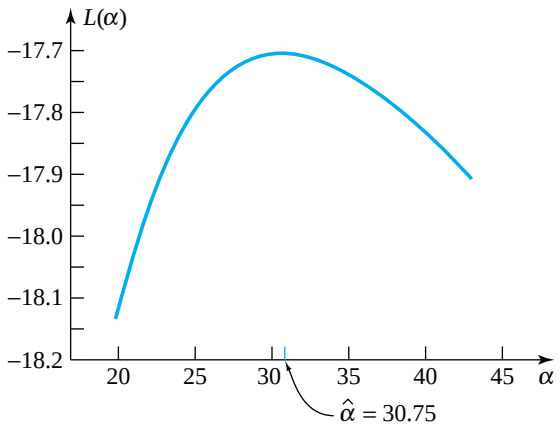

Figure A.17 Plot of a log likelihood based on four iid exponential observations

If Y_1, Y_2, Y_3, and Y_4 are iid exponential variables with means α, an appropriate joint probability density is

$$f(\boldsymbol{y}) = \begin{cases} \displaystyle\prod_{i=1}^{4} \frac{1}{\alpha} e^{-y_i/\alpha} & \text{for each } y_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

So with the data of display (A.29) in hand, the log likelihood function becomes

$$L(\alpha) = -4\ln(\alpha) - \frac{1}{\alpha}(75.4 + 39.4 + 3.7 + 4.5) \qquad \textbf{(A.30)}$$

It is easy to verify (using calculus and/or simply looking at the plot of $L(\alpha)$ in Figure A.17) that $L(\alpha)$ is maximized for

$$\hat{\alpha} = 30.75 = \frac{75.4 + 39.4 + 3.7 + 4.5}{4} = \bar{y}$$

This fact is a particular instance of the general result that the maximum likelihood estimate of an exponential mean is the sample average of the observations.

Maximum likelihood and normal observations

Example 21 is fairly simple, in that only one parameter is involved and calculus can be used to find an explicit formula for the maximum likelihood estimator. The reader might be interested in working through the somewhat more complicated (two-parameter) situation involving n iid normal random variables with means μ and standard deviations σ. Two-variable calculus can be used to show that maximum

likelihood estimates of the parameters based on observations $x_1, x_2, \ldots, x_n$ turn out to be, respectively,

$$\hat{\mu} = \bar{x}$$
$$\hat{\sigma} = \sqrt{\frac{n-1}{n}}\, s$$

The next example concerns an important continuous situation where no explicit formulas for maximum likelihood estimates seem to exist.

Example 22

Maximum Likelihood Estimation Based on iid Weibull Steel Specimen Failure Times

The data in Table A.4 are $n = 10$ ordered failure times for hardened steel specimens that were subjected to a particular rolling fatigue test. These data appeared originally in the paper of J. I. McCool, "Confidence Limits for Weibull Regression With Censored Data" (*IEEE Transactions on Reliability*, 1980). The Weibull probability plot of these data in Figure A.18 suggests the appropriateness of fitting a Weibull model to them (and indicates that β near 2 and α near .25 may be appropriate parameters for such a fitted model).

Notice that the joint density function of $n = 10$ iid Weibull random variables $Y_1, Y_2, \ldots, Y_{10}$ with parameters α and β is

$$f(\boldsymbol{y}) = \begin{cases} \displaystyle\prod_{i=1}^{10} \frac{\beta}{\alpha^{\beta}} y_i^{\beta-1} e^{-(y_i/\alpha)^{\beta}} & \text{for each } y_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

So using the data of Table A.4, the log likelihood

$$\begin{aligned} L(\alpha, \beta) &= 10\ln(\beta) - 10\beta\ln(\alpha) + (\beta - 1)(\ln(.073) + \ln(.098) + \cdots + \ln(.456)) \\ &\quad - \frac{1}{\alpha^{\beta}}((.073)^{\beta} + (.098)^{\beta} + \cdots + (.456)^{\beta}) \\ &= 10\ln(\beta) - 10\beta\ln(\alpha) - 16.267(\beta - 1) - \frac{1}{\alpha^{\beta}}((.073)^{\beta} + (.098)^{\beta} \\ &\quad + \cdots + (.456)^{\beta}) \end{aligned}$$

is indicated. Figure A.19 shows a contour plot of $L(\alpha, \beta)$ and indicates that maximum likelihood estimates of α and β are indeed in the vicinity of $\hat{\beta} = 2.0$ and $\hat{\alpha} = .26$.

Example 22
(*continued*)

Table A.4
Ten Ordered Failure Times of Steel Specimens

.073, .098, .117, .135, .175, .262, .270, .350, .386, .456

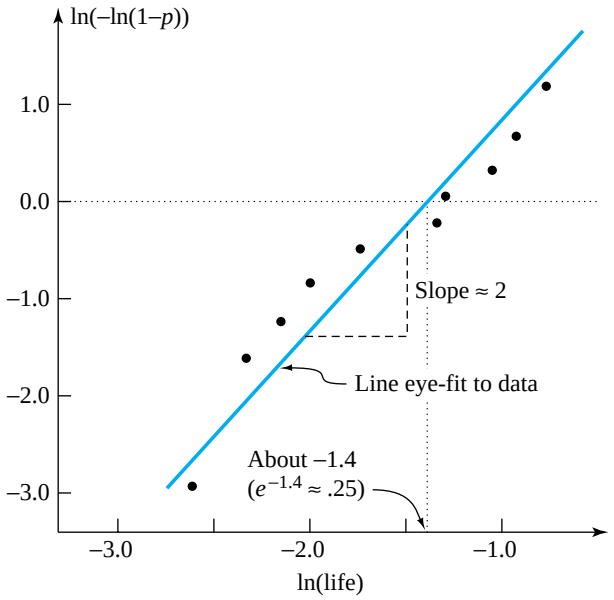

Figure A.18 Weibull probability plot of McCool's steel specimen failure times

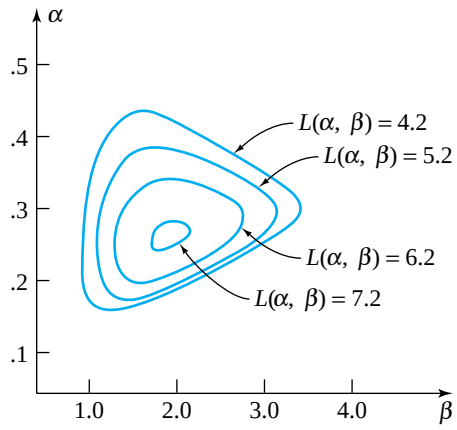

Figure A.19 Contour plot of a Weibull log likelihood for McCool's steel specimen failure times

Analytical attempts to locate the maximum likelihood estimates for this kind of iid Weibull data situation are only partially fruitful. Setting partial derivatives of $L(\alpha, \beta)$ equal to 0, followed by some algebra, does lead to the two equations

$$\beta = \left(\frac{\sum y_i^\beta \ln(y_i)}{\sum y_i^\beta} - \frac{\sum \ln(y_i)}{n} \right)^{-1}$$

$$\alpha = \left(\frac{\sum y_i^\beta}{n} \right)^{1/\beta}$$

which maximum likelihood estimates must satisfy, but these must be solved numerically.

Discrete and continuous likelihood methods have thus far been discussed separately. However, particularly in life-data analysis contexts, statistical engineering studies occasionally yield data that are *mixed*—in the sense that some parts are discrete, while other parts are continuous. If it is sensible to think of the two parts as independent, a combination of things already said here can lead to an appropriate likelihood function and then, for example, to maximum likelihood parameter estimates.

That is, suppose that one has available discrete data, $\boldsymbol{Y}_1 = \boldsymbol{y}_1$, and continuous data, $\boldsymbol{Y}_2 = \boldsymbol{y}_2$, which can be thought of as independently generated—$\boldsymbol{Y}_1$ from a discrete joint distribution with joint probability function

$$f_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{y}_1)$$

and $\boldsymbol{Y}_2$ from a continuous joint distribution with joint probability density

$$f_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{y}_2)$$

Then a sensible likelihood function becomes

A mixed-data likelihood function

$$f_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{y}_1) \cdot f_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{y}_2) \tag{A.31}$$

with corresponding log likelihood

A mixed-data log likelihood function

$$L(\boldsymbol{\Theta}) = \ln\left(f_{\boldsymbol{\Theta}}^{(1)}(\boldsymbol{y}_1) \right) + \ln\left(f_{\boldsymbol{\Theta}}^{(2)}(\boldsymbol{y}_2) \right) \tag{A.32}$$

Armed with equation (A.31) or (A.32), assessments of the compatibility of different parameter vectors $\boldsymbol{\Theta}$ with the data in hand and maximum likelihood model fitting can proceed just as for purely discrete or purely continuous cases.

Example 23

Maximum Likelihood Estimation of a Mean Insulating Fluid Breakdown Time Using Censored Data

Table 2.1 of Nelson's *Applied Life Data Analysis* gives some data on times to breakdown (in seconds) of an insulating fluid at several different voltages. The results of $n = 12$ tests made at 30 kV are repeated below in Table A.5. The last two entries in Table A.5 mean that two tests were terminated at (respectively) 29,200 seconds and 86,100 seconds without failures having been observed. In common statistical jargon, these last two data values are *censored* (at the times 29,200 and 86,100, respectively).

Nelson remarks in his book that exponential distributions are often used to model life length for such fluids. Therefore, consider fitting an exponential distribution with mean α to the data of Table A.5. Notice that the first ten pieces of data in Table A.5 are continuous "exact" failure times, while the last two are essentially discrete pieces of information. Considering first the discrete part of the overall likelihood, the probability that two independent exponential variables exceed 29,200 and 86,100, respectively, is

$$f_\alpha^{(1)}(\boldsymbol{y}_1) = e^{-29{,}200/\alpha} \cdot e^{-86{,}100/\alpha}$$

Then considering the continuous part of the likelihood, the joint density of ten independent exponential variables with mean α is

$$f_\alpha^{(2)}(\boldsymbol{y}_2) = \begin{cases} \dfrac{1}{\alpha^{10}} e^{-\sum y_i/\alpha} & \text{for each } y_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

Putting these two pieces together via equation (A.32), the log likelihood function appropriate here is

$$\begin{aligned} L(\alpha) &= -10\ln(\alpha) - \frac{1}{\alpha}(50 + 134 + 187 + \cdots + \\ &\quad + 15{,}800 + 29{,}200 + 86{,}100) \\ &= -10\ln\alpha - \frac{1}{\alpha}(144{,}673) \end{aligned} \tag{A.33}$$

Table A.5
12 Insulating Fluid Breakdown Times

50, 134, 187, 882, 1450, 1470, 2290, 2930, 4180, 15800, > 29200, > 86100

This function of α is easily seen via elementary calculus to be maximized at

$$\hat{\alpha} = \frac{144{,}673}{10} = 14{,}467.3 \text{ sec}$$

which has the intuitively appealing interpretation of the ratio of the total time on test to the number of failures observed during testing.

A.5.3 Likelihood-Based Large-Sample Inference Methods

One of the appealing things about the likelihood function idea is that in many situations, it is possible to base large-sample significance testing and confidence region methods on the likelihood function. Intuitively, it would seem that those parameter vectors $\boldsymbol{\Theta}$ "most compatible" with the data in hand ought to form a sensible confidence set for $\boldsymbol{\Theta}$. And in significance-testing terms, if a hypothesized value of $\boldsymbol{\Theta}$ (say, $\boldsymbol{\Theta}_0$) has a corresponding value of the likelihood function far smaller than the maximum possible, that circumstance ought to produce a small p-value—that is, strong evidence against $\mathrm{H}_0\colon \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$.

To make this thinking precise, let

The maximum of the log-likelihood function

$$L^* = \max_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta})$$

that is, L^* is the largest possible value of the log likelihood. (If $\hat{\boldsymbol{\Theta}}$ is a maximum likelihood estimate of $\boldsymbol{\Theta}$, then $L^* = L(\hat{\boldsymbol{\Theta}})$.) An intuitively appealing way to make a confidence set for the parameter vector $\boldsymbol{\Theta}$ is to use the set of all $\boldsymbol{\Theta}$'s with log likelihood not too far below L^*,

A likelihood-based confidence set for $\boldsymbol{\Theta}$

$$\left\{\boldsymbol{\Theta} \mid L(\boldsymbol{\Theta}) > L^* - c\right\} \tag{A.34}$$

for an appropriate number c. And a plausible way of deriving a p-value for testing

$$\mathrm{H}_0\colon \boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \tag{A.35}$$

is by trying to identify a sensible probability distribution for

$$L^* - L(\boldsymbol{\Theta}_0) \tag{A.36}$$

when H_0 holds, and using the upper-tail probability beyond an observed value of variable (A.36) as a p-value.

The practical gaps in this thinking are two: how to choose c in display (A.34) to get a desired confidence level and what kind of distribution to use to describe variable (A.36) under hypothesis (A.35). There are no general exact answers to these

questions, but statistical theory does provide at least some indication of approximate answers that are often adequate for practical purposes when large samples are involved. That is, statistical theory suggests that in many large-sample situations, if $\boldsymbol{\Theta}$ is of dimension k, choosing

Constant producing (large sample) approximate γ level confidence for $\{\boldsymbol{\Theta} \mid L(\boldsymbol{\Theta}) > L^* - c\}$

$$c = \tfrac{1}{2}U \tag{A.37}$$

for U the γ quantile of the χ^2_k distribution, produces a confidence set (A.34) of confidence level roughly γ. And similar reasoning suggests that in many large-sample situations, if $\boldsymbol{\Theta}$ is of dimension k, the hypothesis (A.35) can be tested using the test statistic

A test statistic for $H_0: \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ with an approximately χ^2_k reference distribution

$$2\left(L^* - L(\boldsymbol{\Theta}_0)\right) \tag{A.38}$$

and a χ^2_k approximate reference distribution, where large values of the test statistic (A.38) count as evidence against H_0.

Example 23 (*continued*)

Consider the problem of setting confidence limits on the mean time till breakdown of Nelson's insulating fluid tested at 30 kV. In this problem, $\boldsymbol{\Theta}$ is $k = 1$-dimensional. So, for example, making use of the facts that the .9 quantile of the χ^2_1 distribution is 2.706 and that the maximum likelihood estimate of α is 14,467.3, displays (A.33), (A.34), and (A.37) suggest that those α with

$$L(\alpha) > -10\ln(14{,}467.3) - \frac{1}{14{,}467.3}(144{,}673) - \frac{1}{2}(2.706)$$

that is,

$$-10\ln(\alpha) - \frac{1}{\alpha}(144{,}673) > -107.15$$

form an approximate 90% confidence set for α. Figure A.20 shows a plot of the log likelihood (A.33) cut at the level -107.15 and the corresponding interval of α's. Numerical solution of the equation

$$-10\ln(\alpha) - \frac{1}{\alpha}(144{,}673) = -107.15$$

shows the interval for mean time till breakdown to extend from 8,963 sec to 25,572 sec.)

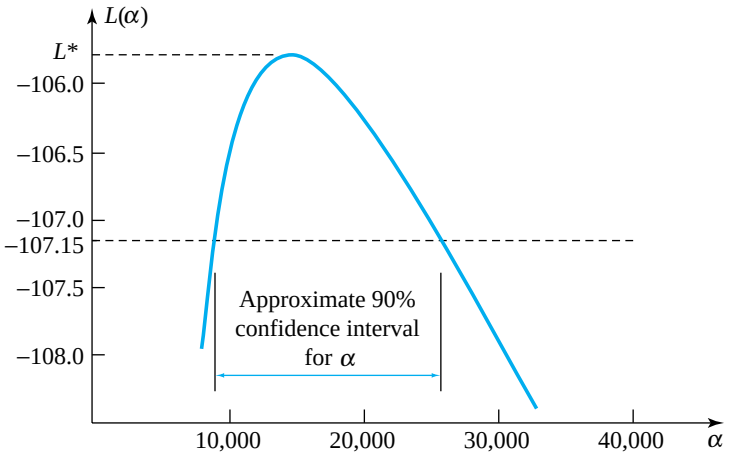

Figure A.20 Plot of the log likelihood for Nelson's insulating fluid breakdown time data and approximate confidence limits for α

The $n = 12$ pieces of data in Table A.5 do not constitute an especially large sample, so the 90% approximate confidence level associated with the interval (8,963, 25,572) should be treated as very approximate. But even so, this interval does give one some feeling about the precision with which α is known based on the data of Table A.5. There is clearly substantial uncertainty associated with the estimate $\hat{\alpha} = 14{,}467.3$.

Cautions concerning the large-sample likelihood-based inference methods

It is not a trivial matter to verify that the χ^2_k approximations suggested here are adequate for a particular nonstandard probability model. In engineering situations where fairly exact confidence levels and/or p-values are critical, readers should seek genuinely expert statistical advice before placing too much faith in the χ^2_k approximations. But for purposes of engineering problem solving requiring a rough, working quantification of uncertainty associated with parameter estimates, the use of the χ^2_k approximation is certainly preferable to operating without any such quantification.

The insulating fluid example involved only a single parameter. As an example of a $k = 2$-parameter application, consider once again the space shuttle O-ring failure example.

Example 19 (*continued*)

Again use the log likelihood (A.23) and the fact that maximum likelihood estimates of α and β in equation (A.21) or (A.22) are $\hat{\alpha} = 15.043$ and $\hat{\beta} = -.2322$. These produce corresponding log likelihood -10.158. This, together with the

Example 19
(*continued*)

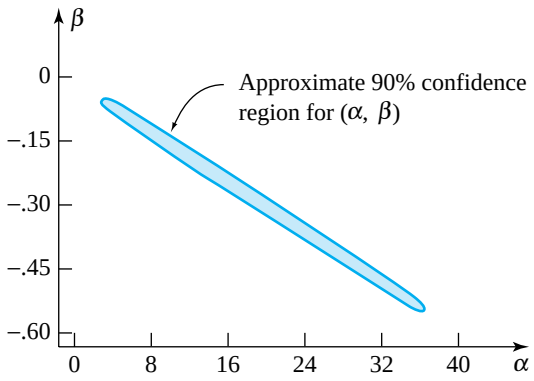

Figure A.21 Likelihood-based approximate confidence region for the parameters of the O-ring failure model

fact that the .9 quantile of the χ^2_2 distribution is 4.605, gives one (from displays (A.34) and (A.37)), that the set of (α, β) pairs with

$$L(\alpha, \beta) > -10.158 - \frac{1}{2}(4.605)$$

that is,

$$L(\alpha, \beta) > -12.4605$$

constitutes an approximate 90% confidence region for (α, β). This set of possible parameter vectors is shown in the plot in Figure A.21. Notice that one message conveyed by the contour plot is that β is pretty clearly negative. Low-temperature launches are more prone to O-ring failure than moderate- to high-temperature launches.

The approximate inference methods represented in displays (A.34) through (A.38) concern the entire parameter vector $\boldsymbol{\Theta}$ in cases where it is multidimensional. It is reasonably common, however, to desire inferences only for particular parameters individually. (For example, in the case of the O-rings, it is the parameter β that determines whether $p(t)$ is increasing, constant, or decreasing in t, and for many purposes β is of primary interest.) It is thus worth mentioning that the likelihood ideas discussed here can be adapted to provide inference methods for a *part* of a parameter vector $\boldsymbol{\Theta}$ of individual interest. An exposition of these adaptations will not be attempted here, but be aware of their existence. For details, refer to more complete expositions of likelihood methods (such as that in Meeker and Escobar's *Statistical Methods for Reliability Data* text).