SECTION 48
RELIABILITY CONCEPTS AND DATA ANALYSIS

William Q. Meeker
Luis A. Escobar
Necip Doganaksoy
Gerald J. Hahn

INTRODUCTION

Relationship between Quality and Reliability. Rapid advances in technology, development of highly sophisticated products, intense global competition, and increasing customer expectations have put new pressures on manufacturers to produce high-quality products. While producers often think of high quality in terms of minimizing scrap and rework, customers are concerned with functionality, reliability, and overall product lifetime. Customers expect a purchased product to meet or exceed life expectations and to be safe. Technically, reliability is often defined as the probability that a system, vehicle, machine, device, or other product will perform its intended function under some specified operating conditions, for a specified period of time. Improving reliability is an important part of the larger overall goal of improving product quality. There are many definitions of quality, but there is general agreement that an unreliable product is *not* a high-quality product. Condra (1993) emphasizes that "reliability is quality over time."

Modern programs for improving reliability of existing products and for assuring continued high reliability for the next generation of products require quantitative methods for predicting and assessing product reliability and for providing early signals and information on root causes of failure. In many cases this will involve the collection and analysis of reliability data from studies such as laboratory tests (or designed experiments) of materials, devices, and components, tests on early prototype units, careful monitoring of early-production units in the field, analysis of warranty data, and systematic longer-term tracking of product in the field. This frequently also involves the careful planning of such programs so as to ensure that the most meaningful information is obtained. Reliability evaluations often present a challenge beyond that normally encountered in quality evaluations because there is usually an elapsed time between when the product is built and when the reliability information is forthcoming.

Applications of Reliability Data Analysis. Some major goals in obtaining reliability data include:

1. Obtaining early identification of failure modes and understanding and removing their root causes—and thereby improving reliability.
2. Obtaining field failure information to help improve the product design, perhaps for the next product release.
3. Determining how long each unit should be run prior to shipment (and at what conditions) in order to avoid premature field failures. This is sometimes referred to as "burn-in."
4. Quantifying reliability to determine whether or not a product is ready for release; i.e., does the product meet the specified level of reliability?
5. Predicting warranty costs.
6. Deciding on the need for recall of a product in the field (and, if so, the extent of the recall).

Although the last four objectives are the ones that analysts are asked to respond to most often, concentrating on them is sometimes premature. The first objective is the most fundamental since one cannot demonstrate a high level of reliability if, in fact, it does not exist, and clearly it is desirable to avoid expensive burn-in and field recalls. However, much of the discussion in this section addresses the last four goals. Section 3 (The Quality Planning Process) and Section 19 (Quality in Research and Development) consider the entire process of assuring high reliability.

Sources and Types of Reliability Data. It is important to distinguish between the following types of reliability data:

- A sequence of reported system repair times (or the times of other system-specific events) for a repairable system.
- The time of failure (or other clearly specified event) for nonrepairable units or components (including *nonrepairable* components within a *repairable* system).

When reliability tests are conducted on larger systems and subsystems (even those that may be repaired), it is essential that detailed component-level information on cause of failure be obtained. This is especially important if the purpose of the data collection is improvement of system reliability, as opposed to mere assessment of overall system reliability. Subsequent subsections describe methods for data analysis for nonrepairable units or components as well as for analyzing system repair data.

Reliability data arise from many different kinds of situations. Examples include:

- Laboratory tests to study durability, wear, or other lifetime properties of particular materials, components, or subsystems.
- Operational life tests on complete systems or subsystems, conducted before a product is released to customers.
- Field operation by customers.

Reliability data are typically censored, i.e., some units remain unfailed, and (at best) only their survival times, but not their failure times, are known. The most common reason for censoring is the need to analyze data before all units fail. The analysis of censored data is more complicated when the censoring times of unfailed units differ. This would happen when different units of the product are placed on test or enter into the field at different times, as is usually the case in analyzing field failure data. It may also be the case when units have different degrees of exposure over time or when one is evaluating failures due to a particular failure mode (in which case failures from other independent modes are treated as censored observations in the data analysis). An important assumption needed for standard analysis of censored data is that the censoring time for a unit is independent of when that unit would have failed. For example, if a unit were removed from the field because it is about to fail, treating it as a censored observation would bias the analysis.

Computer Software. In this section we focus on the methods and applications of reliability data analysis. We provide formulas only when they provide insight and will help readers to understand how to use the method.

Although it is possible to do some of the simplest reliability data analyses by hand, for maximally effective analyses, computer processing along with display using modern high-resolution graphics should be used.

Unfortunately, only a few of the best-known standard data analysis computer packages have adequate capabilities for doing reliability data analysis. As software vendors become more aware of their customers' needs, capabilities in commercial packages can be expected to improve.

SAS PROC RELIABILITY (SAS Institute 1996), SAS JMP (SAS Institute 1995), S-Plus (Statistical Sciences 1996), Minitab 12 (Minitab 1997), and a specialized program called WinSMITH (Abernethy 1996), can do product life data analysis for the simple (single-distribution) situation for which all life data are assumed to be in a single common environment. Ease of use and the character of the user interface vary widely over these packages. In addition, each of these programs, except WinSMITH, can do regression and accelerated life test analyses. SAS JMP can also analyze data with more than one failure mode. SAS PROC RELIABILITY can, in addition, do the nonparametric repairable systems analyses described in this section. Meeker and Escobar (1998) describe S-Plus functions that can be used to do all of the analyses described in this section.

LIFE DATA MODELS

This section deals mainly with nonrepairable components or other products that are replaced rather than repaired upon failure (or time to first failure on repairable products). The following are described:

1. Statistical models for representing time to failure of nonrepairable products.
2. Methods for estimating quantities of interest (e.g., failure probabilities, distribution quantiles) from time-to-failure data, with and without making any distributional assumptions.

We discuss initially models that deal with time to failure in a single-use environment. In our subsequent discussion of accelerated life tests we will describe corresponding regression models that explain the effect that factors like temperature have on life.

Time-to-Failure Model Concepts. The distribution of time to failure T of a product can be characterized by a cumulative distribution function (cdf), a probability density function (pdf), a survival function (sf), or a hazard function (hf). The cdf and pdf are common ways of describing a statistical model in many applications (and are treated in elementary statistics courses). The hf, on the other hand, is not as well known, but it has particular applicability to product life data analysis. These functions are described below and illustrated (for one possible time-to-failure distribution) in Figure 48.1. The choice of which function or functions to use depends on the specific practical problem that is being addressed. All of these functions are important for one purpose or another.

The cdf for T, $F(t) = \Pr(T \le t)$, gives the *probability* that a unit, selected at random from the population or process, will fail before time t. Alternatively, $F(t)$ can be interpreted as the *proportion* of units in the population (or from some stationary process) that will fail before time t. (Here a stationary process is defined as one that generates units that have a $F(t)$ that does not change over time.)

The sf (also known as the *reliability function*) is the complement of the cdf, $S(t) = \Pr(T > t) = 1 - F(t) = \int_t^\infty f(x)\,dx$, and gives the probability of surviving until time t.

The pdf for a continuous random variable T is defined as the derivative of $F(t)$ with respect to t: $f(t) = dF(t)/dt$. Thus, for a positive random variable, $F(t) = \int_0^t f(x)\,dx$. The pdf can be used to represent relative frequency of failure times as a function of time and can be thought of as a smoothed histogram of a large number of observed failure times.

The hf (also known as the hazard rate, the instantaneous failure rate function, and by various other names) is defined by:

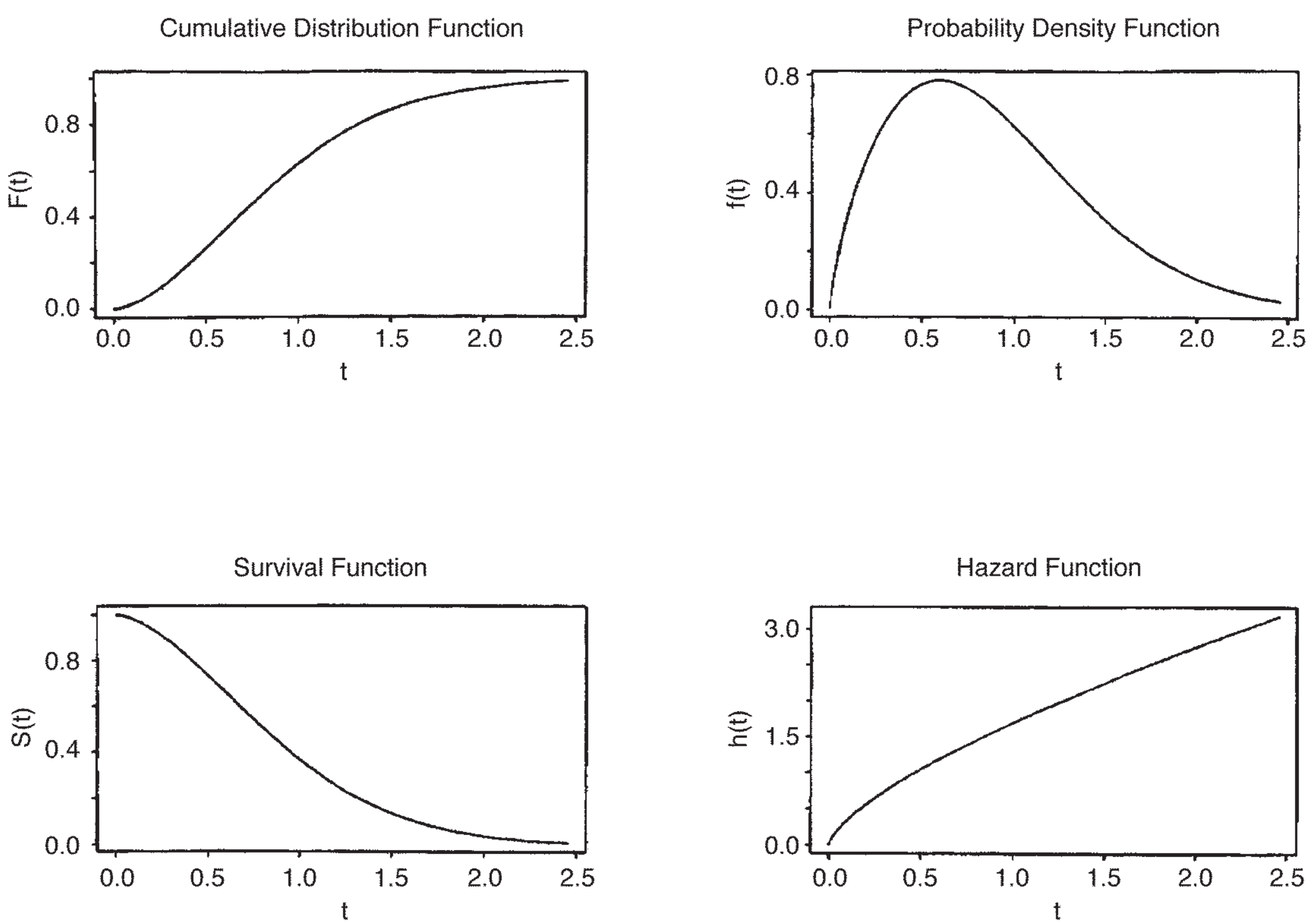

FIGURE 48.1 Typical time-to-failure cdf, pdf, sf, and hf.

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{1 - F(t)}$$

The hf expresses the propensity to fail in the next small interval of time, given survival to time t. That is, for small Δt, $h(t) \times \Delta t \simeq \Pr(t < T \leq t + \Delta t | T > t)$. The hf can be interpreted as a failure rate in the following sense. If there is a large number of items [say $n(t)$] in operation at time t then $n(t) \times h(t)$ is approximately equal to the number of failures per unit time [or $h(t)$ is approximately equal to the number of failures per unit time per unit at risk]. Because of its close relationship with failure processes and maintenance strategies, reliability engineers often model time to failure in terms of $h(t)$. The "bathtub curve" shown in Figure 48.2 provides a useful conceptual model for the hazard of some product populations. There may be early failures of units with quality-related defects (infant mortality). During much of the useful life of a product, the hazard may be approximately constant because failures are caused by external shocks that occur at random. Late-life failures are due to wear-out. Many reliability tests focus on one side or the other of this curve.

Other Quantities of Interest in Reliability Analysis. The mean (also known as the "expectation" or "first moment") of a positive random variable T is a measure of central tendency or "average" time, and it is defined by:

$$E(T) = \int_0^\infty t f(t)\, dt = \int_0^\infty [1 - F(t)]\, dt$$

When the probability density function $f(t)$ is highly skewed, e.g., has a long right tail (as is common in many life data applications), the mean time to failure may differ appreciably from other measures of central tendency like the median time to failure (the time by which exactly half of the population of units will fail). The mean time to failure $E(T)$ is sometimes abbreviated to MTTF.

The traditional parameters of a statistical model (mean and standard deviation) are often *not* of primary interest in reliability studies. Instead, design engineers, reliability engineers, managers, and customers are interested in specific measures of product reliability or particular characteristics of the failure-time distribution, e.g., distribution quantiles. In particular, the quantile t_p is the time at which a specified proportion p of the population will fail. Also, $F(t_p) = p$. For example, $t_{.20}$ is the *time* by which 20 percent of the population will fail. Alternatively, frequently one would like to know $F(t)$, the *probability* of failure associated with a particular number of hours, days, weeks, months, or years of usage, e.g., the probability of a product failing (or not failing) during the first 5 years in the field.

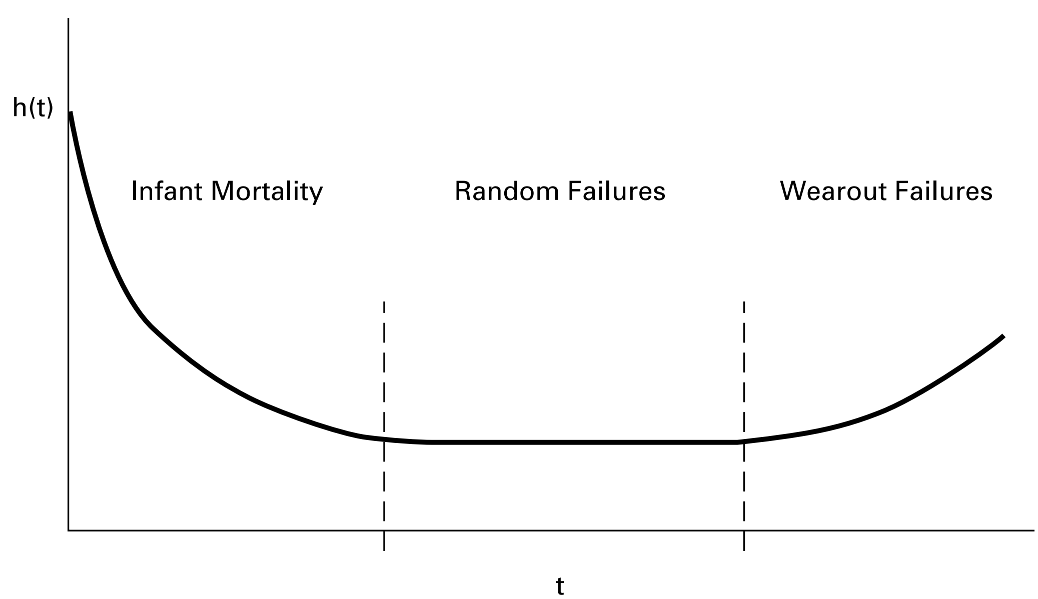

FIGURE 48.2 Bathtub curve hazard function.

Common Parametric Time-to-Failure Distributions

The Exponential Distribution. The exponential distribution cdf is

$$F(t;\, \theta) = 1 - \exp\left(-\frac{t}{\theta}\right) \qquad t > 0$$

where the distribution's single parameter $\theta > 0$ is a scale parameter. The exponential distribution has no shape parameter and, thus, has only a single unique shape. The exponential cdf, pdf, and hazard function are graphed in Figure 48.3 for $\theta = .5$, 1, and 2. The p quantile of the exponential distribution is $t_p = -\theta \log (1 - p)$. The exponential distribution has the important characteristic that its hazard function $h(t) = 1/\theta$ is constant (does not depend on time t). A constant hazard implies that, for an unfailed unit, the probability of failing in the next interval of time is independent of the unit's age. Physically, a constant hazard implies that units in the population are not subject to an infant mortality failure mode and also are not wearing out or otherwise aging. Thus, this model may be appropriate when failures are induced by an external phenomenon that is independent of product life and past history. An example might be the life in service of a dish in a restaurant, whose end of life is the result of breakage. The exponential distribution is commonly, and sometimes incorrectly, used because of its simplicity. It would *not* be appropriate for modeling the life of mechanical components (e.g., bearings) subject to some combination of fatigue, corrosion, or wear or for electronic components that exhibit wear-out properties over their life (e.g., lasers and filament devices).

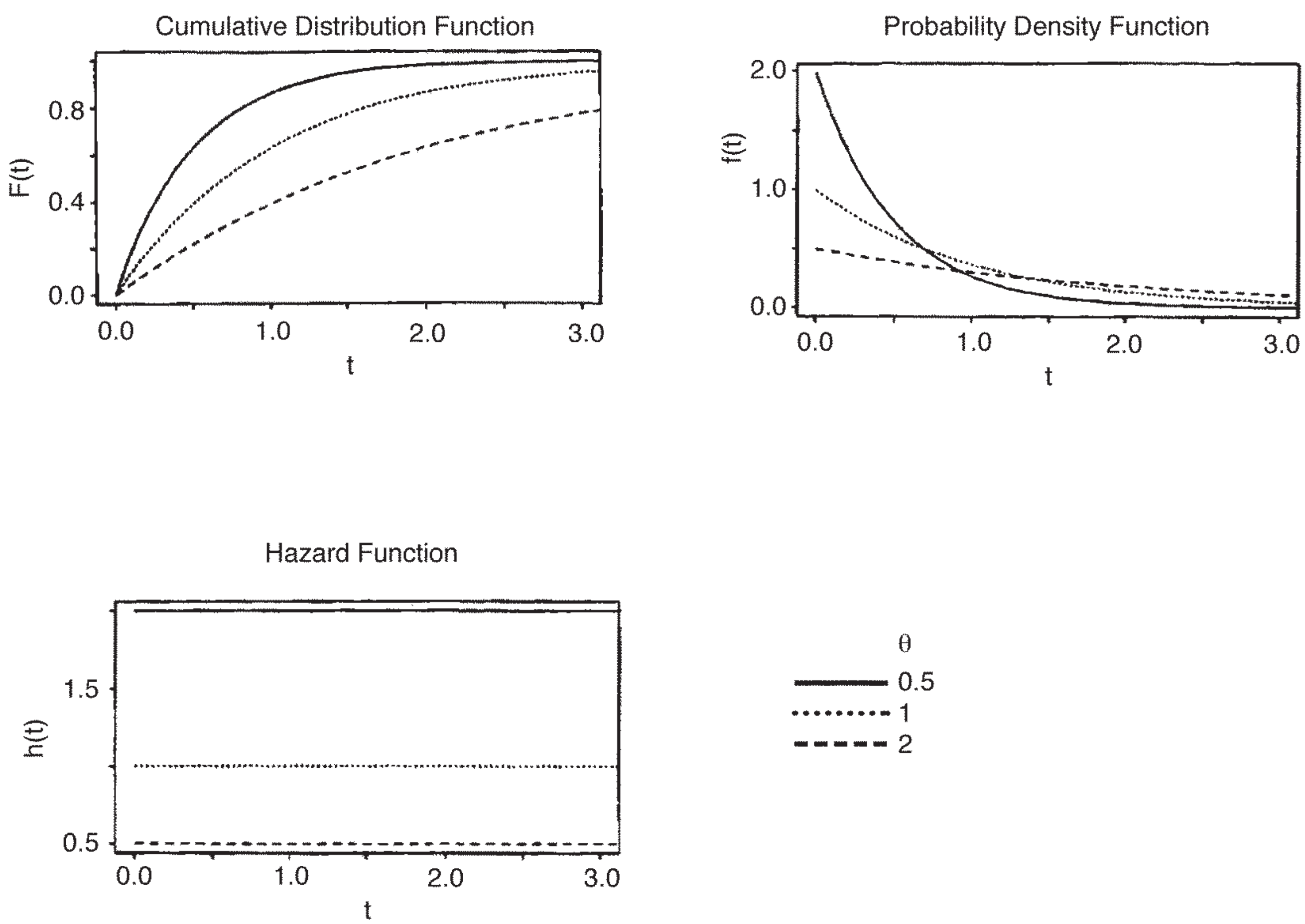

FIGURE 48.3 Exponential distribution cdf, pdf, and hf for $\theta = .5$, 1, and 2.

The Weibull Distribution. The Weibull distribution cdf is

$$F(t;\,\eta,\,\beta) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \qquad t > 0$$

where $\beta > 0$ is a shape parameter and $\eta > 0$ is a scale parameter (which is also approximately the .63 quantile of this distribution). The p quantile of the Weibull distribution is $t_p = \eta[-\log(1-p)]^{1/\beta}$. For example, the Weibull distribution median is $t_{.50} = \eta[-\log(1-.5)]^{1/\beta}$. The Weibull hazard function is $h(t) = (\beta/\eta)(t/\eta)^{\beta-1}$. The exponential distribution is a special case of the Weibull with $\beta = 1$. The practical importance of the Weibull distribution stems from its ability to describe failure distributions with many different commonly occurring shapes. As illustrated in Figure 48.4 with $0 < \beta < 1$, the Weibull hazard function is decreasing and with $\beta > 1$, the Weibull hazard function is increasing. Thus the Weibull distribution can be used to describe either decreasing hazard caused by infant mortality or increasing hazard due to wear-out, but *not* both. The Weibull distribution is often used as a model for the life of insulation and many other products as well as a model for strength of materials. For some applications there is theoretical justification for its use.

The logarithm of a Weibull random variable, $Y = \log(T)$, follows a smallest extreme value distribution so that

$$\Pr[\log(T) \le \log(t)] = F(t;\,\mu,\,\sigma) = \Phi_{\text{sev}}\left(\frac{\log(t)-\mu}{\sigma}\right)$$

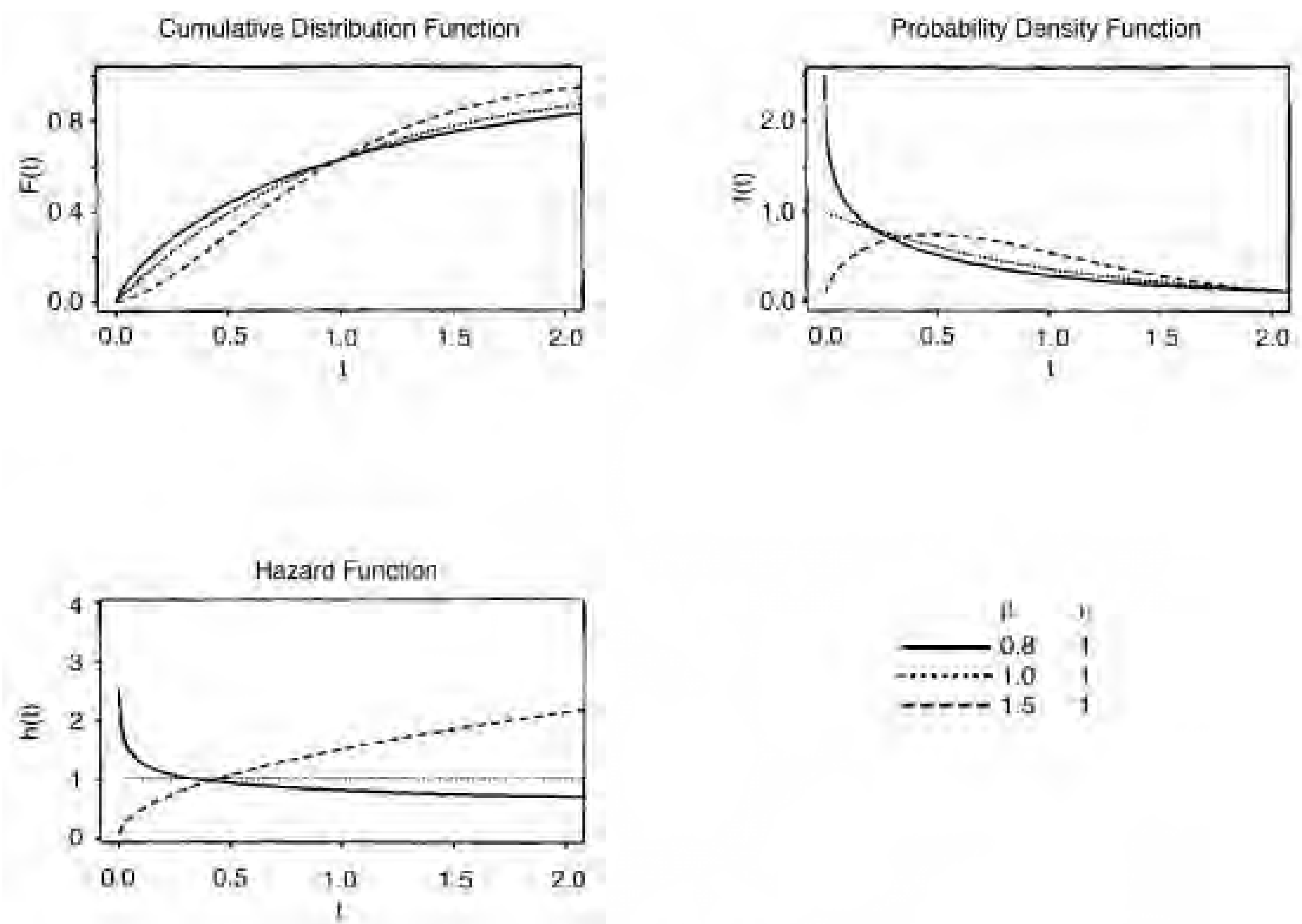

FIGURE 48.4 Weibull distribution cdf, pdf, and hf for scale parameter $\eta=1$ and shape parameter $\beta=.8$, 1.0, and 1.5.

where $y = \log(t)$, $\Phi_{sev}(z) = 1 - \exp[-\exp(z)]$, $\mu = \log(\eta)$ is a location parameter with $-\infty < \mu < \infty$, and $\sigma = 1/\beta > 0$ is a scale parameter. The relationship between the Weibull and the smallest extreme value distribution is similar to the relationship between the normal and lognormal distributions and the relationship will be used later in this section in explaining methods for failure-time regression analysis and accelerated test modeling.

The Lognormal Distribution. The normal distribution, though playing a central role in other statistical applications, is infrequently appropriate as a time-to-failure model. This is because failure can occur at times shortly after 0 hours, but they cannot be negative. Logarithms of failure times, however, are often described well by a normal distribution. This is equivalent to fitting a lognormal distribution. The lognormal cdf is

$$F(t; \mu,\sigma) = \Phi_{nor}\left[\frac{\log(t) - \mu}{\sigma}\right] \qquad t > 0$$

where Φ_{nor} is the cdf for the standard normal distribution (see Section 44). The parameter $\exp(\mu)$ is a scale parameter and the median of the distribution, $\sigma > 0$ is a shape parameter. The p quantile of the lognormal distribution is $t_p = \exp[\mu + \Phi^{-1}_{nor}(p)\,\sigma]$, where Φ^{-1}_{nor} is the inverse of the standard normal distribution. The relationship between the lognormal and normal distributions is often used to simplify the use of the lognormal distribution.

The lognormal distribution is commonly used as a model for the distribution of failure times. As shown in Figure 48.5, the lognormal hazard function always starts at 0, increases to a maximum, and then approaches 0 for large t.

Following from the central limit theorem (described in Section 44), application of the lognormal distribution could be justified for a random variable that arises from the product of a number of identically distributed independent positive random effects. For example, it has been suggested that the lognormal is an appropriate model for time to failure caused by a degradation process with combinations of random rate constants that combine multiplicatively. Correspondingly, the lognormal distribution is widely used to describe failure times due to fatigue and in microelectronic applications.

Other Life Distributions. For more details for these and other distributions that can be used to describe life distributions, see Chapter 8 of Hahn and Shapiro (1967) and Chapters 4 and 5 of Meeker and Escobar (1998).

Multiple Failure Modes. In more complex situations (e.g., products that have *both* infant mortality and wear-out modes, as in the bathtub curve situation, which was described earlier) the data cannot be fit by any of the preceding models. This is often the case when there is more than one failure mode. In such cases, if the individual failure modes can be identified, it is usually better to do the evaluations separately for each of these modes and then combine the results to get total system probabilities. See Chapter 5 of Nelson (1982) or Chapter 15 of Meeker and Escobar (1998) for examples.

ANALYSIS OF CENSORED LIFE DATA

This subsection illustrates simple, but powerful, methods for analyzing life data. The first example uses data from a laboratory life test. The second example illustrates the analysis of multiply censored data from a field tracking study. In both cases we proceed as follows:

1. Analyze the data graphically, using a nonparametric procedure (one that does not require any assumptions about the form of the underlying distribution) plotted on special probability paper. Such plots also allow us to explore the adequacy of possible distributional models.
2. If a parametric distribution provides an adequate description of the available data, estimate the distribution parameters.

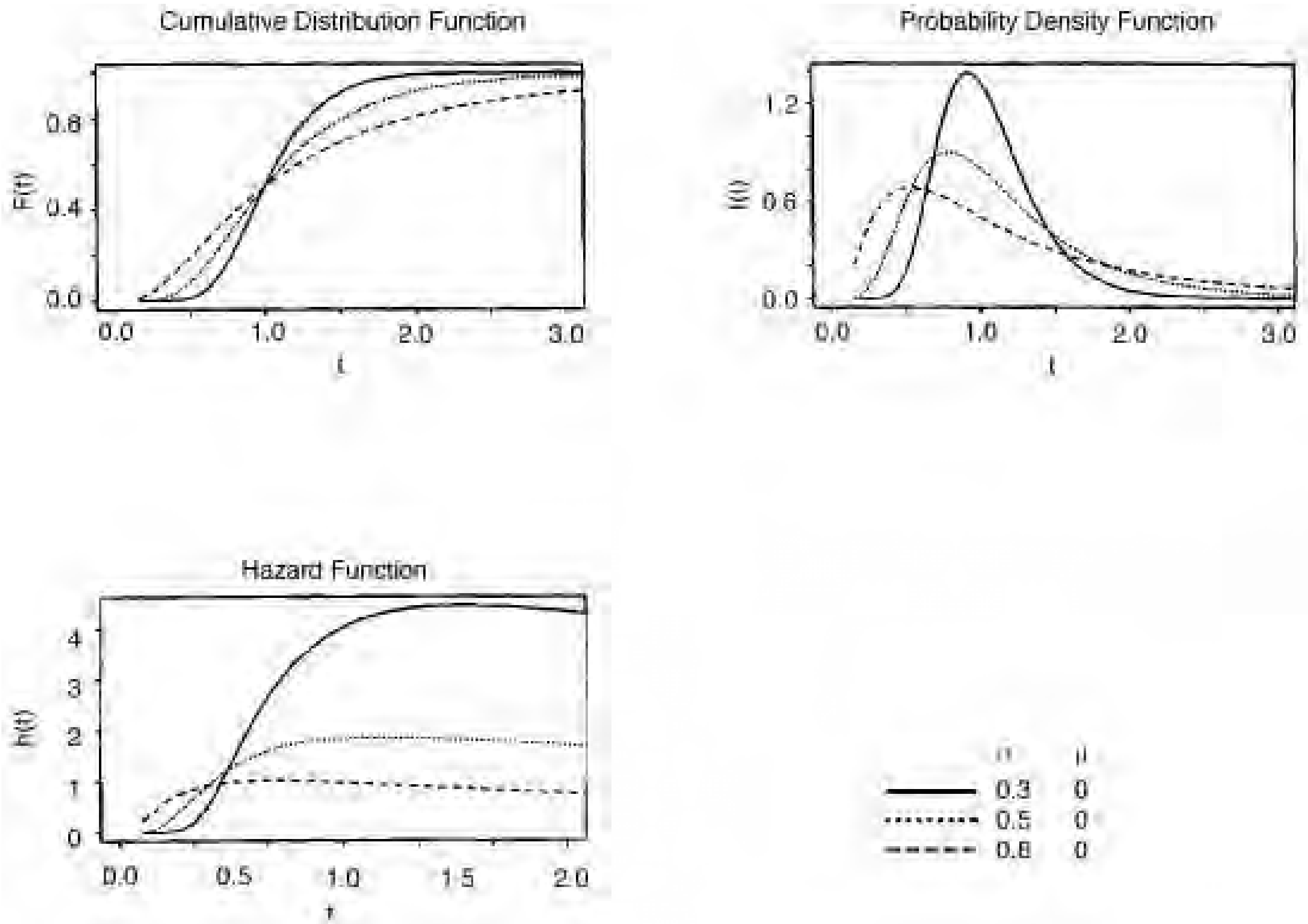

FIGURE 48.5 Lognormal distribution cdf, pdf, and hf for scale parameter $\exp(\mu) = 1$ and for shape parameter $\sigma = .3$, .5, and .8.

3. Using these parameter estimates, estimate the distribution quantiles of interest, population proportion failing at specified times, hazard function values, and appropriate confidence intervals on some or all of these estimates.

Simple Life Test Data (with Single Censoring). For sample life test data with n observations, continuous or very frequent monitoring for failures, and all censoring at the end of the observation period (typical for some laboratory life tests), a nonparametric estimate of $F(t)$ can be computed as

$$\hat{F}(t) = \frac{\text{number of failures up to time } t}{n} \tag{48.1}$$

This is a step function that jumps by an amount $1/n$ at each failure time (unless there are ties, in which case the estimate jumps by the number of tied failures divided by n). This estimate requires no assumption about the form of the underlying distribution. The estimate can be computed only up to the censoring time, and not beyond.

Example: Chain Link Fatigue Life. Parida (1991) gives the results of a load-controlled high-cycle fatigue test conducted on 130 chain links. The 130 links were randomly selected from a population of different heats used to manufacture the links. Each link was tested until failure or until it had run for 80 thousand cycles, which ever came first. There were 10 failures—one each at 33, 46, 50, 59, 62, 71, 74, and 75 thousand cycles and two at 78 thousand cycles. The remaining 120 links had not failed by 80 thousand cycles. Figure 48.6 shows the step-function nonparametric estimate of $F(t)$ for these data.

Probability Plotting. Comparing data in plots like Figure 48.6 with theoretical cumulative distributions like those shown in Figures 48.3, 48.4, and 48.5 is difficult, because the human eye cannot easily compare nonlinear curves. Probability plots display data on special *probability scales* such that a particular theoretical cumulative distribution is a straight line when plotted on such a scale. For example, taking the log of the Weibull quantile gives the straight line $\log(t_p) = \log(\eta) + (1/\beta)\log[-\log(1-p)]$. Thus, appropriately plotting the data on probability paper, and assessing whether the plot reasonably approximates a straight line, provides a simple assessment of the adequacy of an assumed model. In addition, probability plots are commonly used to estimate

- Distribution quantiles
- Probabilities of surviving to specified lifetimes

Figure 48.7 is a Weibull probability plot of the chain link failure data. One common method of making a probability plot is to plot the ith largest time (or number of cycles at failure) value on the time axis versus $(i - .5)/n$ on the probability axis for $i = 1,2,\ldots$. This is equivalent to plotting, at each jump in the step function of Figure 48.6, the time of the step versus the point halfway between steps, as shown in Figure 48.7. Drawing a curve through these points allows one to estimate distribution quantiles and survival probabilities, without assuming any distributional model. However, the points plotted on Figure 48.7 fall nearly along a straight line, indicating that the Weibull distribution provides a good fit to the data. Figure 48.7 contains a special "β" scale that allows one to obtain a graphical estimate of the Weibull shape parameter. This is done by drawing a line through the cross mark on the top of the plot, parallel to the line that seems to best fit the data points. Then the graphical estimate of β is read off the "β" scale. In this case, the graphical estimate of the Weibull shape para-

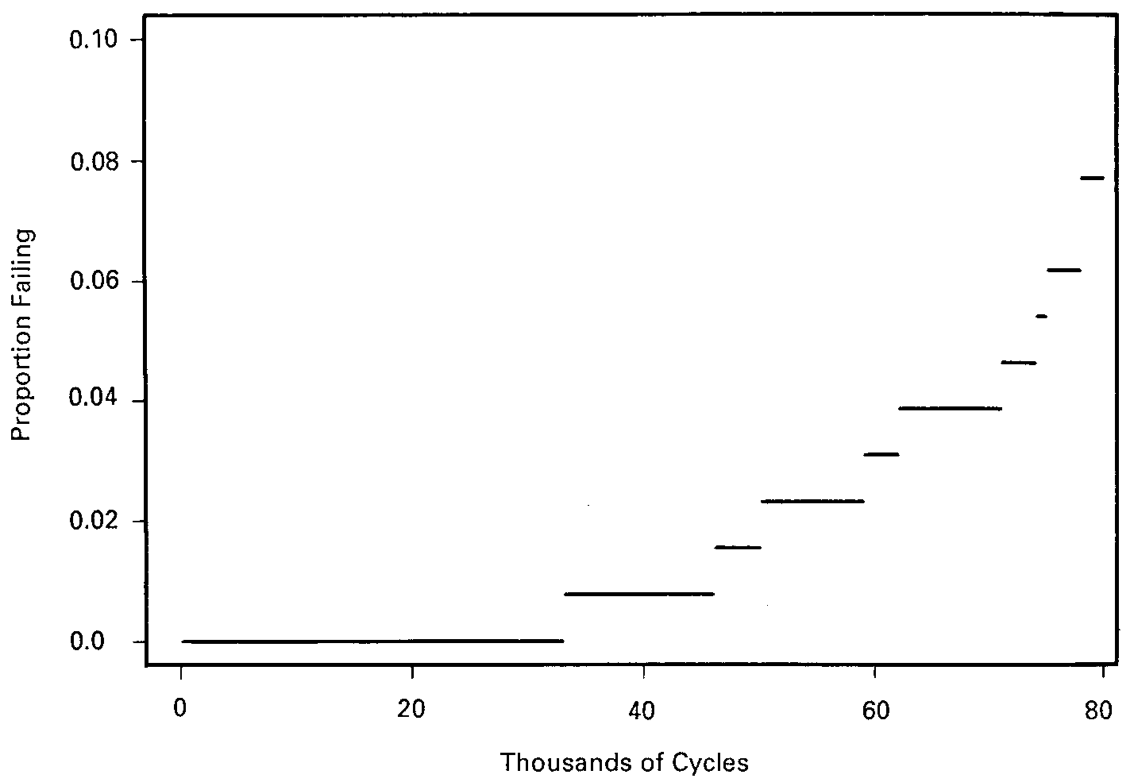

FIGURE 48.6 Nonparametric estimate of $F(t)$ for the chain link failure data.

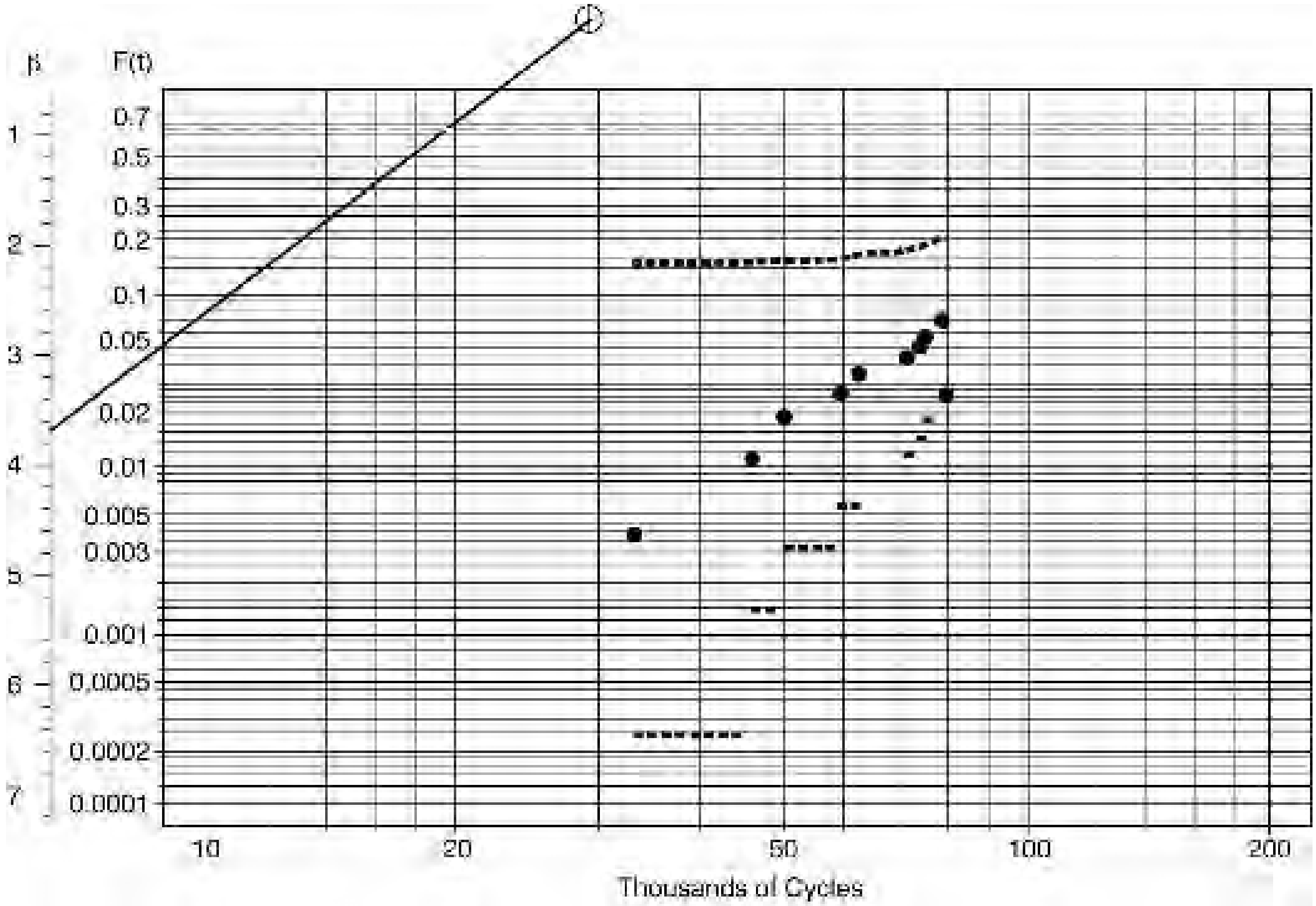

FIGURE 48.7 Weibull probability plot for the chain link failure data with 95 percent simultaneous confidence bands.

meter is $\hat{\beta} \simeq 3.62$. A graphical estimate of the Weibull scale parameter η can be obtained by drawing a line to fit the plotted points and reading off an estimate of the .63 quantile, giving $\hat{\eta} \simeq 185$ in the example. The grid lines on Figure 48.7 are useful for plotting by hand or for reading numbers from the plot. It can be argued that the grid lines interfere with the interpretation of the graph, and their inclusion should be a software option.

The light dotted points on this probability plot are 95 percent simultaneous nonparametric (i.e., without any parametric distribution assumption) confidence bands that help judge the sampling uncertainty in the estimate of $F(t)$. These confidence bands are obtained by using the method described in Nair (1984) and are defined by the inversion of a distributional goodness of fit test. If one can draw a straight line through the confidence bands, as is the case here, then the distribution implied by the probability paper (Weibull for Figure 48.7) cannot be ruled out as a model that might have generated the data.

A similar lognormal plot (not shown here) suggested the data could also have come from a lognormal distribution. The lognormal distribution assumption provides somewhat similar estimates for, say, quantiles within the range of the data, but appreciable differences outside the range.

Maximum Likelihood Estimation. The maximum likelihood (ML) method provides a formal, objective, statistical approach for fitting a life distribution, using an assumed model, and estimating various properties, such as distribution quantiles and failure probabilities. A probability plot can also be used to show the ML estimate of $F(t)$. In fact, fitting a Weibull distribution using ML is an objective way of fitting a straight line through data points on a probability plot. Computer programs are convenient for this task. Least squares (using standard linear regression methods) can also be used to fit such a line to the plotted points. ML is, however, the preferred method, especially with censored data, because of its desirable statistical properties and other theoretical justifications. Figure 48.8 is similar to Figure 48.7, but gives a computer-generated Weibull probability plot for the chain link data with the fitted ML line superimposed. The dotted lines are drawn through a set of 95 percent parametric *pointwise* normal-approximation confidence intervals for $F(t)$ (computed as described in Chapter 8 of Meeker and Escobar 1998). For some purposes, the uncertainty implied by the intervals in Figure 48.8 would suggest that more information is needed. The next subsection discusses methods for choosing the sample size to control the width of confidence intervals for Weibull quantiles.

The probability plot in Figure 48.8 might invite extrapolation beyond the range of the observed times to failure in either direction. It is important to note, however, that such extrapolation is dangerous; the data alone do not tell us about the shape of $F(t)$ outside the range of 30 to 90 thousand cycles. Moreover, the confidence intervals, even though they are wider outside the data range, reflect only the uncertainty due to limited data, and assume that the fitted Weibull model holds beyond, as well as within, the range of the data.

Guidelines for Choosing the Sample Size in Life Tests. Generally, larger samples lead to more precise estimates. This subsection shows how to choose the sample size n to control precision in a simple life test. Certain input information is required, as illustrated in the following example.

Planning Values for Sample Size Determination. For purposes of illustration assume that investigators intend to run a life test for a new mechanical device. The test will run for 30 days and the experimenters, from past experience, expect that about 7 percent of the units will fail in that time. From past experience with similar products, they expect the data to fit a Weibull distribution with a shape parameter near $\beta = 3$.

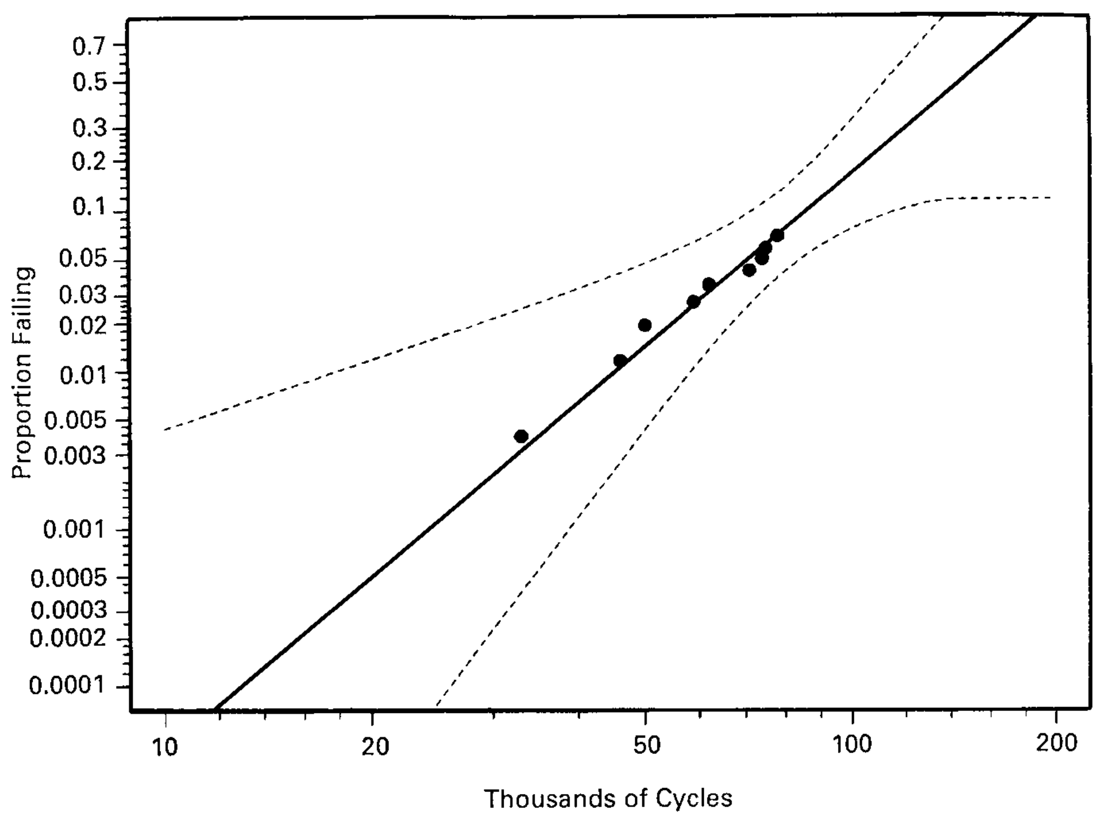

FIGURE 48.8 Weibull probability plot with the Weibull ML estimate and a set of pointwise approximate 95 percent confidence intervals for $F(t)$ for the chain link failure data.

Using Simulation to Evaluate Precision. The plots in Figure 48.9 compare simulations of ML cdf estimates for the above testing situation (Weibull distribution with $\eta = 30[-\log(1-.07)]^{-1/3} = 71.92$ days and $\beta = 3$) with a fixed censoring time of 30 days for sample sizes $n = 40$, 160, 640, and 2560; each simulation is based on 30 random samples of size n. In each plot, the thicker, longer line shows the cdf of the assumed underlying "true" population distribution, from which the simulated data were generated. The other lines show the cdf, as estimated from each of the 30 simulations. A comparison of the plots shows that with a small sample size, the estimated line can depart appreciably from the true line. Looking at the horizontal dashed line allows us to see the improvement in precision for estimating the time at which 1 percent of the population will fail as the sample size increases. For example, for $n = 40$, the 30 estimates of $t_{.01}$ range from 5.9 to 28.0 days, while for $n = 2560$, the range is 11.2 to 14.2 days. Thus, such simulations allow us to assess the precision that can be expected in the estimates from a planned study and from this choose an appropriate sample size.

Analytical Formula for Sample Size. If one prefers an analytic approach to find the sample size needed to estimate the quantile t_p of a Weibull distribution with a specified degree of precision, one can use a simple approximate formula, supplemented by a set of curves. Thus, in our example, we might wish to estimate the .01 quantile within 40 percent of its true value with 95 percent confidence. The approximate formula is

$$n = \frac{z^2_{(1-\alpha/2)} V^{\square}}{[\log(R_T)]^2}$$

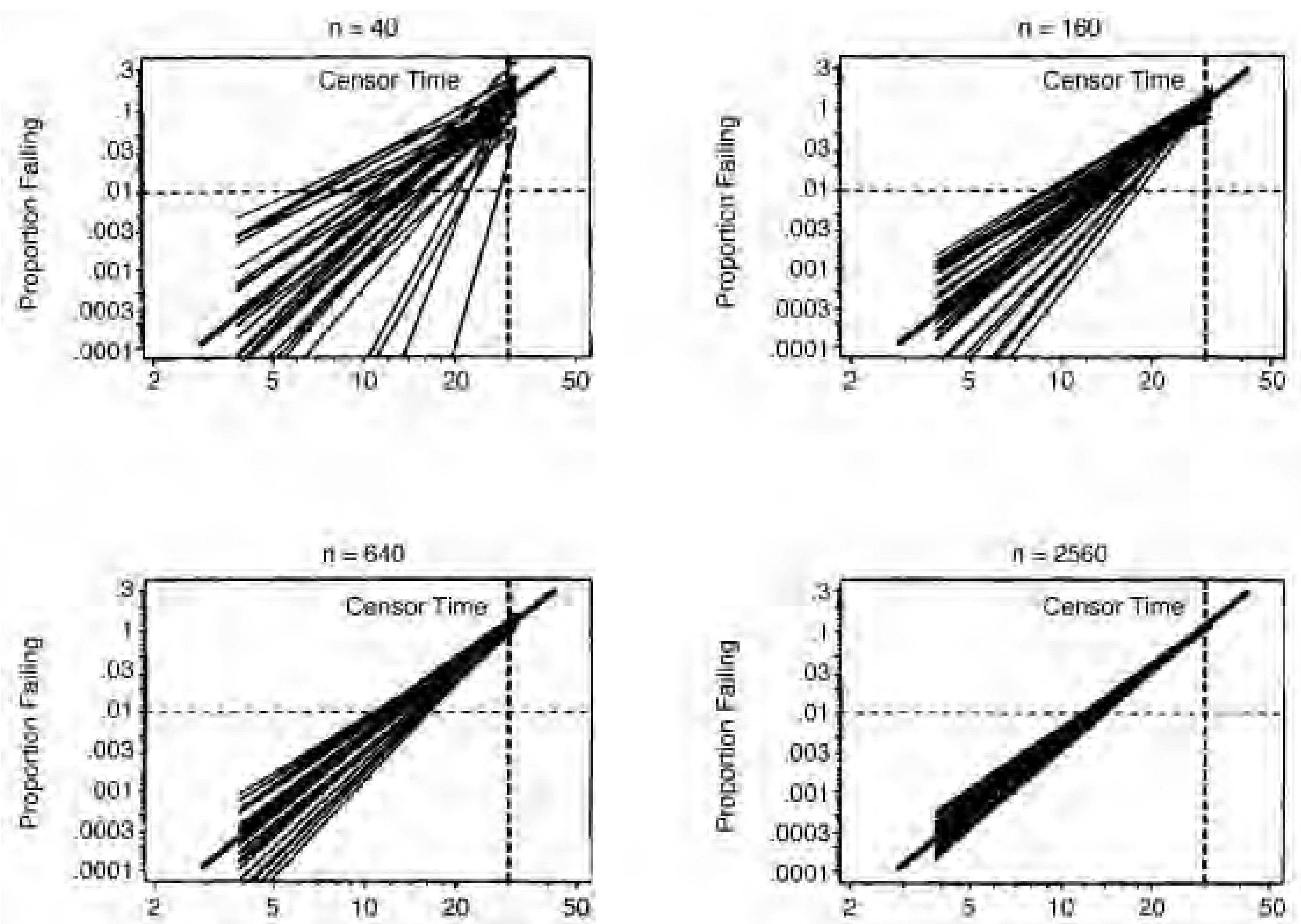

FIGURE 48.9 Weibull probability plots showing the cdf, and the estimated .01 quantile as estimated by ML, for 30 simulated samples of size n=40, 160, 640, and 2560 from a Weibull distribution with η=71.92 and β=3.

where $z_{(1-\alpha/2)}$ is the $1 - \alpha/2$ quantile of the standard normal distribution, $1 - \alpha$ is the desired level of confidence, $R_T > 1$ is the target precision, and $V^{\square}$ is the value read from Figure 48.10, multiplied by $(\beta^{\square})^2$ where $\beta^{\square}$ is a "planning value" for the Weibull distribution parameter β. In our example $\beta^{\square} = 3$. For a 95 percent confidence interval $\alpha = .05$, $1 - \alpha/2 = .975$ so $z_{.975} = 1.96$. Typical values of R_T range from 1.2 to 2, where the desire is to have confidence interval endpoints not more than $100(R_T - 1)$ percent above and below the point estimate for t_p. In our example $R_T = 1.4$. Figure 48.10 gives a plot of the large-sample approximate variance factor $\beta^2 \times V$ for estimating t_p. A smaller variance implies better estimation precision. Each line gives $\beta^2 \times V$ as a function of the quantile p whose value is to be estimated from the life test; in our example $p = .01$. There are different lines for different values of p_c, the expected proportion failing during the life test. In our example, $p_c = .07$. The curves in Figure 48.10 are then used to read $\beta^2 \times V$ using p and p_c and then solve for the value $V^{\square}$ of V, using $\beta^{\square}$. If the model and planning values are correct, the actual precision achieved will be smaller than the target precision R_T with probability approximately .5.

For the example, entering Figure 48.10 with $p_c = .07$ and $p = .01$, we read the ordinate $V^{\square} \times (\beta^{\square})^2 \simeq 70$, so $V^{\square} \simeq 70/(3)^2 = 7.778$. Thus, with $R_T = 1.4$,

$$n = \frac{z^2_{(1-\alpha/2)} V^{\square}}{[\log (R_T)]^2} = \frac{(1.96)^2(7.778)}{[\log (1.4)]^2} \simeq 264$$

is the number of items that should be tested. In summary, in order to estimate the 0.01 quantile within 40 percent of its true value with 95 percent confidence, we need to test approximately 264 units up to 30 days (under the assumption of a Weibull distribution with shape parameter 3 and an expected 7 percent failures after 30 days).

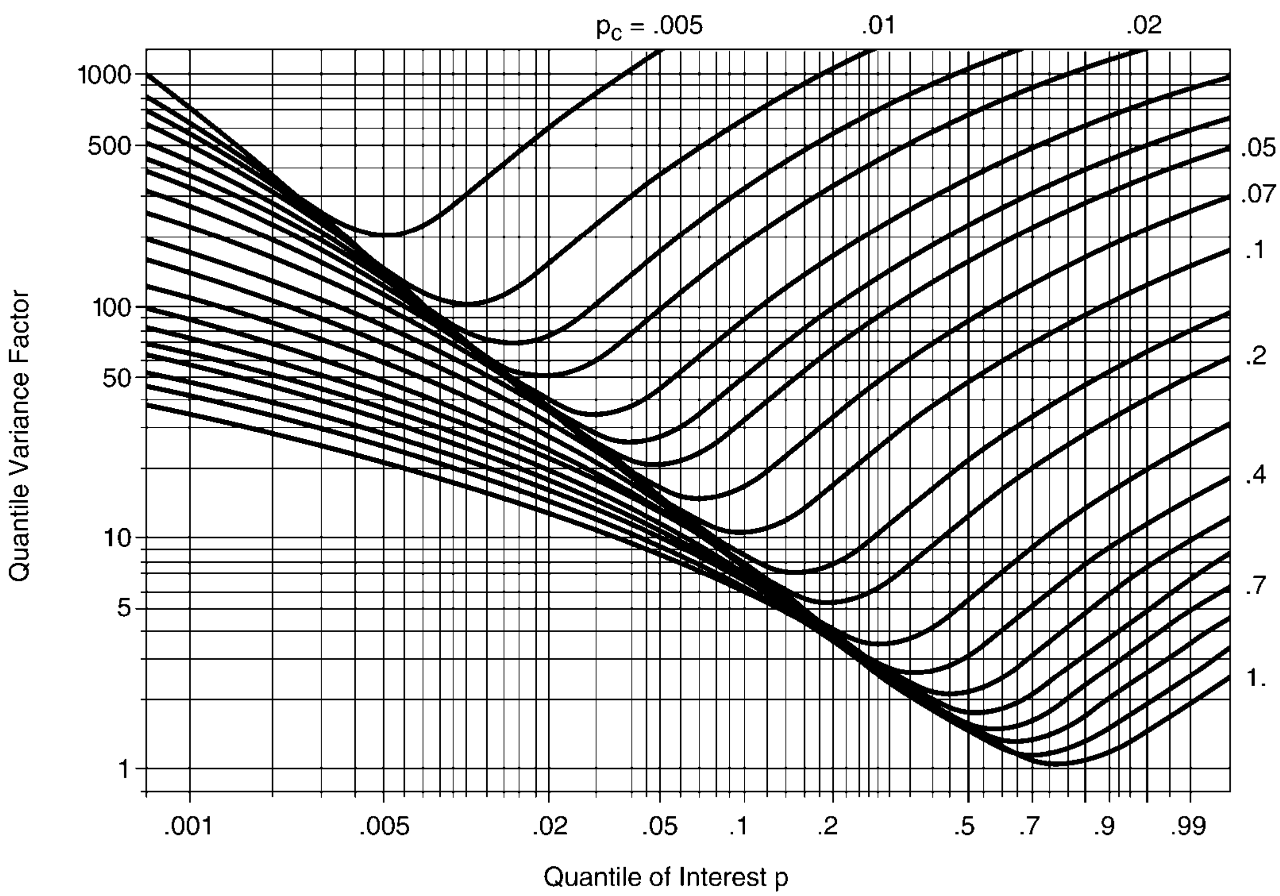

FIGURE 48.10 Large-sample approximate variance factor $\beta^2 \times V$ for ML estimation of Weibull distribution quantiles as a function of p_c, the population proportion failing by censoring time t_c and p, the quantile of interest.

Analysis of Multiply Censored Data. Multiple censoring can arise when units go into service at different times or units accumulate varying running times over a particular elapsed time period (likely in the field), or in a variety of other situations (such as the analysis of a particular failure mode, for which units that fail independently due to other modes are taken as still running with regard to the mode under evaluation). In this case, often some running times for unfailed units are less than some failure times. Multiple censoring is encountered frequently in practice, especially in dealing with field life data. For such data, the cumulative failure probability $F(t)$ cannot be estimated directly using (48.1). Instead an alternative estimator, to be described below, is used. The ML estimates and approximate confidence intervals that were previously mentioned still do apply.

Example: Shock Absorber Failure Data. O'Connor (1985) gives failure data for shock absorbers. At the time of analysis, failures had been reported at 6700, 9120, 12,200, 13,150, 14,300, 17,520, 20,100, 20,900, 22,700, 26,510, and 27,490 km. There were 27 units in service that had not failed. The running times for these units were 6950, 7820, 8790, 9660, 9820, 11,310, 11,690, 11,850, 11,880, 12,140, 12,870, 13,330, 13,470, 14,040, 17,540, 17,890, 18,450, 18,960, 18,980, 19,410, 20,100, 20,150, 20,320, 23,490, 27,410, 27,890, and 28,100 km.

To estimate $F(t)$ with multiply censored data, we use the following nonparametric procedure (i.e., no distribution model assumed). Let $t_1, t_2, \ldots, t_r$ denote the times at which failures occurred, let d_i denote the number of units that failed at t_i, and let n_i denote the number units that are unfailed just before t_i. The estimator of $F(t)$ for all values of t between t_i and t_{i+1} is

$$F(t_i) = 1 - \prod_{j=1}^{i}\left(1 - \frac{d_j}{n_j}\right) \qquad i = 1, \ldots, r \tag{48.2}$$

This is the well-known product-limit or Kaplan-Meier estimator. It can be used only within the range of the failure data. The justification for this estimator is based on the relationship between the cdf and the discrete-time hazard function and is described more completely in Chapter 3 of Meeker and Escobar (1998). Calculations of the estimated cumulative failure probability $F(t_i)$ and the estimated survival probability $S(t_i)$ to 12,200 km for the shock absorber data are illustrated in Table 48.1. This table could have gone up to 27,490 km—the largest failure time.

Figure 48.11 shows the shock absorber data plotted on Weibull probability paper, with the points plotted at each of the failure times at a height halfway between the jumps in the nonparametric cdf estimate in Table 48.1 (extended to 27,490 km). Figure 48.11, like Figure 48.8, also shows the Weibull ML cdf estimate and a set of approximate 95 percent pointwise confidence intervals for $F(t)$. The plot shows that the Weibull distribution fits the data well. A similar lognormal probability plot

TABLE 48.1 Nonparametric Estimates of $F(t)$ for the Shock Absorber Data up to 12,200 km

t_i, km	d_j	r_j	n_j	d_j/n_j	$1-d_j/n_j$	$\hat{S}(t_i)$	$\hat{F}(t_i)$
6,700	1	0	38	1/38	37/38	.97368	.0263
6,950	0	1	37				
7,820	0	1	36				
8,790	0	1	35				
9,120	1	0	34	1/34	33/34	.94505	.0549
9,660	0	1	33				
9,820	0	1	32				
11,310	0	1	31				
11,690	0	1	30				
11,850	0	0	29				
11,880	0	1	28				
12,140	0	1	27				
12,200	1	0	26	1/26	25/26	.90870	.0913

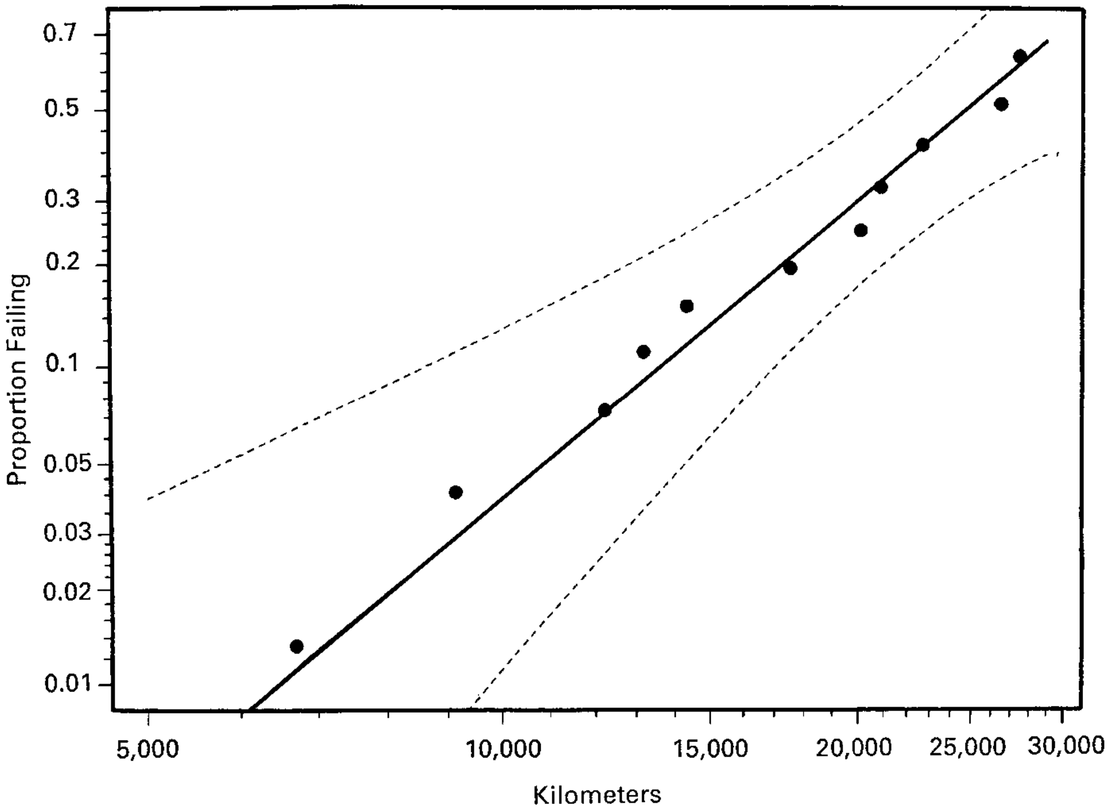

FIGURE 48.11 Weibull probability plot of shock absorber failure times with maximum likelihood estimates and approximate 95 percent pointwise confidence intervals for $F(t)$.

(not shown here) did not fit nearly as well. Although the plotted points represent a nonparametric estimate that is *not* based on any assumed distribution, the straight line drawn through the points on Weibull probability paper, by ML, freehand, or some other method is a parametric estimate that, unlike the nonparametric estimate, allows extrapolation in time. Extrapolation outside the range of the data is, however, dangerous because the assumed distribution model, though providing a reasonable fit near the center of the data, may no longer apply beyond the extremes.

Chapter 4 of Nelson (1982) describes hazard plotting, an alternative graphical method of nonparametric estimation for multiply censored data. The two methods are very similar and generally provide similar estimates.

ACCELERATED LIFE TEST MODELS AND DATA ANALYSIS

Estimating the time-to-failure distribution or long-term performance of components of *high-reliability* products is particularly difficult. Most modern products are designed to operate without failure for years, decades, or longer. Thus, we might expect that few units will fail in a test of practical length at normal use conditions. For example, the design and construction of a new communications satellite or a new appliance may allow only 8 months to test components that are expected to be in service for 15 or 20 years. For such applications, Accelerated Life Tests (ALTs) are used widely, particularly to obtain timely information on the reliability of simple product components, materials, and to provide early identification (and removal) of failure modes, thus improving reliability. There are difficult practical and statistical issues involved in accelerating life, especially for a complicated product that can fail in different ways. Here we describe accelerated life tests for components, materials, or products with a single failure mode. Extensions are described in Chapter 7 of Nelson (1990). As previously suggested, the analysis of life data with various independent failure modes often calls for considering failure modes one at a time, treating failures from other modes as if they are unfailed, and then combining the results.

Often, information from tests at high levels of stress (e.g., use rate, temperature, voltage, or pressure) is extrapolated, through a physically reasonable statistical model, to obtain estimates of life or long-term performance at lower, normal levels of stress. In some cases, stress is increased or otherwise changed during the course of a test (step-stress and progressive-stress ALTs). ALT results are used in the reliability-design process to assess or demonstrate component and subsystem reliability, certify components, detect failure modes, compare different manufacturers, and so forth as *speedily* as possible. ALTs have become increasingly important because of rapidly changing technologies, more complicated products with more components, the pressures for rapid new product introduction, and higher customer expectations for better reliability (in place of a traditional "ship and fix later" approach).

Methods of Acceleration. Three different methods of accelerating a reliability test are

- Increase the use rate (or cycling rate) of the product. A typical toaster is designed for a median lifetime of 20 years with a usage rate of twice each day. If, instead, we test the toaster 365 times each day, we could reduce the median lifetime (for many potential failure modes) to about 40 days. Also, because it is not necessary to have all units fail in a life test, useful reliability information might be obtained in a matter of days instead of months. This form of acceleration is the simplest to handle, since the methods for a single distribution can generally be applied directly (after making an appropriate time transformation). Thus, this situation will not be discussed further. The key assumption is that the increased use rate will not introduce new failure modes or change the damage per cycle rate for existing failure modes (e.g., due to overheating of cycled components).
- Increase the aging rate of the product. For example, increasing temperature or humidity can accelerate the chemical processes of certain failure modes such as chemical degradation (resulting in eventual weakening) of an adhesive mechanical bond or the growth of a conducting filament across an insulator (eventually causing a short circuit).
- Increase the level of stress (e.g., voltage or pressure) under which test units operate. A unit will fail when its *strength* drops below applied stress. Thus, a unit with degrading strength at a high level of stress will generally fail more rapidly than it would have failed at a low level of stress.

Combinations of these methods of acceleration can also be employed. A factor like voltage will, for some failure modes, both increase the rate of an electrochemical reaction (thus accelerating the aging rate) and increase stress relative to strength. In such situations, when the effect of stress is complicated, there may not be enough knowledge to provide an adequate physical model for acceleration. Empirical models may or may not be useful. See Chapter 2 of Nelson (1990) or Chapter 18 of Meeker and Escobar (1998) for more discussion.

Acceleration Models. Analysis of accelerated life tests, other than those that just increase the product use rate, requires a physically reasonable model relating acceleration variables like temperature, voltage, pressure, and size to time to failure. One then fits the model to the data to describe the effect that the factors have on failure.

Physical Acceleration Models. For well-understood failure modes, it may be possible to use a model based on physical/chemical theory that will be capable not only of describing the failure-causing process over the range of the data, but also allow extrapolation to use conditions based on accelerated testing. Usually the actual relationship between acceleration variables and the actual failure mechanism is extremely complicated. Often, however, it is possible to find a physical simplification that will adequately capture and describe the dominant aspect of the process. For example, failure may be affected by a complicated chemical process with many steps, but there may be one rate-limiting (or dominant) step, and a good understanding of this part of the process may provide a model that is adequate for extrapolation. See Meeker and LuValle (1995) and Chapter 2 of Nelson (1990) for examples.

Empirical Acceleration Models. When there is little understanding of the chemical or physical processes leading to failure, it may be impossible to develop a model based on physical/chemical theory. An empirical model may be the only alternative. However, such a model might provide an excellent fit to the available data, but may be highly incorrect when extrapolated to the use conditions of real interest. In some situations there may be extensive experience with an empirical model and this may provide the needed basis for extrapolation to use conditions.

Some guidelines for the use of acceleration models include:

- Use physical theory as much as possible, and choose accelerating variables (e.g., temperature, voltage, or humidity) that correspond to factors that cause actual failures.
- Investigate previous attempts to accelerate failure modes similar to the ones of interest. There are, for example, many research reports and papers on physics of failure. Chapter 2 of Nelson (1990) or Chapter 18 of Meeker and Escobar (1998) provide examples and references.
- Design accelerated tests to minimize, as much as possible, the degree of extrapolation required [see Meeker and Hahn (1985), Chapter 6 of Nelson (1990), and Chapter 20 of Meeker and Escobar (1998)]. Highly accelerated testing can cause failure modes that would not occur at use conditions. If extraneous failures are not recognized they can lead to seriously incorrect conclusions. Note that product life data can be analyzed, using maximum likelihood methods, even though at some test conditions the majority, or even all, of the test units remain unfailed. This allows us to be much more conservative in establishing testing conditions than would otherwise be the case, (i.e., include tests at or close to the use conditions). See Meeker and Hahn (1985) or Nelson (1990) for details.
- Most accelerated tests are used to obtain information about a single, relatively simple, failure mode (or corresponding degradation measure). If there is more than one failure mode, it is likely that the different failure modes will be accelerated at different rates and, unless this is accounted for in the modeling, testing, and analysis, the resulting inferences could be seriously incorrect.
- In practice, acceleration relationships are difficult to verify in their entirety. The accelerated life test data should be used to look for departures from the assumed acceleration model. It is important to recognize, however, that the available data will generally provide very little power to detect anything but the most serious departures, and typically there is no useful diagnostic information close to use conditions.
- Simple models with the right shape have generally proven to be more useful for extrapolation to use conditions than elaborate multiparameter models. For example, even if there is some evidence of curvature in ALT data, a simple linear model relating the accelerated stress variable to time to failure will generally provide more accurate extrapolation than a fitted quadratic model.
- Sensitivity analyses should be used to assess the effect of perturbing uncertain inputs (e.g., inputs related to model assumptions). For example, it is useful to compare the extrapolations to use conditions with different assumed time-to-failure distributions, or different physically reasonable models relating stress to life.
- Include some test units at conditions at which no failures would be expected during the duration of the test (e.g., at or near use conditions). If failures do, indeed, take place at these conditions it would raise serious concerns about reliability (and suggest that the assumed model is inadequate).
- Accelerated life test programs should be planned and conducted by teams including individuals knowledgeable about the product and its use environment, the physical/chemical/mechanical aspects of the failure mode, and the statistical aspects of the design and analysis of reliability experiments.

The next subsection describes a simple temperature-acceleration model and illustrates the analysis of data from an ALT.

Elevated Temperature Acceleration. High temperature has been said to be the enemy of reliability. Increasing temperature is one of the most commonly used methods to accelerate failures.

The Arrhenius relationship is a well-known model describing the effect that temperature has on *R,* the rate of a simple chemical reaction. This model for reaction rate can be written as

$$R(\text{temp}) = \gamma_0 \exp\left(\frac{-E_a}{k_B \times \text{temp K}}\right) = \gamma_0 \exp\left(\frac{-E_a \times 11{,}605}{\text{temp K}}\right)$$

where temp K = temp °C + 273.15 is temperature in the absolute Kelvin scale, $k_B = 1/11{,}605$ is Boltzmann's constant in units of electron volts per degrees Celsius and E_a is the reaction activation energy in electron-volts. The parameters E_a and γ_0 are generally unknown product or material characteristics. In some simple situations, the failure-causing process can be modeled adequately by such a simple chemical reaction. In such cases, if log (T) follows a distribution with a location and scale parameter (so that the time to failure T could, for example, be described by a lognormal distribution or a Weibull distribution) with parameters μ and σ then the life distribution can be written as a function of temperature:

$$\Pr(T \le t;\ \text{temp}) = \Phi\left[\frac{\log(t) - \mu(\text{temp})}{\sigma}\right]$$

where $\Phi = \Phi_{\text{nor}}$, and $\Phi = \Phi_{\text{sev}}$ (see earlier discussion) for the lognormal and Weibull distributions, respectively, $\mu(\text{temp}) = \beta_0 + \beta_1 x$, $x = 11{,}605/(\text{temp K})$, and $\beta_1 = E_a$. This is known as the Arrhenius failure-time model. One would then use life data at various accelerated temperatures and the resulting observed times to failure to estimate β_0, β_1, and σ.

Although the Arrhenius model does not apply to all temperature-acceleration problems and will be adequate over only a limited range of temperatures (depending on the particular application), it is used widely in many areas of application. Nelson (1990, p. 76) comments that "…in certain applications (e.g., motor insulation), if the Arrhenius relationship…does not fit the data, the data are suspect rather than the relationship."

Similar life-stress regression relationships have been used for other accelerating variables like voltage, pressure, humidity, cycling rate, and specimen size. See Chapter 2 of Nelson (1990) and Chapter 18 of Meeker and Escobar (1998) for examples and further references.

Planning an ALT. Usually accelerated life tests need to be conducted within stringent cost and time constraints. Careful planning is essential. Resources need to be used efficiently, and the degree of extrapolation minimized, to the greatest degree possible. Meeker and Hahn (1985), Chapter 6 of Nelson (1990), and Chapter 20 of Meeker and Escobar (1998) describe methods for planning statistically efficient ALTs that meet practical constraints. During the test planning phase of a study, experimenters should try to explore the kind of results that they might obtain as a function of the assumed model and proposed test plan. Simulation methods, such as described in discussion of the guidelines for choosing sample size, also provide useful insights in this process.

Strategy for Analyzing ALT Data. This subsection outlines a useful procedure for analyzing ALT data when multiple units are life tested at three or more conditions (e.g., three or more temperatures). See, for example, the data in Table 48.2.

The basic idea is to start by analyzing the data at each test condition separately and then to fit a model that ties together the data at the different conditions. Briefly, the strategy, as illustrated in the example that follows, is to:

1. Construct probability plots of the data at each test condition (level of the accelerating factor) separately, to suggest and explore the adequacy of possible distributional models.
2. At each test condition with two or more failures, fit models, as suggested by the previous step, individually to the data using the ML method. Plot the ML lines on a multiple probability plot for each of these conditions. Use the plotted points and fitted lines to assess the reasonableness of assumptions (such as a constant standard deviation at each test condition, e.g., Figure 48.12).

TABLE 48.2 Failure Intervals or Censoring Times and Testing Temperatures from an ALT Experiment on a New-Technology Integrated Circuit Device

Temperature °C	Number of devices	Hours Lower	Hours Upper	Status
150	50		1536	Censored
175	50		1536	Censored
200	50		96	Censored
250	1	384	788	Failed
250	3	788	1536	Failed
250	5	1536	2304	Failed
250	41		2304	Censored
300	4	192	384	Failed
300	27	384	788	Failed
300	16	788	1536	Failed
300	3		1536	Censored

3. Fit a model to the assumed relationship between life and the accelerating variable. Plot the fitted model on probability paper, together with a plot of the data (e.g., Figure 48.13).
4. Perform residual analyses and other diagnostic checks of the model assumptions (as in standard regression analysis). See Chapter 5 of Nelson (1990). (This is not done for the example here.)
5. Obtain the desired estimates and confidence intervals, and assess their reasonableness (e.g., Figure 48.14).

For further examples and further discussion of methods for analyzing ALT data, see Nelson (1990) or Meeker and Escobar (1998).

Example. Table 48.2 gives the results of an ALT on a new-technology integrated circuit (IC) device. The device inspection process involved an electrical diagnostic test conducted periodically to determine whether or not failure had occurred. This test was expensive because it was manual and required much time on a special machine. Thus, only a few inspections could be conducted on each device. One common method of planning the times for such inspections is to choose a first inspection time and then space the further inspections to be equally spaced on a log scale. In this case, the first inspection was after 1 day with subsequent inspections at 2 days, 4 days, and so on (except for a day when the person doing the inspection had to leave early). Tests were run at 150, 175, 200, 250, and 300°C. Life tests in which failures are recorded only at inspection times lead to interval-censored data (sometimes called "interval" or "read out" data). When a unit fails the test, all that is known is that there has been a failure in the elapsed time since the previous inspection. Interval-censored data require special statistical methods that are described in Chapter 9 of Nelson (1982), Chapter 3 of Nelson (1990), and Chapters 3, 7, and 19 of Meeker and Escobar (1998). Table 48.2 gives the interval in which failure occurred for the devices that failed. Correspondingly, for devices that did not fail, Table 48.2 gives the running time at the last inspection.

At the time of analysis, failures had been found only at the two highest temperatures. After early failures had been observed at 250 and 300°C there was some concern that no failures would be observed at 175°C before the time at which decisions would have to be made. Thus the 200°C test was started later than the others and only limited running time was accumulated by the time of the analysis.

The developers were interested in estimating the activation energy of the suspected failure mode and the long-life reliability of the components. In particular, they wanted to estimate the proportion of devices in the product population that would fail by 100 thousand hours (about 11 years) at the use condition of 100°C.

Figure 48.12 is a lognormal probability plot and estimated ML cdf of the failures at 250 and 300°C along with the ML estimates of the individual lognormal cdfs. The differing slopes in the plot suggest the possibility that the lognormal shape parameter sigma changes from 250 to 300°C. Such a change could be caused by a change in failure mode. Failure modes with a higher activation energy, that might never be seen at low levels of temperature, can appear at higher levels of temperature (or other acceleration factors). An approximate 95 percent confidence interval on $\sigma_{250}/\sigma_{300}$ is [1.01, 3.53] (calculations not shown here), suggests that the difference could be real. These results also suggested that detailed physical failure mode analysis should be done for at least some of the failed units and that, perhaps, the accelerated test should be extended until some failures are observed at lower levels of temperature. However, in the subsequent analyses we shall assume that σ is constant at the different temperatures.

Table 48.3 gives Arrhenius-lognormal model ML estimation results and confidence bounds for the new-technology IC device under the assumed model. Figure 48.13 is a lognormal probability plot of the data at 250°C and 300°C showing the Arrhenius-lognormal model cdf fit to the data. This figure also shows lognormal cdf estimates from the fitted model for the other test temperatures as well as the estimated cdf, and an approximate 95 percent confidence interval at the use condition of 100°C. We note that the upper bound of the 95 percent confidence interval on the proportion failing by 100,000 hours at this temperature is less than .0001 (0.01 percent).

Figure 48.14 is an Arrhenius plot of the lognormal model fit to the new-technology IC device ALT data, showing the times at which 1, 10, and 50 percent of the units are expected to fail as a function of time. This plot shows the rather extreme extrapolation needed to make inferences at the use conditions of 100°C. If the projections are close to reality, it appears unlikely that there will be any failures below 200°C during the remaining 3000 hours of testing, and, as mentioned before, this was the reason for testing additional units at 200°C.

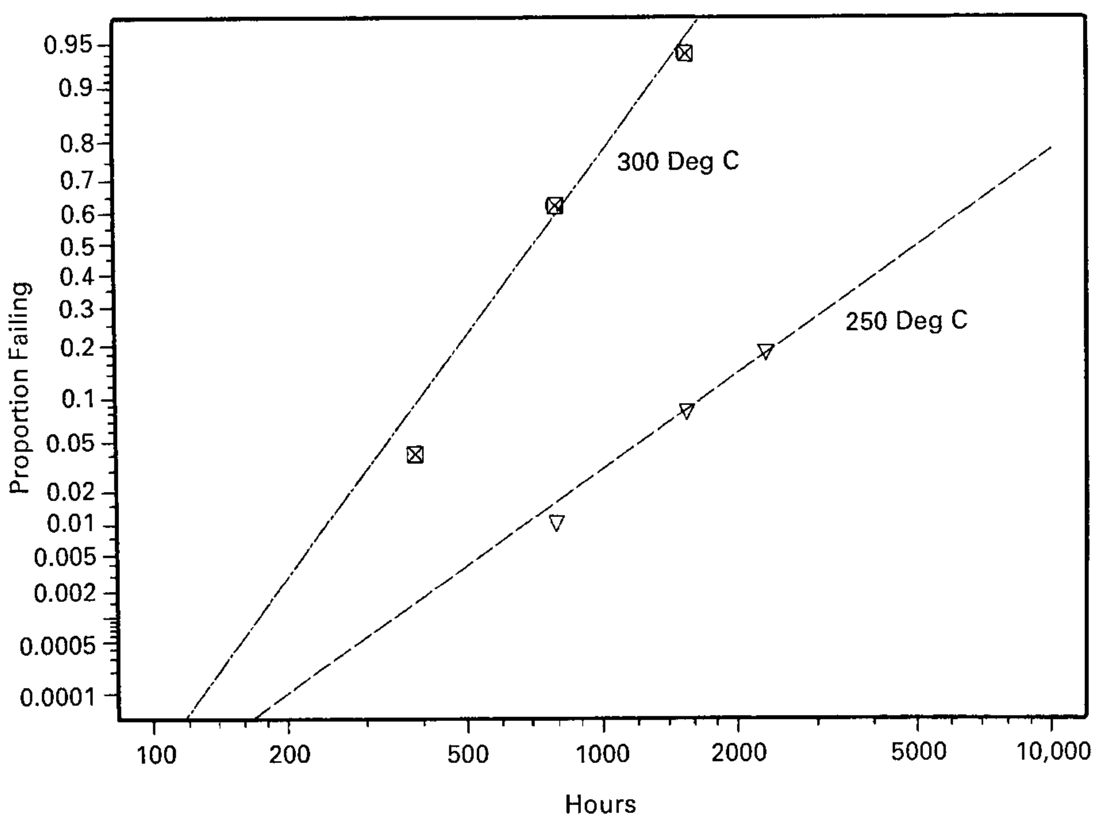

FIGURE 48.12 Lognormal probability plot and estimated ML cdf of the failures at 250 and 300°C for the new-technology integrated circuit device ALT experiment.

TABLE 48.3 Individual Lognormal ML Estimation Results for the New-Technology IC Device*

			Approximate 95% confidence intervals	
Parameter	ML estimate	Standard error	Lower	Upper
β_0	−10.2	1.5	−13.2	−7.2
β_1	.83	.07	.68	.97
σ	.52	.06	.42	.64

*The loglikelihood is $\mathscr{L} = -88.36$. The confidence intervals are based on the normal approximation method.

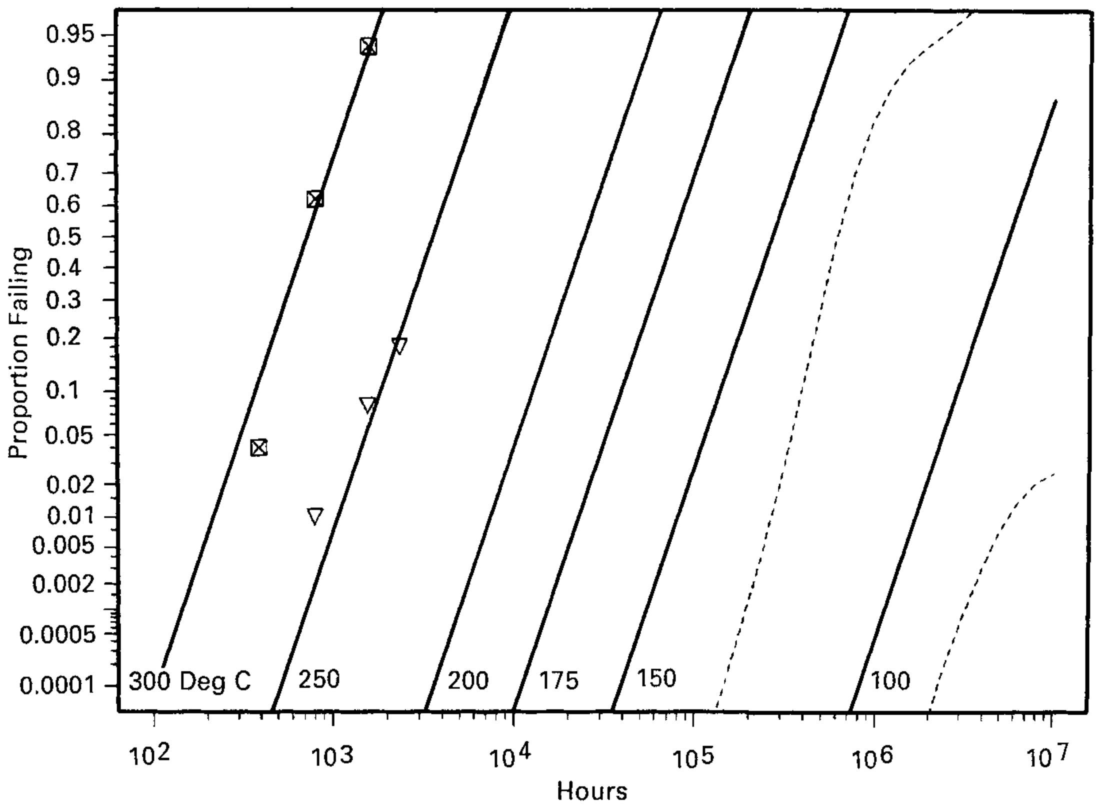

FIGURE 48.13 Lognormal probability plot showing the Arrhenius-lognormal model ML estimation results for the new-technology IC device at the test temperatures and (with the approximate 95 percent confidence interval) at 100°C.

SYSTEM RELIABILITY CONCEPTS

The system failure probability $F_T(t)$ is the probability that the system fails before time t. The failure probability of the system is a function of time in operation t (or other measure of use), the operating environment(s), the system structure, and the reliability of system components, interconnections, and interfaces (including, for example, human operators).

This subsection describes several simple system structures. Not all systems fall into one of these categories, but the examples provide building blocks to illustrate the basics of using system structure to compute system reliability. Complicated system structures can generally be decomposed into collections of the simpler structures presented here. Thus the methods for evaluation of system relia-

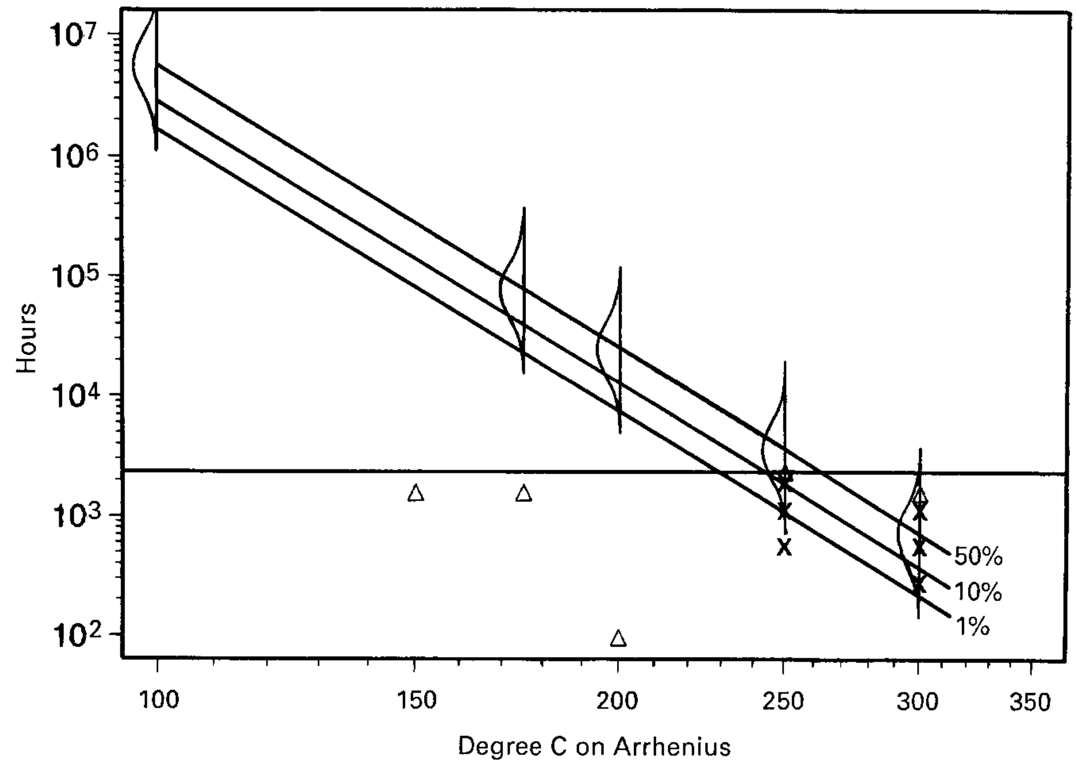

FIGURE 48.14 Arrhenius plot showing the Arrhenius-lognormal model ML estimates for the new-technology IC device. Failure and censored observation times are indicated by × and Δ respectively.

bility can be adapted to more complicated structures. For more information, see texts such as O'Connor (1985), Høyland and Rausand (1994), or Meeker and Escobar (1998).

Terminology for System Reliability. Consider the time to failure of a system (i.e., all components starting a time 0) with s independent components. The cdf for component i is denoted by $F_i(t)$. The corresponding reliability (or survival probability) is $S_i(t) = 1 - F_i(t)$. The cdf for the system is denoted by F_T. This cdf is determined by the F_i's and the system structure, that is, $F_T(t) = g[F_1(t), \ldots, F_s(t)]$. To simplify the presentation, this function will be expressed as $F_T = g(F_1, \ldots, F_s)$.

Systems with Components in Series. A series structure with s components works if and only if all the components work. Examples of systems with components in series include chains, high-voltage multicell batteries, inexpensive computer systems, and inexpensive decorative tree lights using low-voltage bulbs. For a system with two independent components in a series, illustrated in Figure 48.15, the cdf is

$$\begin{aligned} F_T(t) &= \Pr(T \le t) = 1 - \Pr(T > t) = 1 - \Pr(T_1 > t \text{ and } T_2 > t) \\ &= 1 - [\Pr(T_1 > t)\ \Pr(T_2 > t)] = 1 - (1 - F_1)(1 - F_2) \end{aligned} \tag{48.3}$$

FIGURE 48.15 A system with two components in series.

For a system with s independent components $F_T(t) = 1 - \Pi_{i=1}^{s} (1 - F_i)$ and for a system with s independent components, each of which has the same time to failure distribution [for example, multicell batteries ($F = F_i$, $i = 1, \ldots, s$)], $F_T(t) = 1 - (1 - F)^s$. The system hazard function, for a series system of s independent components, is the sum of the component hazard functions:

$$h_T(t) = \sum_{i=1}^{s} h_i(t)$$

Figure 48.16 shows the relationship between system reliability $1 - F_T(t)$ and individual component reliability $1 - F_T(t)$ for systems with different numbers of identical independent components in series. This figure shows that extremely high component reliability is needed to maintain high system reliability if the system has many components in series. If the system components are not independent, then the first line of (48.3) still gives $F_T(t)$, but the evaluation has to be done with respect to the bivariate distribution of T_1 and T_2 with a similar generalization to a multivariate distribution for more than two components.

Importance of Part Count in Product Design. An important rule of thumb in reliability engineering design practice is "keep the part count small," meaning keep the number of individual parts (or components in series) in a system to a minimum. Besides the cost of purchasing and handling of additional individual parts, there is also an important reliability motivation for having a smaller number of parts in a product. For example, the design of a new-technology computer modem uses a higher level of microelectronic integration and requires only 20 discrete parts instead of the 40 parts required in the previous generation. As a specific example, the system hazard function $h_T(t)$ becomes particularly simple if a constant hazard rate (or equivalently, an exponential time-to-failure distribution) provides an adequate model for time to failure for the system components or parts. In this case, as a rough approximation, assuming that all failures are due to part failures, and that the parts have the same hazard function, the new design with only 20 parts will experience only half of the failures of the old design. Allowing that failures can occur at interfaces and interconnections between parts with the same frequency in the new and old designs would further increase the reliability improvement because of the larger number of such interfaces with a higher number of parts. With nonconstant hazard function (more common in practice), and parts with different hazard functions, the idea is similar.

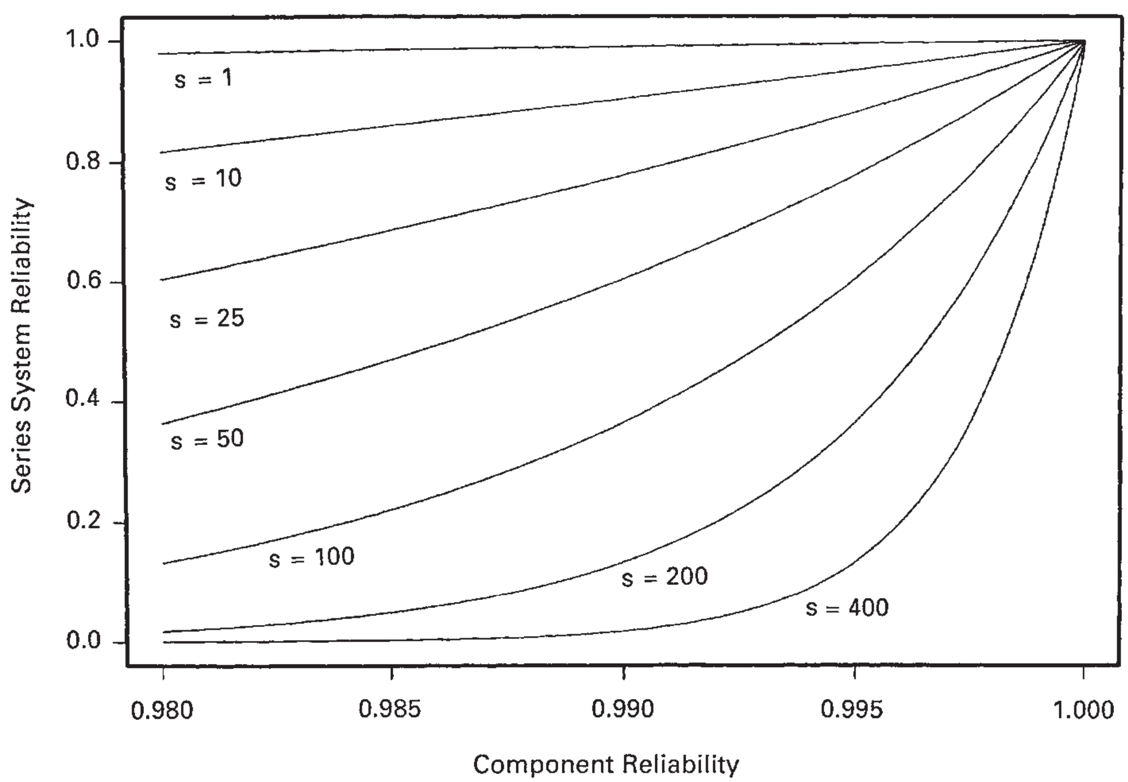

FIGURE 48.16 Reliability of a system with *s* identical independent components in series.

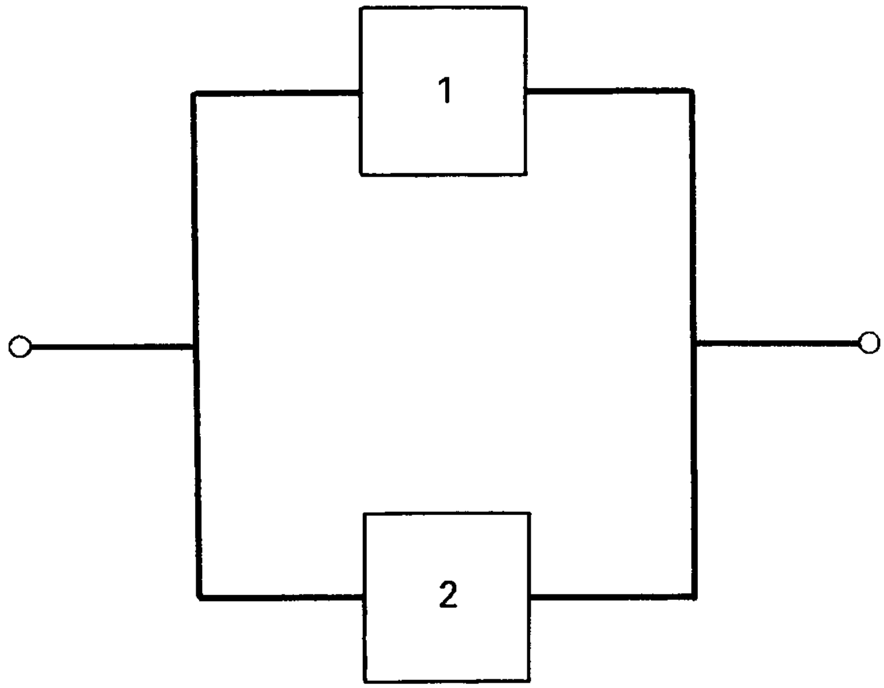

FIGURE 48.17 A system with two components in parallel.

This illustration assumes, of course, that the reliability of the parts in the new system will be the same (or at least similar to) the reliability of the individual parts of the old system, and that the stress in operation on each part remains the same. If the new system uses parts from an immature production process with low part reliability, or the operating stress on individual components is increased, the new system could have lower reliability.

Systems with Components in Parallel. Parallel redundancy is often used to improve the reliability of weak links or critical parts of larger systems. A parallel structure with s components works if at least one of the components works. Examples of systems with s components in parallel include automobile headlights, RAID computer disk array systems, stairwells with emergency lighting, overhead projectors with a backup bulb, and multiple light banks in a classroom. For two independent parallel components, illustrated in Figure 48.17,

$$\begin{aligned} F_T(t) &= \Pr\,(T \le t) = \Pr\,(T_1 \le t \text{ and } T_2 \le t) \\ &= \Pr\,(T_1 \le t)\,\Pr\,(T_2 \le t) = F_1F_2 \end{aligned} \tag{48.4}$$

For s independent components $F_T(t) = \Pi_{i=1}^{s} F_i$ and for s independent identically distributed components ($F_i = F$, $i = 1, \ldots, s$), $F_T(t) = F^s$. Figure 48.18 shows the relationship between system reliability $1 - F_T(t)$ and individual component reliability $1 - F(t)$ for different numbers of identical independent components in parallel. The figure shows the dramatic improvement in reliability that parallel redundancy can provide. If the components are not independent, then the first line of (48.4) still gives $F_T(t)$, but the evaluation has to be done with respect to the bivariate distribution of T_1 and T_2. A similar generalization applies for more than two nonindependent components.

REPAIRABLE SYSTEM DATA

The previous discussion dealt with reliability data analysis for nonrepairable components (or devices). Since a nonrepairable component can fail only once, time to failure data from a sample of

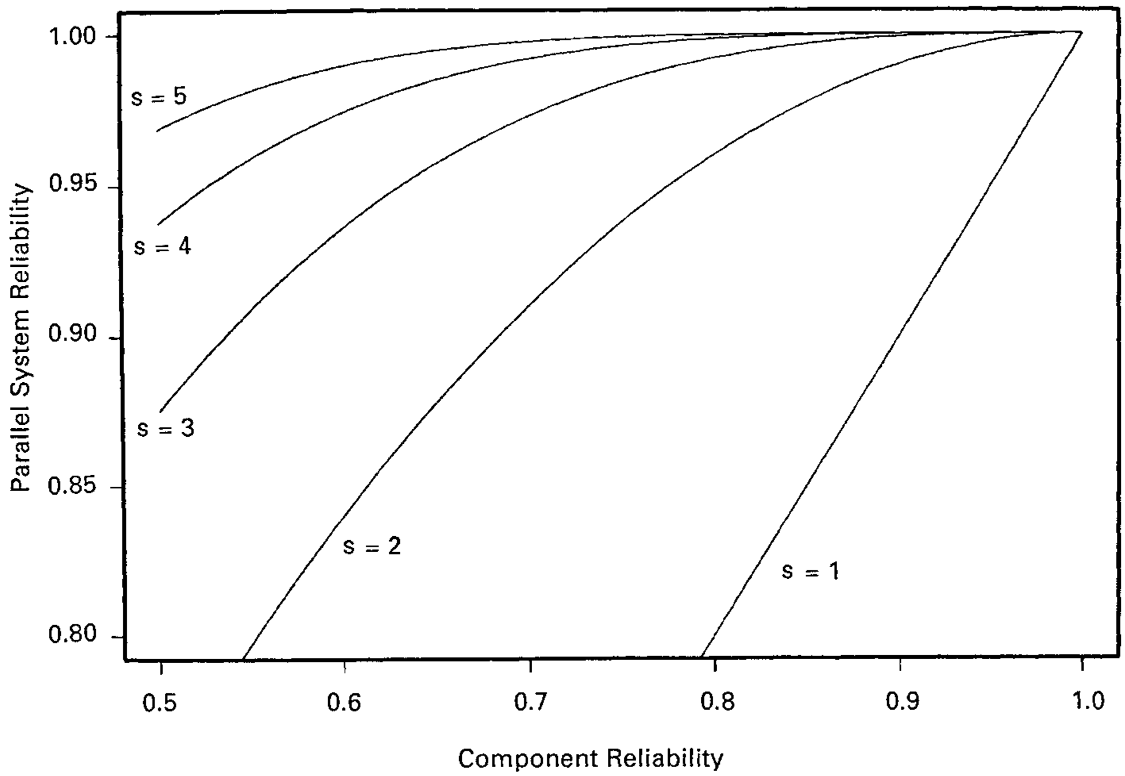

FIGURE 48.18 Reliability of a system with s independent components with identical time to failure distributions, in parallel.

nonrepairable components consist of the times to first failure for each component. In most instances involving nonrepairable components, the assumption of independent and identically distributed failure times is a reasonable one and suitable lifetime distributions (such as the Weibull or lognormal) are used to describe the distribution of failure times. In contrast, repairable system data typically consist of multiple repair times (or times between repairs) on the same system since a repairable system can be placed back in service after repair.

The purpose of many reliability studies is to describe the trends and patterns of repairs on failures over time for an *overall system* or collection of systems. Data consist of a sequence of system repair times for similar systems. When a single component or subsystem in a larger system is repaired or replaced after a failure, the distribution of the time to the next system repair will depend on the overall state of the system at the time just before the repair and the nature of the repair. Thus, repairable system data, in many situations, should be described with models that allow for changes in the state of the system over time or for dependencies between repairs over time.

Repairable system data can be viewed as sequence of repair times $T_1, T_2, \ldots$. The model for such data is sometimes called a "point process." Some applications have repair data on only one system. In most applications there are data from a collection of systems, typically monitored over a fixed observation period (t_0, t_a), where, often, $t_0 = 0$. The observation period often differs from system to system. In some cases, exact repair times are recorded. In other cases, the numbers of repairs within time intervals are reported.

From the repair data, one would typically like to estimate:

- The distribution of the times between repairs, $\tau_j = T_j - T_{j-1}$ $(j = 1, 2, \ldots)$ where $T_0 = 0$
- The number of repairs in the interval $(0, t)$ as a function of t
- The expected number of repairs in the interval $(0, t)$ as a function of t
- The recurrence rate of replacements as a function of time t

These questions lead to the analyses in the following example.

Example: Times of Replacement of Diesel Engine Valve Seats. The following data will be used to illustrate the methods. Repair records for a fleet of 41 diesel engines were kept over time. Table 48.4 gives the times of replacement (in number of days of service) of the engine's valve seats. This is an example of data on a group of systems. The data were given originally in Nelson and Doganaksoy (1989) and also appear in Nelson (1995). Questions to be answered from these data (not all of which are discussed here) include:

- How many replacements can the fleet owner expect, on the average, during the first year of operation of a locomotive?
- Does the replacement rate increase or decrease with diesel engine age and at what rate?
- How many replacement valves will be needed in the next two calendar years for this fleet?
- Can the valve replacement times be modeled as a renewal process (so that simple methods for independent observations can be used for further analysis and prediction)?

Simple data plots provide a good starting point for analysis of system repair data. Figure 48.19 is an event plot of the valve seat repair data showing the observation period and the reported replacement times. Note that the length of the observation period differed from engine to engine.

Initial evaluations of the data suggested some differences in repair vulnerability among locomotives. This is a separate subject of study, and is ignored in the following evaluations.

Nonparametric Model for Point Process Data. For data on a single system, the cumulative number of repairs up to time t is denoted as $N(t)$. The corresponding model, used to describe a population of systems from the same population, is based on the mean cumulative function (MCF) at time t. The MCF is defined as the average or expected number of repairs per system before time

TABLE 48.4 Diesel Engine Age at Time of Replacement of Valve Seats

Engine ID	Number of days observed	Engine age at replacement time, days	Engine ID	Number of days observed	Engine age at replacement time, days
251	761		403	593	
252	759		404	589	573
327	667	98	405	606	165 408 604
328	667	326 653 653	406	594	249
329	665		407	613	344 497
330	667	84	408	595	265 586
331	663	87	409	389	166 206 348
389	653	646	410	601	
390	653	92	411	601	410 581
391	651		412	611	
392	650	258 328 377 621	413	608	
393	648	61 539	414	587	
394	644	254 276 298 640	415	603	367
395	642	76 538	416	585	202 563 570
396	641	635	417	587	
397	649	349 404 561	418	578	
398	631		419	578	
399	596		420	586	
400	614	120 479	421	585	
401	582	323 449	422	582	
402	589	139 139			

Source: Nelson and Doganaksoy (1989).

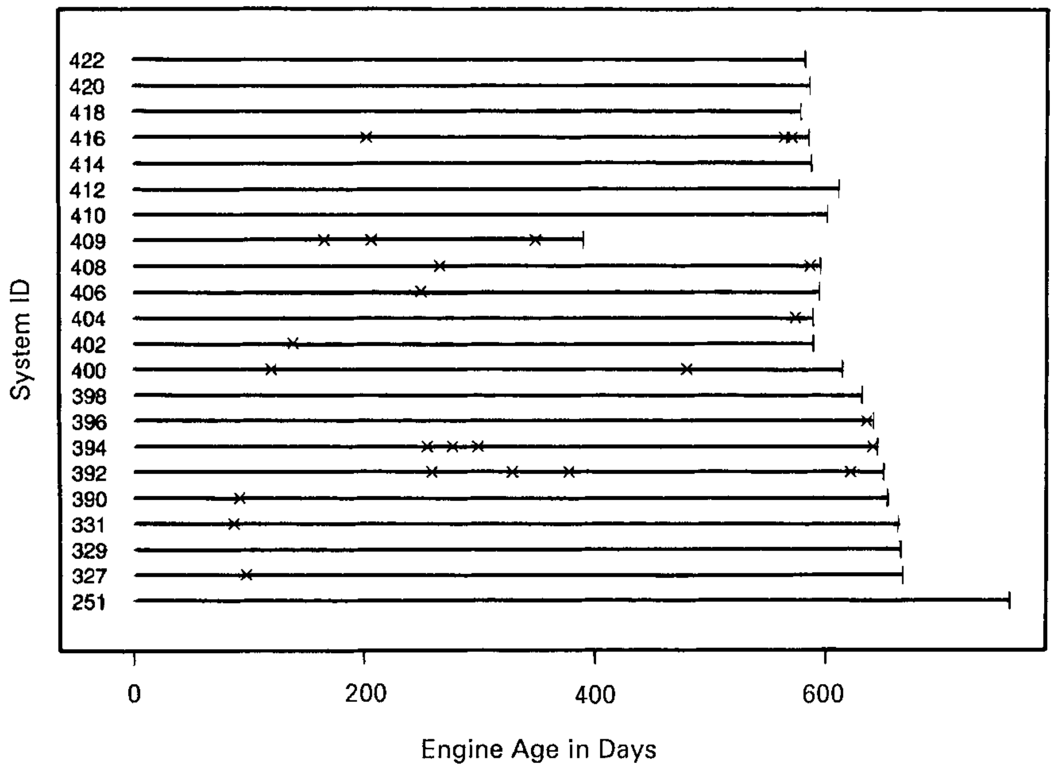

FIGURE 48.19 Valve seat replacement event plot for a subset of 22 of the 43 diesel engines.

t, i.e., $\mu(t) = E[N(t)]$, where the expectation is over the entire population of systems. Assuming that $\mu(t)$ is differentiable,

$$v(t) = \frac{dE\,[N(t)]}{dt} = \frac{d\mu(t)}{dt}$$

defines the recurrence rate of repairs per system for the system population. This can also be interpreted as an *average* rate of repair occurrence for individual systems.

Although data on number of repairs (or other specific events related to reliability) are encountered frequently, the methods given here can be used to model other quantities accumulating in time, including continuous variables like cost. Then, for example, if $C(t)$ is the cumulative repair cost for a system up to time t, then $\mu(t) = E[C(t)]$ is the average cumulative cost per system in the time interval $(0, t)$.

Nonparametric Estimation of the MCF. Given $n \geq 1$ repairable systems, the following method can be used to estimate the MCF. The method is nonparametric in the sense that it does not require specification of a parametric model for the repair-time point process. The method assumes that the sample data are taken randomly from a population of MCF functions. It is also assumed that the time at which we stop observing a system does not depend on the process. Thus, it is important that the time at which a unit is censored is not systematically related to any factor related to the repair time distribution. Biased estimators will, for example, result if units follow a staggered entry into service (e.g., one unit put into service each month) and if there has been a design change that has increased the repair probability of the more recent systems introduced into service. Then newer systems have a more stressful life and will fail earlier, causing an overly optimistic trend over time on the estimated recurrence rate $v(t)$. In such cases, data from different production periods must be analyzed separately or the change in the recurrence rate needs to be modeled as a function of system age and calendar time.

Let $N_i(t)$ denote the cumulative number of system repairs for system i at time t and let t_{ij}, $j = 1, \ldots, m_i$ be the failure (or repair, or other event) times for system i. A simple estimator of the MCF at time t would be the sample mean of the available $N_i(t)$ values for the systems still operating at time t. This estimator is simple, but appropriate only if all systems are still operating at time t. Thus, this method can be used in the diesel engine example up to $t = 389$ days, or up to 578 days if we ignore the data on engine 409. A more appropriate estimator, allowing for multiple censoring, and providing an unbiased estimate of the MCF is described in Nelson (1988), Nelson (1995), Lawless and Nadeau (1995), and Meeker and Escobar (1998). A plot of the MCF estimate versus age indicates whether the reliability of the system is increasing, decreasing, or unchanging over time.

MCF Estimate for the Valve Seat Replacements. Figure 48.20 shows the estimate of the valve seat MCF as a function of engine age in days. The estimate increases sharply after 650 days, but this is based on only a small number (i.e., 10) of systems that had a total operating period exceeding 650 days. The uncertainty in the estimate for longer times is reflected in the width of the confidence intervals [the computation of the estimate beyond 389 days and the confidence limits is explained in Nelson (1995) and Chapter 16 of Meeker and Escobar (1998)].

Nonparametric Comparison of Two Samples of Repair Data. Decisions often need to be made on the basis of a comparison between two manufacturers, product designs, environments, etc. Doganaksoy and Nelson (1998) describe methods for comparing samples from two different groups of systems.

OTHER TOPICS IN RELIABILITY

Sources of Reliability Data and Information. Component and subsystem reliability information is needed as input to reliability models. Such information comes from a number of different sources. For example:

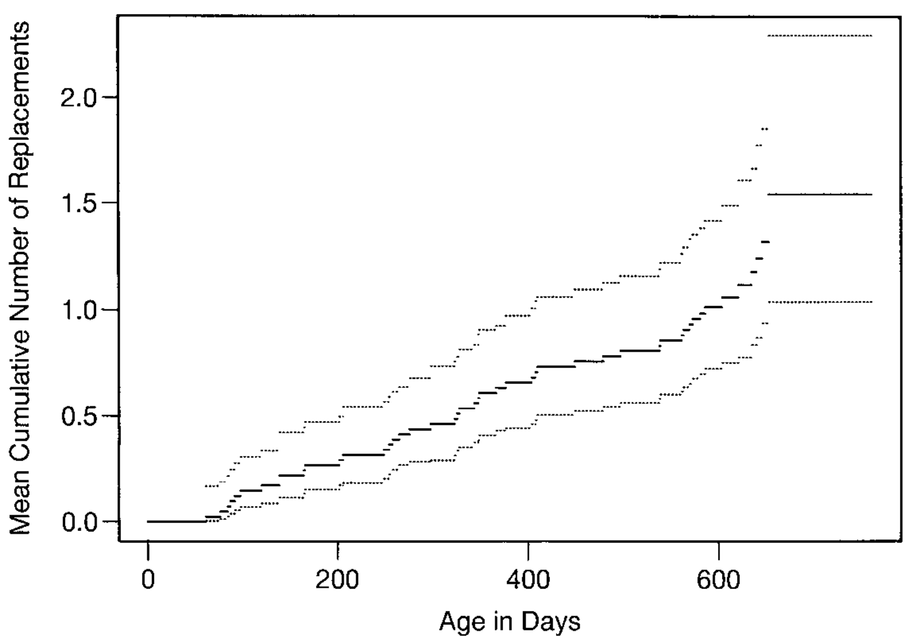

FIGURE 48.20 Estimate of the mean cumulative number of diesel engine valve seat repairs, with a 95 percent confidence interval.

- Laboratory tests are used widely, especially to test new materials and components, when there is little past experience. Such testing is generally expensive and may have limited applicability to product field reliability, but it may provide early warning of reliability problems. Care must be taken to assure that failure modes relevant to field use are obtained and that, if the goal is reliability prediction, the test conditions can be accurately related to actual field conditions. As described earlier in this section, laboratory tests are often accelerated with the goal of getting component reliability information more quickly.
- Carefully collected field data, when available, may provide the most accurate information about how components and systems behave in the field. It is important, however, to ensure that field data are consistently and unambiguously recorded to allow speedy and easy analysis. Field warranty data often contain no information on units that do not fail and may be biased, for example, because units in the harshest environments tend to fail sooner. Also, field data generally come too late to help avoid costly reliability problems.
- Reliability handbooks and data banks can be useful [e.g., Klinger, Nakada, and Menendez (1990) and MIL-HDBK 217F (1991)]. One common complaint about such handbooks, however, is that data become obsolete by the time they are published, or shortly after.
- Expert knowledge is often used when no other source of information is available.

Unless data are collected from carefully planned and conducted studies and properly maintained and analyzed, obtaining unbiased estimates, and quantifying uncertainty may be impossible.

FMEA and FMECA. Products and systems often have complicated designs that are the result of the efforts of several design teams. Management for system reliability requires a global process to assure up front that the product/system reliability will meet customer requirements.

Failure Modes and Effect Analysis (FMEA) is a systematic, structured method for identifying system failure modes and assessing the effects or consequences of the identified failure modes. Failure Modes and Effect Criticality Analysis (FMECA) considers, in addition, the importance of identified failure modes, with respect to safety, successful completion of system mission, or other criteria. The goal of FMEA/FMECA is to identify and possibly remove possible failure modes at a specified level of system architecture. These methods are typically used in product/system design review processes. The use of FMEA/FMECA usually begins in the early stages of product/system conceptualization and design. Then the FMEA/FMECA evolves over time along with changes in the product/system design and accumulation of information about product/system performance in preproduction testing and field experience. FMEA/FMECA is used during the product/system design phase to help guide decision making. It is also used to develop product/system guidelines for system repair and maintenance procedures, to improve system safety, and to provide direction for reliability improvement efforts.

Operationally, FMEA/FMECA begins by defining the scope of the analysis, specified by the system level at which failures are to be considered. FMEA/FMECA can be conducted at various levels in a product or a system. It might be done initially for individual subsystems. Then the results can be integrated to provide an FMEA/FMECA for an entire system comprising many subsystems. For example, an FMEA to study the reliability of a telecommunications relay repeater might consider, as basic components, each discrete device in the electronic circuit (e.g., ICs, capacitors, resistors, diodes). At another level, an FMEA for a large telecommunications network might consider as components all of the network nodes and node interconnections (ignoring the electronic detail within each node).

The next step in the FMEA/FMECA process is the identification of all components that may fail during the life of the product. This is followed by identification of all component interfaces or connections that might fail. In many applications, environmental and human factors–related failures are considered in defining failure modes. Finally the effects of the identified failure modes are delineated. Determining the effect of failure modes and combinations of failure modes uses the detailed specification of the relationship among the product/system components (system structure). Special worksheets and/or computer software can be used to organize all of the information.

MIL-STD-1629A (1980) and books like Høyland and Rausand (1994) and Sundararajan (1991) outline in more detail and provide examples of the procedures for performing an FMEA/FMECA analysis. Høyland and Rausand (1994) also list several computer programs designed to facilitate such analyses.

Fault Trees. The FMEA/FMECA process described above is sometimes referred to as a “bottom-up” approach to reliability modeling. Fault tree analysis, on the other hand, quantifies system failures using a “top-down” approach. First, one or more critical “top events” (such as loss of system functionality) are defined. Then in a systematic manner, the combination (or combinations) of conditions required for that event to occur are delineated. Generally this is done by identifying how failure-related events at a higher level are caused by lower-level “primary events” (e.g., failure of an individual component) and “intermediate events” (e.g., failure of a subsystem). Information from an FMEA analysis might be used as input to this step. The information is organized in the form of a “fault tree diagram” with the top event at the top of the diagram. Events at different levels of the tree are connected by logic gates defined by Boolean logic (e.g., AND, OR, Exclusive OR gates).

A complete fault tree can be used to model the probability of critical system events. Additional inputs required for this analysis include the probabilities or the conditional probabilities of the primary events. With this information and the detailed system structure specification provided by the fault tree, it is possible to compute critical event probabilities. See O’Connor (1985), Sundararajan (1991), Høyland and Rausand (1994), or Lewis (1996) for examples.

Fault tree diagrams are, in one sense, similar to the reliability block diagrams presented earlier in this subsection. It is generally possible to translate from one to the other. Fault tree analysis, however, differs in its basic approach to system reliability. Reliability block diagrams are structured around the event that the system does *not* fail. Fault tree analysis focuses on the critical top events like loss of system functionality or other safety-critical events. The tree shows, directly, the root causes of these top events, and other contributing events, at all levels within the scope of the analysis. The structure and logic of the fault tree itself provides not only a mechanism for quantitative reliability assessment, but also clearer insight into possible approaches for reliability improvement.

Designed Experiments to Improve Reliability. Experimental design can be an important tool for reliability improvement. Different kinds of experiments can obtain information for assessing and improving reliability. These include life tests and accelerated life tests, such as those described earlier in this section, to assess the durability or lifetime distribution of materials and components, for which there is one or a small number of identifiable failure modes. Information from such tests is used to make design decisions that will result in acceptable (or better) reliability for the tested components.

At another level one may conduct experiments to define system configurations or operating factors to enhance reliability by making components more robust to varying manufacturing conditions or operating/environmental conditions that might be encountered in service. Generally such experiments also focus on a single failure mode or a few failure modes for a single component or subsystem. Multifactor robust-design experiments (RDE) provide methods for systematic and efficient reliability improvement. These are often conducted on prototype units and subsystems and focus on failure modes involving interfaces and interactions among components and subsystems. Among many possible product design factors that may impact a system’s reliability, RDEs empirically identify the important ones and find levels of the product design factors that yield consistent high quality and reliability. Graves and Menten (1996) provide an excellent description of experimental strategies that can be used to help design products with higher reliability. Byrne and Quinlan (1993) describe an experiment to design a flexible push cable that will have long life under varying operating conditions. Tseng, Hamada, and Chiao (1995) describe the use of a factorial experiment to choose manufacturing variables that will extend the life of fluorescent light bulbs. Phadke (1989) describes an experiment to optimize performance, and thus reliability, of a copying machine subsystem.

The need to quickly develop new high-reliability products has motivated the development of other new-product testing methods. The purpose of these testing methods is to rapidly identify and eliminate potential reliability problems (failure modes or other system weaknesses) early in the design stage of product development. One such testing method is known as STRIFE (*str*ess-l*ife*) testing. The basic idea of STRIFE testing is to aggressively stress and test prototype or early production units (using, for example, temperature cycling, power cycling, and vibration) to force failures. Although it may be useful to test only one or two units, appropriately selected units can provide important additional information on unit-to-unit variation. Bailey and Gilbert (1981) describe an example in which the complete STRIFE test and improvement program was successfully completed in three weeks. Kececioglu and Sun (1995) also describe STRIFE testing and present several interesting case studies. Schinner (1996) provides a detailed description of preproduction "accelerated tests" that can be used to discover or assess the potential impact of product failure modes.

Reliability Demonstration. Reliability demonstration is used to assess whether product reliability meets specifications. The reliability specification is often stated in terms of the mean or a specified percentile of the life distribution of the product. For example, an industrial purchaser of a component may require that the percentage of failing parts during the first 5 years of service must not exceed 0.5 percent.

Demonstration of product reliability usually involves some type of a life test. The basic considerations in planning a reliability demonstration test are similar to those of acceptance sampling (see Section 46), although some special aspects arise as noted below. A reliability demonstration test plan specifies the number of units to be placed on test, test time, and criteria for successful demonstration. Acceptance test plans involving attribute data [such as MIL-STD-105E (1989) or the corresponding civilian standard ANSI/ASQC (1993*a*) Z1.4] are applicable sometimes to reliability demonstration when the outcome of a test is either a success or a failure, such as in demonstrating start-up reliability of a system. Likewise, standard acceptance test plans for the normal distribution [MIL-STD-414 (1957) or the corresponding civilian standard ANSI/ASQC (1993*b*) Z1.9] can sometimes be used for reliability demonstration when all units are tested to failure, provided distributional assumptions of the plan are satisfied (e.g., logs of failure times are often modeled with a normal distribution). However, special test plans for reliability demonstration are needed in situations where testing is to be terminated before all units fail. Abernethy (1996) describes small-sample Weibull demonstration plans. Special test plans for reliability are also needed to accommodate life distributions other than the normal (such as the Weibull) which are widely used in reliability data analysis.

Reliability demonstration test plans for a wide range of situations have been documented in various handbooks and standards published by the U.S. Defense Department. For example, U.S. Department of Defense (1960), *Quality Control and Reliability Handbook H-108,* provides failure-terminated (test is stopped upon occurrence of preassigned number of failures), time-terminated (test duration is predetermined) and sequential test plans for the exponential distribution. Similar test plans for the Weibull distribution are contained in U.S. Department of Defense (1965) *Handbook TR 7.* Schilling (1982) provides a thorough overview of common test plans used in reliability demonstration.

Screening and Burn-in. Many products, ranging from electronic components to electronic systems to automobiles, experience reliability problems in the early break-in part of life. Such "infant mortality" failures often occur in a small proportion of units with manufacturing defects. After all or most of these defective units have failed, population failure rates typically reach a low level until wear-out failures begin to occur (see earlier discussion of "bathtub curve"). For some products (e.g., desktop computers), the wear-out phase of life is well beyond the technological life of the product. However, in many applications involving critical systems or where repair is impossible or highly expensive (e.g., satellite systems, undersea systems, and critical aircraft systems), infant mortality failures present an important obstacle to achieving the needed level of reliability.

Manufacturers, in their quality and reliability improvement programs, strive to eliminate or at least reduce the rate of occurrence of infant mortality failures. In such applications, screening or burn-in techniques are often used to identify and remove defective components or to remove products containing defects before they are released to customers.

Burn-in tests can be viewed as a type of 100 percent inspection or screening of the product population. All units are run for a period of time before shipment or installation. To accelerate the process, components such as integrated circuits may be run at high levels of temperature and/or other stresses. The ability to use acceleration is much more limited for systems and subsystems. See Jensen and Petersen (1982) for an engineering approach to this subject. Also, see Nelson (1990, p. 43) for more information and references.

Environmental Stress Screening (ESS) was developed as a means of accelerated burn-in for units at the system or subsystem (e.g., circuit pack) level. ESS uses mild, but complicated, stressing such as combinations of temperature cycling, physical vibration, and perhaps stressful operational regimes (e.g., running computer chips at higher than usual clock speeds and lower than usual voltages) to help identify the weak units. Tustin (1990) gives a motivational description of the methodology and several references. MIL-STD-2164 (1985) provides a useful overview of ESS procedures. Kececioglu and Sun (1995) provide a comprehensive treatment of ESS. Nelson (1990, p. 39) gives additional references, including military standards.

Burn-in and ESS are inspection/screening schemes. In line with the modern quality goal of eliminating reliance on mass inspection, most manufacturers prefer not to do burn-in or ESS. Such methods are also expensive and may not be totally effective. By improving reliability through continuous improvement of the product design and the manufacturing process, it may be possible to reduce or eliminate reliance on screening tests except, perhaps, in the most critical applications.

Reliability Growth. Reliability of a system continually evolves during its design, development, testing, production, and field use. This ongoing change is referred to as “reliability growth” (although, in fact, reliability may actually improve or deteriorate). Growth usually results from efforts to discover design (or manufacturing) flaws and implement fixes to affect all future manufactured units. Thus, later-generation systems should have better reliability than their predecessors. Sometimes this process is referred to as “test, analyze, and fix” (TAAF). The basic idea is to find and fix reliability problems that had not been found earlier. Reliability growth models and data-fitting methods allow predictions of product reliability due to such improvements. Reliability growth analysis usually involves fitting a reliability growth model (such as the popular Duane plot) to failure data on systems built over time. These empirical models are then used to predict the reliability of future generations of systems. For further information and references on reliability growth, see MIL-HDBK-189 (1981), Klion (1992), Ascher and Feingold (1984), and Chapter 11 of Pecht (1995).

Software Reliability. State-of-the-art reservation, banking, billing, accounting, and other financial and business systems depend on complicated software systems. Additionally, modern hardware systems of all kinds, from automobiles and televisions to communications networks and spacecraft, contain complicated circuitry. Most of these electronic systems depend heavily on software for important functionality and flexibility. For many systems, software reliability has become an important limiting factor in system reliability.

The Institute of Electrical and Electronic Engineers defines software reliability as “the probability that software will not cause a system failure for a specified time under specified conditions.” Software reliability differs from hardware reliability in at least one important way. Hardware failures can often be traced to some combination of a physical fault and/or physical/chemical degradation that progresses over time, perhaps accelerated by stress, shocks, or other environmental or operating conditions. Software failures, on the other hand, are generally caused by inherent faults in the software that are usually present all along. Actual failure may not occur until a particular set of inputs is used or until a particular system state or level of system load is reached. The state of the software itself does not change without intervention.

Software errors differ in their criticality. Those who work with personal computers know that from time to time the system will stop functioning for unknown reasons. The cause is often software-related (i.e., it would not have occurred if the software had been designed to anticipate the conditions that caused the problem). Restarting the computer and the application will seem to make the problem disappear. Important data in the application being used at the time of the failure may or may not have been lost. Future versions of the operating system or the application software may correct such problems. In safety-critical systems (e.g., medical, air traffic control, or military systems) software failures can have much more serious (e.g., life-threatening) consequences.

For some purposes, statistical methods for software reliability are similar to those used in predicting reliability growth of a system or analyzing data from a repairable system, treated earlier in this subsection. Software data often consist of a sequence of times of failures (or some other specific event of interest) in the operation of the software system. Software reliability data are collected for various reasons, including assessment of the distribution of times between failures, tracking the effect of continuing efforts to find and correct software errors, making decisions on when to release a software product, assessing the effect of changes to improve the software development process, etc.

Numerous models have been suggested and developed to model software reliability data. The simplest of these describe the software failure rate as a smooth function of time in service and other factors, such as system load and amount of testing or use to which the system has been exposed. In an attempt to be more mechanistic and to incorporate information from the fix process directly into the software reliability model, many of these models have a parameter corresponding to the number of faults remaining in the system. In some models, the failure rate would be proportional to the number of faults. When a "repair" is made, there is some probability that the fault is fixed and, perhaps, a probability that a new fault is introduced.

For more information on software reliability and software reliability models, see Musa, Iannino, and Okumoto (1987), Shooman (1983), Neufelder (1993), Chapter 6 of Pecht (1995), or Azem (1995).

ACKNOWLEDGMENTS

Parts of this section were taken from Meeker and Escobar (1998) with permission from John Wiley & Sons, Inc. We would like to thank Elaine Miller and Denise Riker for their help in editing and typing parts of the manuscript.

REFERENCES

Abernethy, R. B. (1996). *The New Weibull Handbook.* Self-published, 536 Oyster Road, North Palm Beach, FL 33408-4328.

ANSI/ASQC (1993*a*). Z1.4-1993: *Sampling Procedures and Tables for Inspection by Attributes.* Available from American National Standards Institute, Customer Service, 11 West 42nd St., New York, NY 10036.

ANSI/ASQC (1993*b*). Z1.9-1993: *Sampling Procedures and Tables for Inspection by Variables for Percent Nonconforming.* Available from American National Standards Institute, Customer Service, 11 West 42nd St., New York, NY 10036.

Ascher, H., and Feingold, H. (1984). *Repairable Systems Reliability.* Marcel Dekker, New York.

Azem, A. (1995). *Software Reliability Determination for Conventional and Logic Programming.* Walter de Gruyter, New York.

Bailey, R. A., and Gilbert, R. A. (1981). "STRIFE Testing for Reliability Improvement." *Proceedings of the Institute of Environmental Sciences,* pp. 119–121.

Byrne, D., and Quinlan, J. (1993). "Robust Function for Attaining High Reliability at Low Cost." *1993 Proceedings Annual Reliability and Maintainability Symposium,* pp. 183–191.

Condra, L. W. (1993). *Reliability Improvement with Design of Experiments.* Marcel Dekker, New York.

Doganaksoy, N., and Nelson, W. (1998). "A Method to Compare Two Samples of Recurrence Data." *Life Data Analysis,* vol 4, pp. 51-63 .

Graves, S., and Menten, T. (1996). "Designing Experiments to Measure and Improve Reliability." *Handbook of Reliability Engineering and Management,* 2nd ed. W. G. Ireson, C. F. Coombs, and R. Y. Moss, eds. McGraw-Hill, New York, Chapter 11.

Hahn, G. J., and Shapiro, S. S. (1967). *Statistical Models in Engineering.* John Wiley & Sons, New York.

Høyland, A., and Rausand, M. (1994). *System Reliability Theory: Models and Statistics Methods.* John Wiley & Sons, New York.

Jensen, F., and Petersen, N. E. (1982). *Burn-in, An Engineering Approach to Design and Analysis of Burn-in Procedures.* John Wiley & Sons, New York.

Kececioglu, D., and Sun, F. (1995). *Environmental Stress Screening: Its Quantification, Optimization and Management.* PTR Prentice Hall, Englewood Cliffs, NJ.Klinger, D. J., Nakada, Y., and Menendez, M. A. (1990). *AT&T Reliability Manual.* Van Nostrand Reinhold, New York.

Klion, J. (1992). *Practical Electronic Reliability Engineering.* Van Nostrand Reinhold, New York.

Lawless, J. F., and Nadeau, C. (1995). "Some Simple Robust Methods for the Analysis of Recurrent Events." *Technometrics,* vol 37, pp. 158–168.

Lewis, E. E. (1996). *Introduction to Reliability Engineering.* John Wiley & Sons, New York.

Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data.* John Wiley & Sons, New York.

Meeker, W. Q., and Hahn, G. J. (1985). "How to Plan Accelerated Life Tests: Some Practical Guidelines." *ASQC Basic References in Quality Control: Statistical Techniques.* Available from the American Society for Quality, 310 W. Wisconsin Ave., Milwaukee, WI 53203, vol. 10.

Meeker, W. Q., and LuValle, M. J. (1995). "An Accelerated Life Test Model Based on Reliability Kinetics." *Technometrics,* vol. 37, pp. 133–146.

MIL-HDBK-189 (1981). *Reliability Growth Management.* Available from Naval Publications and Forms Center, 5801 Tabor Ave., Philadelphia, PA 19120.

MIL-HDBK-217F (1991). *Reliability Prediction for Electronic Equipment.* Available from Naval Publications and Forms Center, 5801 Tabor Ave., Philadelphia, PA 19120.

MIL-STD-105E (1989). *Military Standard, Sampling Procedures and Tables for Inspection by Attributes.* Available from Naval Publications and Forms Center, 5801 Tabor Ave., Philadelphia, PA 19120.

MIL-STD-414 (1957). *Military Standard, Sampling Procedures and Tables for Inspection by Variables for Percent Defective.* Available from Naval Publications and Forms Center, 5801 Tabor Ave., Philadelphia, PA 19120.

MIL-STD-1629A (1980). *Failure Modes and Effects Analysis.* Available from Naval Publications and Forms Center, 5801 Tabor Ave., Philadelphia, PA 19120.

MIL-STD-2164 (1985). *Environmental Stress Screening Process for Electronic Equipment.* Available from Naval Publications and Forms Center, 5801 Tabor Ave., Philadelphia, PA 19120.

Minitab (1997). *Version 12 User's Guide,* Minitab Inc., State College, PA.

Musa, J. D., Iannino, A., and Okumoto, K. (1987). *Software Reliability: Measurement, Prediction, Application.* McGraw-Hill, New York.

Nair, V. N. (1984). "Confidence Bands for Survival Functions with Censored Data: a Comparative Study." *Technometrics,* vol. 26, pp. 265–275.

Nelson, W. (1982). *Applied Life Data Analysis.* John Wiley & Sons, New York.

Nelson, W. (1988). "Graphical Analysis of System Repair Data." *Journal of Quality Technology,* vol. 20, pp. 24–35.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses.* John Wiley & Sons, New York.

Nelson, W. (1995). "Confidence Limits for Recurrence Data—Applied to Cost or Number of Product Repairs." *Technometrics,* vol. 37, pp. 147–157.

Nelson, W., and Doganaksoy, N. (1989). "A Computer Program for an Estimate and Confidence Limits for the Mean Cumulative Function for Cost or Number of Repairs of Repairable Products." TIS Report 89CRD239, General Electric Company Research and Development, Schenectady, NY.

Neufelder, A. (1993). *Ensuring Software Reliability.* Marcel Dekker, New York.

O'Connor, P. D. T. (1985). *Practical Reliability Engineering,* 2nd ed., John Wiley & Sons, New York.

Parida, N. (1991). "Reliability and Life Estimation from Component Fatigue Failures below the Go-No-Go Fatigue Limit." *Journal of Testing and Evaluation,* vol. 19, pp. 450–453.

Pecht, M., ed. (1995). *Product Reliability, Maintainability, and Supportability Handbook.* CRC Press, Boca Raton, FL.

Phadke, M. S. (1989). *Quality Engineering Using Robust Design.* Prentice Hall, Englewood Cliffs, NJ.

SAS Institute (1995). *JMP User's Guide,* Version 3.1. SAS Institute Inc., Cary, NC.

SAS Institute (1996). *SAS/STAT™ Software: Changes and Enhancements through Release 6.11.* SAS Institute Inc., Cary, NC. (Proc RELIABILITY is part of the SAS/QC software.)

Schilling, E. G. (1982). *Acceptance Sampling in Quality Control.* Marcel Dekker, New York.

Schinner, C. (1996). "Accelerated Testing." *Handbook of Reliability Engineering and Management,* 2nd ed., W. G. Ireson, C. F. Coombs, and R. Y. Moss, eds. McGraw-Hill, New York, Chapter 12.

Shooman, M. L. (1983). *Software Engineering: Design, Reliability, and Management.* McGraw-Hill, New York.

Statistical Sciences, (1996). *S-PLUS User's Manual,* Volumes 1, 2, Version 3.4. Statistical Sciences, Inc., Seattle.

Sundararajan, C. (1991). *Guide to Reliability Engineering.* Van Nostrand Reinhold, New York.

Tseng, T. S., Hamada, M., and Chiao, C. H. (1995). "Using Degradation Data from a Factorial Experiment to Improve Fluorescent Lamp Reliability." *Journal of Quality Technology,* vol. 27, pp. 363–369.

Tustin, W. (1990). "Shake and Bake the Bugs Out." *Quality Progress,* September 1990, pp. 61–64.

U.S. Department of Defense (1960). "Sampling Procedures and Tables for Life and Reliability Testing." *Quality Control and Reliability (Interim) Handbook (H-108).* Office of the Assistant Secretary of Defense (Supply and Logistics), Washington, DC.

U.S. Department of Defense (1965). "Factors and Procedures for Applying MIL-STD-105D Sampling Plans to Life and Reliability Testing." *Quality Control and Reliability Assurance Technical Report (TR 7).* Office of the Assistant Secretary of Defense (Installations and Logistics), Washington, DC.