product-limit estimator) of $H(t)$ without assuming a distributional form is

$$\widehat{H}(t_i) = -\sum_{j=1}^{i} \log(1 - \widehat{p}_j) \approx \sum_{j=1}^{i} \widehat{p}_j = \sum_{j=1}^{i} \frac{d_j}{n_j} = \widehat{\widehat{H}}(t_i)$$

$\widehat{\widehat{H}}(t_i)$ is known as the Nelson–Aalen estimator of $H(t_i)$. Thus $\widehat{\widehat{F}}(t_i) = 1 - \exp[-\widehat{\widehat{H}}(t_i)]$ is another nonparametric estimator for $F(t_i)$.

(a) Give conditions to assure a good agreement between $\widehat{H}(t_i)$ and $\widehat{\widehat{H}}(t_i)$ and thus between $\widehat{F}(t_i)$ and $\widehat{\widehat{F}}(t_i)$.

(b) Use the delta method to compute approximate expressions for $\text{Var}[\widehat{H}(t_i)]$ and $\text{Var}[\widehat{\widehat{H}}(t_i)]$. Comment on the expression(s) you get.

(c) Compute Nelson–Aalen estimate of $F(t)$ and compare with the estimate computed in Exercise 3.20. Describe similarities and differences.

(d) Show that $\widehat{\widehat{H}}(t_i) < \widehat{H}(t_i)$ and that $\widehat{\widehat{F}}(t_i) < \widehat{F}(t_i)$.

(e) Describe suitable modifications of the estimator that can be used when failure and censoring times are grouped into common intervals.

CHAPTER 4

Location-Scale-Based Parametric Distributions

Objectives

This chapter explains:

- Important ideas behind parametric models in the analysis of reliability data.
- Motivation for important functions of model parameters that are of interest in reliability studies.
- The location-scale family of probability distributions.
- Properties and the importance of the exponential distribution.
- Properties and the importance of log-location-scale distributions such as the Weibull, lognormal, and loglogistic distributions.
- How to generate pseudorandom data from a specified distribution (such random data are used in simulation evaluations in subsequent chapters).

Overview

This chapter introduces some basic ideas of parametric modeling and the most important parametric distributions. Parametric distributions are used extensively in subsequent chapters. Section 4.1 explains some of the basic concepts and motivation for using parametric models. Section 4.2 describes important functions of parameters like failure probabilities and distribution quantiles. Section 4.3 introduces the important location-scale family of distributions. Sections 4.4–4.11 give detailed information on these and the important log-location-scale distributions. Subsequent chapters require at least a basic understanding of the characteristics and notation for the exponential, Weibull, and lognormal distributions. Applications for the other distributions follow without difficulty. Physical motivation for these and the other distributions is helpful in practical modeling applications. Section 4.12 describes alternative choices for parameters. Section 4.13 describes methods for generating simulated values from a specified distribution. In various parts of this book we will use simulation to develop

4.1 INTRODUCTION

As we saw in Chapter 3, it is possible to make certain kinds of inferences without having to assume a particular parametric form for a failure-time distribution. There are, however, many problems in reliability data analysis where it is either useful or essential to use a parametric distribution form. This chapter describes a number of simple probability distributions that are commonly used to model failure-time processes. Chapter 5 does the same for other important and useful, but more complicated, distributions. The discussion in these chapters concerns underlying continuous-time models, although much of the material also holds for discrete-time models.

As explained in Chapter 2, a natural model for a continuous random variable, say, T, is the cumulative distribution function (cdf). Specific examples given in this chapter and in Chapter 5 are of the form $\Pr(T \leq t) = F(t; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters. In this book, we use T to denote positive random variables like failure time, so that $T > 0$; correspondingly, we will use Y to denote unrestricted random variables so that $-\infty < Y = \log(T) < \infty$. Unlike the "basic parameters" in $\boldsymbol{\pi}$ and $\boldsymbol{p}$ used in the "nonparametric" formulation in Chapters 2 and 3, the parametric models described in this chapter will have a $\boldsymbol{\theta}$ containing a small fixed number of parameters. The most commonly used parametric probability distributions have between one and four parameters, although there are some distributions with more than four parameters. More complicated models could contain many more parameters involving mixtures, competing failure modes, or other combinations of distributions or models that include explanatory variables. One simple example that we will use later in this chapter is the exponential distribution for which

$$\Pr(T \leq t) = F(t; \theta) = 1 - \exp\left(-\frac{t}{\theta}\right), \quad t > 0 \tag{4.1}$$

where θ is the single scalar parameter of the distribution (equal to the mean or first moment, in this example).

Use of parametric distributions complements nonparametric techniques and provides the following advantages:

- Parametric models can be described concisely with just a few parameters, instead of having to report an entire curve.
- It is possible to use a parametric model to extrapolate (in time) to the lower or upper tail of a distribution.
- Parametric models provide smooth estimates of failure-time distributions.

In practice it is often useful to do various parametric and nonparametric analyses of a data set.

4.2 QUANTITIES OF INTEREST IN RELIABILITY APPLICATIONS

Starting in Chapter 7, we will focus on the problem of *estimating* the parameters $\boldsymbol{\theta}$ and important functions of $\boldsymbol{\theta}$. In this section we describe ideas behind parameterization of a probability distribution and describe a number of particular functions of parameters that are of interest for reliability analysis.

In most practical problems, interest centers on quantities that are functions of $\boldsymbol{\theta}$ and the ML estimates of these functions will *not* depend on the particular parameterization that is used to specify the parametric model. The quantities of interest discussed here extend the list introduced in Chapter 2, and now these quantities will be expressed as functions of the small set of parameters $\boldsymbol{\theta}$. Specifically, for distributions of positive and continuous random variables (there are similar definitions for discrete and/or nonpositive random variables):

- The "probability of failure" $p = \Pr(T \leq t) = F(t; \boldsymbol{\theta})$ by a specified t. For example, if T is the time of failure of a unit, then p is the probability that the unit will fail before t.
- The "p quantile" of the distribution of T is the smallest value t such that $F(t; \boldsymbol{\theta}) \geq p$. We will express the p quantile as $t_p = F^{-1}(p; \boldsymbol{\theta})$. For the failure-time example, t_p is the time at which $100p\%$ of the units in the product population will have failed. The median is equal to $t_{.5}$.
- The "hazard function" (hf) is defined as

$$h(t) = \frac{f(t; \boldsymbol{\theta})}{1 - F(t; \boldsymbol{\theta})}. \tag{4.2}$$

  As described in Section 2.1.1, the hazard function is of particular interest in reliability applications because it indicates, for surviving units, the propensity to fail in the following small interval of time, as a function of age.
- The mean life (also known as the "average," "expectation," or "first moment") of T

$$\mathrm{E}(T) = \int_0^{\infty} t f(t; \boldsymbol{\theta})\, dt = \int_0^{\infty} [1 - F(t; \boldsymbol{\theta})]\, dt \tag{4.3}$$

  is a measure of the center of $f(t; \boldsymbol{\theta})$. When $f(t; \boldsymbol{\theta})$ is highly skewed, the mean may differ appreciably from other measures of central tendency like the median. The mean is sometimes, but not always, one of the distribution parameters. For some pdfs, the value of the integral will be infinite. Then it is said that the mean of T "does not exist." When T is time to failure, the mean is sometimes referred to as the MTTF, for mean time to failure.
- The variance (also known as the "second central moment") of T

$$\mathrm{Var}(T) = \int_0^{\infty} [t - \mathrm{E}(T)]^2 f(t; \boldsymbol{\theta})\, dt$$

  is a measure of spread of the distribution of T. Var(T) is the average squared deviation of T from its mean. Again, if the value of the integral is infinite, it is said that the variance of T "does not exist." The quantity $\mathrm{SD}(T) = \sqrt{\mathrm{Var}(T)}$, known as the "standard deviation" of T, is easier to interpret because it has the same units as T.

- The unitless quantity $\gamma_2 = \mathrm{SD}(T)/\mathrm{E}(T)$, known as the "coefficient of variation" of T, is useful for comparing the relative amount of variability in different distributions. The quantity $1/\gamma_2 = \mathrm{E}(T)/\mathrm{SD}(T)$ is sometimes known as the "signal-to-noise ratio."
- The unitless quantity

$$\gamma_3 = \frac{\int_0^\infty [t - \mathrm{E}(T)]^3 f(t; \boldsymbol{\theta})\, dt}{[\mathrm{Var}(T)]^{3/2}},$$

known as the "standardized third central moment" or "coefficient of skewness" of T, is a measure of the skewness in the distribution of T. When a distribution is symmetric, $\gamma_3 = 0$. It is, however, possible to have $\gamma_3 = 0$ for a distribution that is not perfectly symmetric (e.g., the Weibull distribution, discussed in in Section 4.8, has $\gamma_3 = 0$ when $\beta = 3.602$, but the distribution is only approximately symmetric). Usually, however, when γ_3 is positive (negative), the distribution of T is skewed to the right (left).

For reliability applications, quantiles, failure probabilities, and the hazard function are typically of higher interest than distribution moments. In subsequent chapters we will describe *point estimation* and, at the same time, emphasize methods of obtaining *confidence intervals* (for scalars) and *confidence regions* (for simultaneous inference on a vector of two or more quantities) for parameters and important functions of parameters. Confidence intervals and regions quantify the uncertainty in parameter estimates arising from the fact that inferences are generally based on only a finite number of observations from the process or population of interest.

4.3 LOCATION-SCALE AND LOG-LOCATION-SCALE DISTRIBUTIONS

A random variable Y belongs to the location-scale family of distributions if its cdf can be expressed as

$$\Pr(Y \le y) = F(y; \mu, \sigma) = \Phi\left(\frac{y - \mu}{\sigma}\right),$$

where Φ does not depend on any unknown parameters. In this case we say that $-\infty < \mu < \infty$ is a location parameter and that $\sigma > 0$ is a scale parameter. Substitution shows that Φ is the cdf of Y when $\mu = 0$ and $\sigma = 1$. Also, Φ is the cdf of $(Y - \mu)/\sigma$. Location-scale distributions are important for a number of reasons including:

- Many of the widely used statistical distributions are either location-scale distributions or closely related. These distributions include the exponential, normal, Weibull, lognormal, loglogistic, logistic, and extreme value distributions.
- Methods of data analysis and inference, statistical theory, and computer software developed for the location-scale family can be applied to any of the members of the family.
- Theory for location-scale distributions is relatively simple.

In cases where Φ does depend on one or more unknown parameters (as with a number of the distributions described in Chapter 5), Y is not a member of the location-scale family, but the location-scale structure and notation will still be useful for us.

A random variable T belongs to the log-location-scale family distribution if $Y = \log(T)$ is a member of the location-scale family. The Weibull, lognormal, and loglogistic distributions are the most important members of this family.

4.4 EXPONENTIAL DISTRIBUTION

When T has an exponential distribution, we indicate this by $T \sim \mathrm{EXP}(\theta, \gamma)$. The two-parameter exponential distribution (to distinguish it from the more commonly used one-parameter exponential distribution) has cdf, pdf, and hf

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

$$f(t; \theta, \gamma) = \frac{1}{\theta}\exp\left(-\frac{t - \gamma}{\theta}\right),$$

$$h(t; \theta, \gamma) = \frac{1}{\theta}, \quad t > \gamma,$$

where $\theta > 0$ is a scale parameter and γ is both a location and a threshold parameter. For $\gamma = 0$ this is the well-known one-parameter exponential distribution (and often known simply as the exponential distribution). When T has this simpler distribution, we indicate it by $T \sim \mathrm{EXP}(\theta)$. The cdf, pdf, and hf are graphed in Figure 4.1 for $\theta = .5, 1$, and 2 and $\gamma = 0$.

For integer $m > 0$, $\mathrm{E}[(T - \gamma)^m] = m!\, \theta^m$. Thus the mean and variance of the exponential distribution are, respectively, $\mathrm{E}(T) = \gamma + \theta$ and $\mathrm{Var}(T) = \theta^2$. The p quantile of the exponential distribution is $t_p = \gamma - \log(1 - p)\, \theta$.

The one-parameter exponential distribution, where $\gamma = 0$, is the simplest distribution that is commonly used in the analysis of reliability data. The exponential distribution has the important characteristic that its hf is constant (does not depend on time t). A constant hf implies that, for an unfailed unit, the probability of failing in the next small interval of time is independent of the unit's age. Physically, a constant hf suggests that the population of units under consideration is not wearing out or otherwise aging. The exponential distribution is a popular distribution for some kinds of electronic components (e.g., capacitors or robust, high-quality integrated circuits). This exponential distribution would *not* be appropriate for a population of electronic components having failure-causing quality defects (such defects are difficult to rule out completely and are a leading cause of electronic system reliability problems). On the other hand, the exponential distribution might be useful to describe failure times for components that exhibit physical wearout if the wearout does not show up until long after the expected technological life of the system in which the compo-

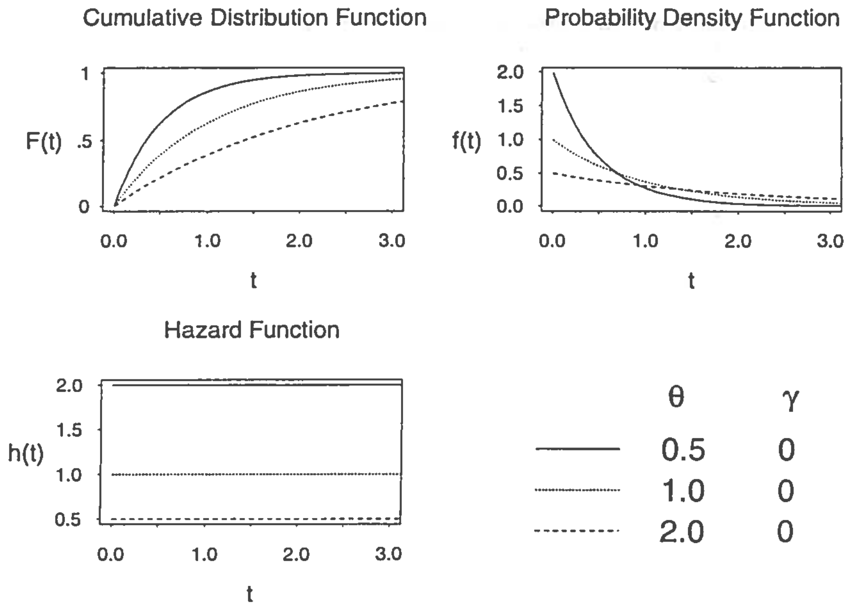


Figure 4.1. Exponential cdf, pdf, and hf for $\theta = .5$, 1, and 2 and $\gamma = 0$.

nent would be installed (e.g., electronic components in computing equipment having failures caused by random external events).

Under very special circumstances, the exponential distribution may be appropriate for the times between system failures, arrivals in a queue, and other interarrival time distributions. Specifically, the exponential distribution is the distribution of interval times of a homogeneous Poisson process. See Chapter 3 of Thompson (1988) and Chapter 16 for more information on homogeneous Poisson processes.

The exponential distribution is usually *inappropriate* for modeling the life of mechanical components (e.g., bearings) subject to some combination of fatigue, corrosion, or wear. It is also usually inappropriate for electronic components that exhibit wearout properties over their technological life (e.g., lasers and filament devices). A distribution with an increasing hf is, in such applications, usually more appropriate. Similarly, for populations containing mixtures of good and bad units the population hf may decrease with life because, as the bad units fail and leave the population, only the stronger units are left.

4.5 NORMAL DISTRIBUTION

When Y has a normal distribution, we indicate this by $Y \sim \text{NOR}(\mu, \sigma)$. The normal distribution is a location-scale distribution with cdf and pdf

$$F(y; \mu, \sigma) = \Phi_{\text{nor}}\left(\frac{y-\mu}{\sigma}\right),$$

$$f(y; \mu, \sigma) = \frac{1}{\sigma}\phi_{\text{nor}}\left(\frac{y-\mu}{\sigma}\right), \quad -\infty < y < \infty,$$

where $\phi_{\text{nor}}(z) = (1/\sqrt{2\pi})\exp(-z^2/2)$ and $\Phi_{\text{nor}}(z) = \int_{-\infty}^{z} \phi_{\text{nor}}(w)\,dw$ are, respectively, the pdf and cdf for the standardized NOR($\mu = 0, \sigma = 1$) distribution. Here $-\infty < \mu < \infty$ is a location parameter and $\sigma > 0$ is a scale parameter. When there is no useful simplification of the hf definition in (4.2), as with the normal distribution, the definition will not be repeated. The normal distribution pdf, cdf, and hf are graphed in Figure 4.2 for $\mu = 5$ and $\sigma = .3, .5, .8$.

For integer $m > 0$, $\text{E}[(Y-\mu)^m] = 0$ if m is odd and $\text{E}[(Y-\mu)^m] = m!\sigma^m / [2^{m/2}(m/2)!]$ if m is even. From this, the mean and variance of the normal distribution are, respectively, $\text{E}(Y) = \mu$ and $\text{Var}(Y) = \sigma^2$. The p quantile of the normal distribution is $y_p = \mu + \Phi_{\text{nor}}^{-1}(p)\sigma$, where $\Phi_{\text{nor}}^{-1}(p) = z_p$ is the p quantile of the standard normal distribution.

As a model for variability, the normal distribution has a long history of use in many areas of application. This is due to the simplicity of normal distribution theory and the central limit theorem. The central limit theorem states that the distribution of the sum of a large number of independent identically distributed random quantities

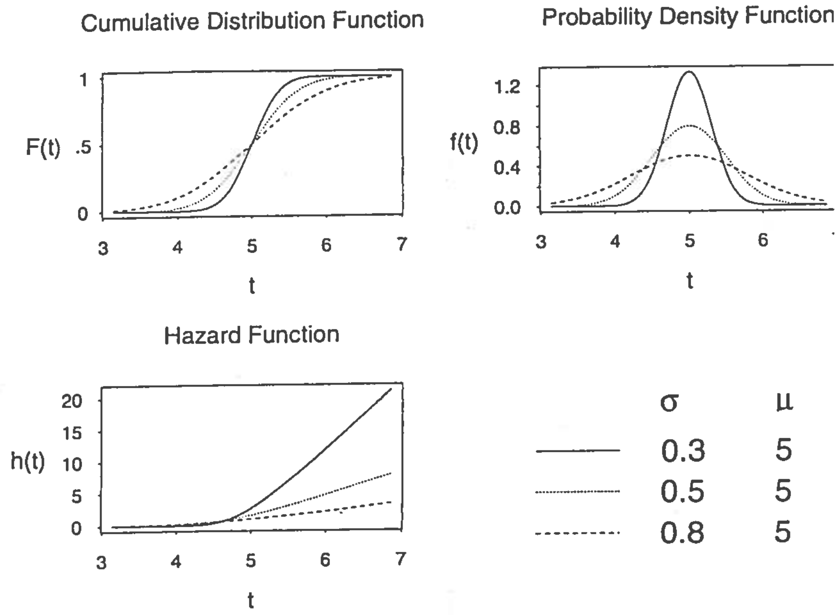


Figure 4.2. Normal cdf, pdf, and hf with location parameter (mean) $\mu = 5$ and scale parameter (standard deviation) $\sigma = .3, .5$, and .8.

has, approximately, a normal distribution. In reliability data analysis, the use of the normal distribution is, however, less common. As seen from Figure 4.2, the normal distribution has an increasing hf that begins to increase rapidly near, but before, the point of median life. The normal distribution has proved to be a useful distribution for certain life data when $\mu > 0$ and the coefficient of variation (σ/μ) is small. Examples include electric filament devices (e.g., incandescent light bulbs and toaster heating elements) and strength of wire bonds in integrated circuits (component strength is often used as an easy-to-obtain surrogate measure or indicator of eventual reliability). Also, as described in Section 4.6, the normal distribution is often a useful model for the logarithms of failure times (see the next section).

4.6 LOGNORMAL DISTRIBUTION

When T has a lognormal distribution, we indicate this by $T \sim \text{LOGNOR}(\mu, \sigma)$. If $T \sim \text{LOGNOR}(\mu, \sigma)$ then $Y = \log(T) \sim \text{NOR}(\mu, \sigma)$. The lognormal cdf and pdf are

$$F(t; \mu, \sigma) = \Phi_{\text{nor}}\left[\frac{\log(t) - \mu}{\sigma}\right], \tag{4.4}$$

$$f(t; \mu, \sigma) = \frac{1}{\sigma t}\phi_{\text{nor}}\left[\frac{\log(t) - \mu}{\sigma}\right], \quad t > 0, \tag{4.5}$$

where ϕ_{nor} and Φ_{nor} are pdf and cdf for the standardized normal. The median $t_{.5} = \exp(\mu)$ is a scale parameter and $\sigma > 0$ is a shape parameter. The lognormal cdf, pdf, and hf are graphed in Figure 4.3 for $\sigma = .3, .5$, and $.8$ and $\mu = 0$, corresponding to the median $t_{.5} = \exp(\mu) = 1$.

The most common definition of the lognormal distribution uses base e (natural) logarithms. Base 10 (common) logarithms are also used in some areas of application. Bottom-line answers for important reliability metrics (e.g., estimates of failure probabilities, failure rates, and quantiles) will not depend on the base that is used. The definition of the parameters μ (mean of the *logarithm* of T) and σ (standard deviation of the *logarithm* of T) will, however, depend on the base that is used. For this reason it is important to make consistent use of one particular base. In this book we will generally use base e (natural) logarithms for the lognormal distribution definition.

For integer $m > 0$, $\text{E}(T^m) = \exp(m\mu + m^2\sigma^2/2)$. From this it follows that the mean and variance of the lognormal distribution are, respectively, $\text{E}(T) = \exp(\mu + .5\sigma^2)$ and $\text{Var}(T) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$. The quantile function of the lognormal distribution is $t_p = \exp[\mu + \Phi^{-1}_{\text{nor}}(p)\sigma]$.

The lognormal distribution is a common model for failure times. Following from the central limit theorem (mentioned in Section 4.5), 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 quantities. 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

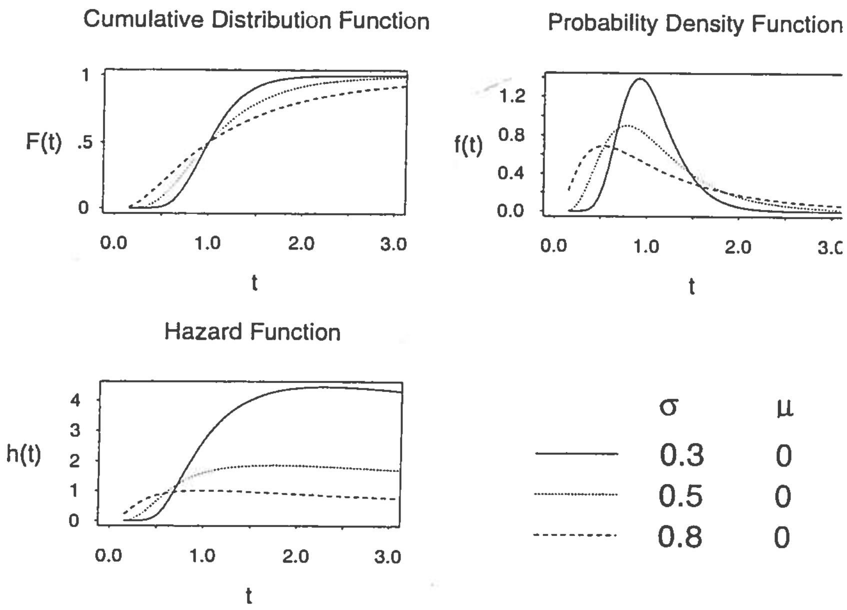


Figure 4.3. Lognormal cdf, pdf, and hf for scale parameter $t_{.5} = \exp(\mu) = 1$ and for shape parameter $\sigma = .3, .5$, and $.8$.

multiplicatively (e.g., see the models in Chapter 13). The lognormal distribution is widely used to describe time to fracture from fatigue crack growth in metals. As shown in Figure 4.3 (also see Exercise 4.19), the lognormal $h(t)$ starts at 0, increases to a point in time, and then decreases eventually to zero. For large σ, $h(t)$ reaches a maximum early in life and then decreases. For this reason, the lognormal distribution is often used as a model for a population of electronic components that exhibits a decreasing hf. It has been suggested that early-life "hardening" of certain kinds of materials or components might lead to such an hf. The lognormal distribution also arises as the time to failure distribution of certain degradation processes, as described in Chapter 13. The lognormal distribution described in this section is sometimes referred to as the "two-parameter lognormal distribution" to distinguish it from the three-parameter lognormal distribution described in Section 5.10.2.

4.7 SMALLEST EXTREME VALUE DISTRIBUTION

When the random variable Y has a smallest extreme value distribution, we indicate this by $Y \sim \text{SEV}(\mu, \sigma)$. The SEV cdf, pdf, and hf are

$$F(y; \mu, \sigma) = \Phi_{\text{sev}}\left(\frac{y - \mu}{\sigma}\right),$$

$$f(y;\mu,\sigma) = \frac{1}{\sigma}\phi_{\mathrm{sev}}\left(\frac{y-\mu}{\sigma}\right),$$

$$h(y;\mu,\sigma) = \frac{1}{\sigma}\exp\left(\frac{y-\mu}{\sigma}\right), \qquad -\infty < y < \infty,$$

where $\Phi_{\mathrm{sev}}(z) = 1 - \exp[-\exp(z)]$ and $\phi_{\mathrm{sev}}(z) = \exp[z - \exp(z)]$ are the cdf and pdf, respectively, for standardized SEV ($\mu = 0, \sigma = 1$). Here $-\infty < \mu < \infty$ is the location parameter and $\sigma > 0$ is the scale parameter. The SEV cdf, pdf, and hf are graphed in Figure 4.4 for $\mu = 50$ and $\sigma = 5, 6$, and 7.

The mean, variance, and quantile functions of the smallest extreme value distribution are $\mathrm{E}(Y) = \mu - \sigma\gamma$, $\mathrm{Var}(Y) = \sigma^2\pi^2/6$, and $y_p = \mu + \Phi_{\mathrm{sev}}^{-1}(p)\,\sigma$, where $\Phi_{\mathrm{sev}}^{-1}(p) = \log[-\log(1-p)]$ and $\gamma \approx .5772$ is Euler's constant.

Figure 4.4 shows that the smallest extreme value distribution pdf is skewed to the left. Although most failure-time distributions are skewed to the right, distributions of strength will sometimes be skewed to the left (because of a few weak units in the lower tail of the distribution, but a sharper upper bound for the majority of units in the upper tail of the strength population). The SEV distribution may have physical justification arising from an extreme value theorem. Namely, it is the limiting standardized distribution of the minimum of a large number of random variables from a certain class of distributions (this class includes the normal distribution as a special case). If σ is small relative to μ the SEV distribution can be used as a life

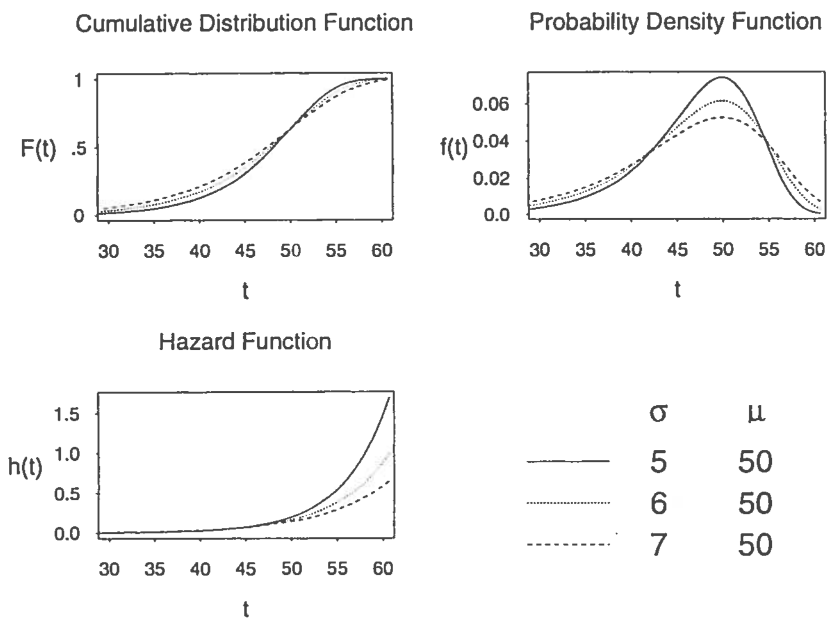


Figure 4.4. Smallest extreme value cdf, pdf, and hf with $\mu = 50$ and $\sigma = 5, 6$, and 7.

distribution. The exponentially increasing hf suggests that the SEV would be suitable for modeling the life of a product that experiences very rapid wearout after a certain age. The distributions of logarithms of failure times can often be modeled with the SEV distribution; see Section 4.8. Also see the closely related Gompertz–Makeham distribution in Section 5.8.

4.8 WEIBULL DISTRIBUTION

The Weibull distribution cdf is often written as

$$\Pr(T \le t;\eta,\beta) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \qquad t > 0. \tag{4.6}$$

For this parameterization, $\beta > 0$ is a shape parameter and $\eta > 0$ is a scale parameter as well as the .632 quantile. The practical value of the Weibull distribution stems from its ability to describe failure distributions with many different commonly occurring shapes. As illustrated in Figure 4.5, for $0 < \beta < 1$, the Weibull has a decreasing hf. With $\beta > 1$, the Weibull has an increasing hf.

For integer $m > 0$, $\mathrm{E}(T^m) = \eta^m\Gamma(1 + m/\beta)$, where $\Gamma(\kappa) = \int_0^\infty z^{\kappa-1}\exp(-z)\,dz$ is the gamma function. From this it follows that the mean and variance of the Weibull

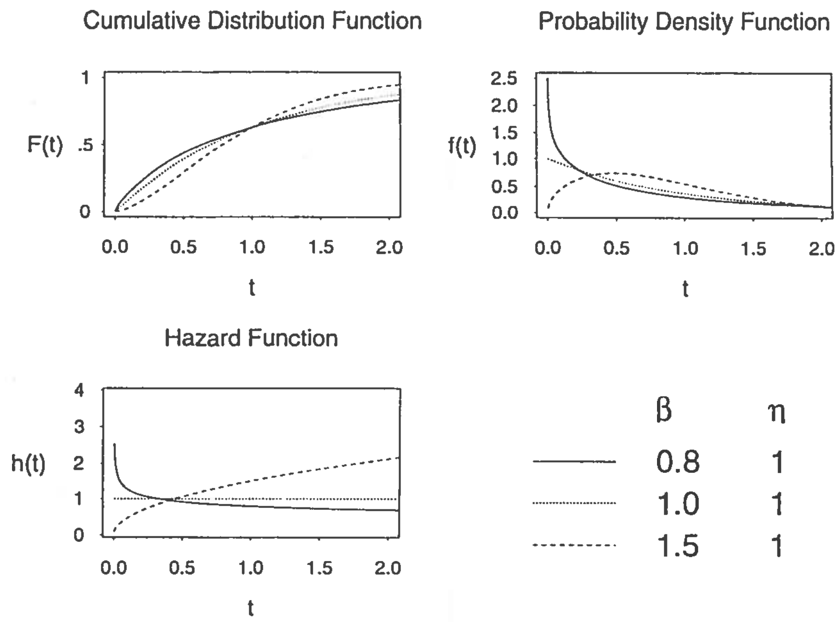


Figure 4.5. Weibull cdf, pdf, and hf for $t_{.632} = \eta = \exp(\mu) = 1$ and $\beta = 1/\sigma = .8, 1$, and 1.5.

distribution are, respectively, $E(T) = \eta\Gamma(1 + 1/\beta)$ and $Var(T) = \eta^2[\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)]$. The Weibull p quantile is $t_p = \eta[-\log(1 - p)]^{1/\beta}$. Note that when $\beta = 1$, the cdf in (4.6) reduces to an exponential distribution with scale parameter $\theta = \eta$.

It is convenient to use a simple alternative parameterization for the Weibull distribution. This alternative parameterization is based on the relationship between the Weibull distribution and the smallest extreme value distribution described in Section 4.7. In particular, if T has a Weibull distribution, then $Y = \log(T) \sim SEV(\mu, \sigma)$, where $\sigma = 1/\beta$ is the scale parameter and $\mu = \log(\eta)$ is the location parameter. Thus when T has a Weibull distribution, we indicate this by $T \sim WEIB(\mu, \sigma)$. In this form, the Weibull cdf, pdf, and hf can be written as

$$F(t; \mu, \sigma) = \Phi_{sev}\left[\frac{\log(t) - \mu}{\sigma}\right], \qquad (4.7)$$

$$f(t; \mu, \sigma) = \frac{1}{\sigma t}\phi_{sev}\left[\frac{\log(t) - \mu}{\sigma}\right] = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right],$$

$$h(t; \mu, \sigma) = \frac{1}{\sigma\exp(\mu)}\left[\frac{t}{\exp(\mu)}\right]^{1/\sigma-1} = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}, \quad t > 0.$$

Then the Weibull p quantile is $t_p = \exp[\mu + \Phi_{sev}^{-1}(p)\,\sigma]$. The Weibull/SEV relationship parallels the lognormal/normal relationship. The SEV parameterization is useful because location-scale distributions are easier to work with in general. As mentioned in Section 4.3, transforming the Weibull distribution into an SEV distribution allows the use of general results for location-scale distributions, which apply directly to all such distributions, including the Weibull, lognormal, and some other distributions.

The theory of extreme values shows that the Weibull distribution can be used to model the minimum of a large number of independent positive random variables from a certain class of distributions. Thus extreme value theory also suggests that the Weibull distribution may be suitable. The more common justification for its use is empirical: the Weibull distribution can be used to model failure-time data with decreasing or increasing hf. The Weibull distribution described in this section is sometimes referred to as the "two-parameter Weibull distribution" to distinguish it from the three-parameter Weibull distribution described in Section 5.10.2.

4.9 LARGEST EXTREME VALUE DISTRIBUTION

When Y has a largest extreme value distribution, we indicate this by $Y \sim LEV(\mu, \sigma)$. The largest extreme value distribution cdf, pdf, and hf are

$$F(y; \mu, \sigma) = \Phi_{lev}\left(\frac{y - \mu}{\sigma}\right),$$

$$f(y; \mu, \sigma) = \frac{1}{\sigma}\phi_{lev}\left(\frac{y - \mu}{\sigma}\right),$$

$$h(y; \mu, \sigma) = \frac{\exp\left(-\frac{y-\mu}{\sigma}\right)}{\sigma\left\{\exp\left[\exp\left(-\frac{y-\mu}{\sigma}\right)\right] - 1\right\}}, \quad -\infty < y < \infty,$$

where $\Phi_{lev}(z) = \exp[-\exp(-z)]$ and $\phi_{lev}(z) = \exp[-z - \exp(-z)]$ are cdf and pdf for the standardized $LEV(\mu = 0, \sigma = 1)$ distribution. Here $-\infty < \mu < \infty$ is a location parameter and $\sigma > 0$ is a scale parameter. The LEV cdf, pdf, and hf are graphed in Figure 4.6 for $\mu = 10$ and $\sigma = 5, 6$, and 7.

The mean, variance, and quantile functions of the largest extreme value distribution are $E(Y) = \mu + \sigma\gamma$, $Var(Y) = \sigma^2\pi^2/6$, and $y_p = \mu + \Phi_{lev}^{-1}(p)\,\sigma$, where $\Phi_{lev}^{-1}(p) = -\log[-\log(p)]$. Note the close relationship between LEV and SEV: if $Y \sim LEV(\mu, \sigma)$ then $-Y \sim SEV(-\mu, \sigma)$ and $\Phi_{lev}^{-1}(p) = -\Phi_{sev}^{-1}(1 - p)$.

The theory of extreme values shows that the LEV distribution can be used to model the maximum of a large number of random variables from a certain class of distributions (which includes the normal distribution). As shown in Figure 4.6, the largest extreme value pdf is skewed to the right. The LEV hf always increases but is bounded in the sense that $\lim_{t\to\infty} h(t; \mu, \sigma) = 1/\sigma$. Although most failure-time distributions are skewed to the right, the LEV distribution is not commonly used as a model for failure times. This is because the LEV distribution (like the SEV and normal

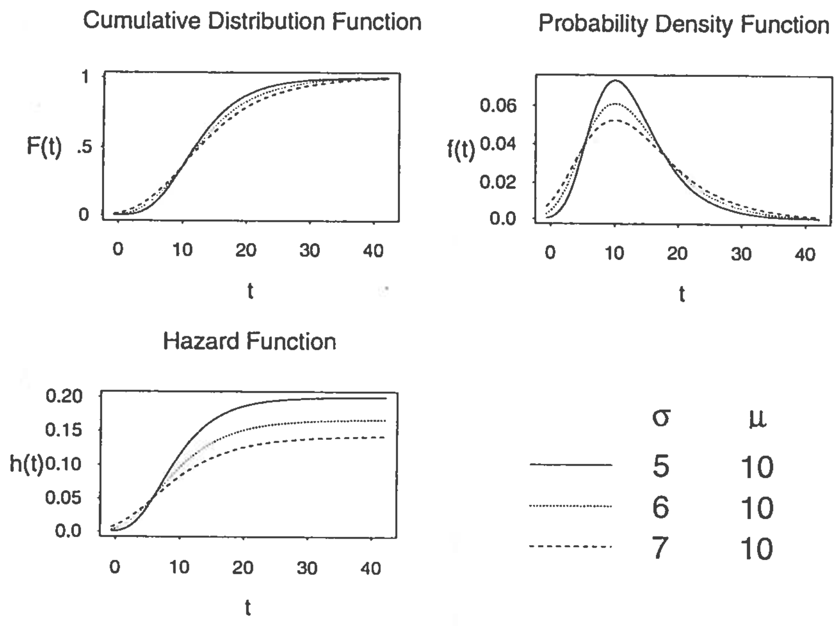


Figure 4.6. Largest extreme value cdf, pdf, and hf with $\mu = 10$ and $\sigma = 5, 6$, and 7.

distributions) has positive probability of negative observations and there are a number of other right-skewed distributions that do not have this property. Nevertheless, the LEV distribution could be used as a model for life if σ is small relative to $\mu > 0$.

4.10 LOGISTIC DISTRIBUTION

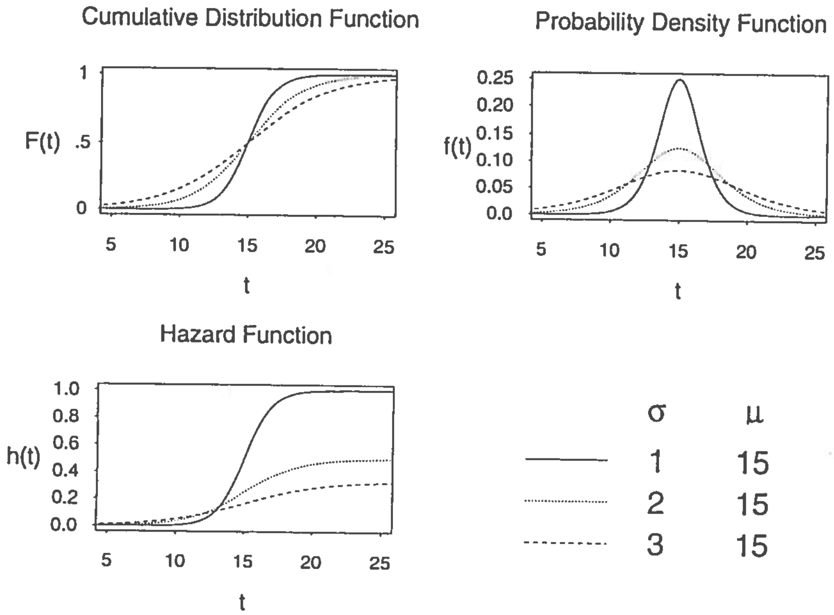
When Y has a logistic distribution, we indicate this by $Y \sim \text{LOGIS}(\mu, \sigma)$. The logistic distribution is a location-scale distribution with cdf, pdf, and hf

$$F(y; \mu, \sigma) = \Phi_{\text{logis}}\left(\frac{y-\mu}{\sigma}\right),$$

$$f(y; \mu, \sigma) = \frac{1}{\sigma}\phi_{\text{logis}}\left(\frac{y-\mu}{\sigma}\right),$$

$$h(y; \mu, \sigma) = \frac{1}{\sigma}\Phi_{\text{logis}}\left(\frac{y-\mu}{\sigma}\right), \quad -\infty < y < \infty,$$

where $\Phi_{\text{logis}}(z) = \exp(z)/[1+\exp(z)]$ and $\phi_{\text{logis}}(z) = \exp(z)/[1+\exp(z)]^2$ are the cdf and pdf, respectively, for a standardized LOGIS($\mu = 0, \sigma = 1$). Here $-\infty < \mu < \infty$ is a location parameter and $\sigma > 0$ is a scale parameter. The logistic cdf, pdf, and hf are graphed in Figure 4.7 for location parameter $\mu = 15$ and scale parameter $\sigma = 1$, 2, and 3.



For integer $m > 0$, $\text{E}[(Y-\mu)^m] = 0$ if m is odd, and $\text{E}[(Y-\mu)^m] = 2\sigma^m(m!)[1-(1/2)^{m-1}]\sum_{i=1}^{\infty}(1/i)^m$ if m is even. From this $\text{E}(Y) = \mu$ and $\text{Var}(Y) = \sigma^2\pi^2/3$. The p quantile is $y_p = \mu + \Phi_{\text{logis}}^{-1}(p)\sigma$, where $\Phi_{\text{logis}}^{-1}(p) = \log[p/(1-p)]$ is the p quantile of the standard logistic distribution.

The shape of the logistic distribution is very similar to that of the normal distribution; the logistic distribution has slightly "longer tails." In fact, it would require an extremely large number of observations to assess whether data come from a normal or logistic distribution. The main difference between the distributions is in the behavior of the hf in the upper tail of the distribution, where the logistic hf levels off, approaching $1/\sigma$ for large y. For some purposes, the logistic distribution has been preferred to the normal distribution because its cdf can be written in a simple closed form. With modern software, however, it is not any more difficult to compute probabilities from a normal cdf.

4.11 LOGLOGISTIC DISTRIBUTION

When T has a loglogistic distribution, we indicate this by $T \sim \text{LOGLOGIS}(\mu, \sigma)$. If $T \sim \text{LOGLOGIS}(\mu, \sigma)$ then $Y = \log(T) \sim \text{LOGIS}(\mu, \sigma)$. The loglogistic cdf, pdf, and hf are

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

$$f(t; \mu, \sigma) = \frac{1}{\sigma t}\phi_{\text{logis}}\left[\frac{\log(t)-\mu}{\sigma}\right],$$

$$h(t; \mu, \sigma) = \frac{1}{\sigma t}\Phi_{\text{logis}}\left[\frac{\log(t)-\mu}{\sigma}\right], \quad t > 0,$$

where ϕ_{logis} and Φ_{logis} are the pdf and cdf, respectively, for a standardized LOGIS, defined in Section 4.10. The median $t_{.5} = \exp(\mu)$ is a scale parameter and $\sigma > 0$ is a shape parameter. The LOGLOGIS cdf, pdf and hf are graphed in Figure 4.8 for scale parameter $\exp(\mu) = 1$ and $\sigma = .2, .4$, and .6.

For integer $m > 0$, $\text{E}(T^m) = \exp(m\mu)\Gamma(1+m\sigma)\Gamma(1-m\sigma)$, where $\Gamma(x)$ is the gamma function. From this $\text{E}(T) = \exp(\mu)\Gamma(1+\sigma)\Gamma(1-\sigma)$ and $\text{Var}(T) = \exp(2\mu)[\Gamma(1+2\sigma)\Gamma(1-2\sigma) - \Gamma^2(1+\sigma)\Gamma^2(1-\sigma)]$. Note that for values of $\sigma \geq 1$, the mean of T does not exist and for $\sigma \geq 1/2$, the variance of T does not exist. The p quantile function is $t_p = \exp[\mu + \Phi_{\text{logis}}^{-1}(p)\sigma]$, where $\Phi_{\text{logis}}^{-1}(p)$ is defined in Section 4.10.

Corresponding to the similarity between the logistic and normal distributions,

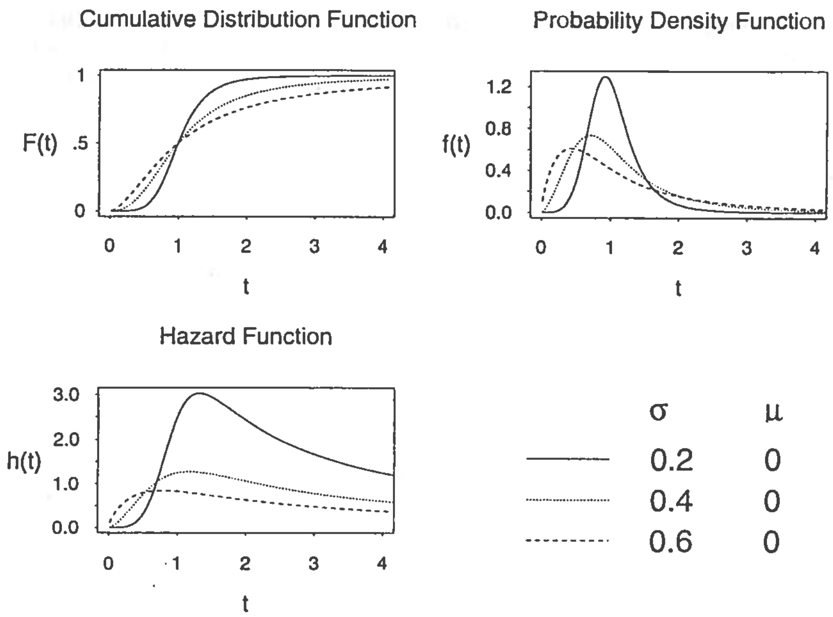


Figure 4.8. Loglogistic cdf, pdf, and hf for $t_{.5} = \exp(\mu) = 1$ and $\sigma = .2, .4,$ and $.6$.

4.12 PARAMETERS AND PARAMETERIZATION

The choice of $\boldsymbol{\theta}$, a set of parameters (the values of which are usually unknown) to describe a particular model, is somewhat arbitrary and may depend on tradition, on physical interpretation, or on having a model parameterization with desirable computational properties for estimating parameters. For example, the exponential distribution can be written in terms of its mean θ, as in (4.1), or its constant hazard $\lambda = 1/\theta$. The μ, σ notation for the Weibull distribution allows us to see connections with other location-scale-based distributions. The traditional parameters of a normal distribution are $\theta_1 = \mu$ and $\theta_2 = \sigma > 0$, the mean and standard deviation, respectively. An alternative with no restrictions on the range of the parameters would be $\theta_1 = \mu$ and $\theta_2 = \log(\sigma)$. Another parameterization, which may have better numerical properties for estimation with heavily censored data sets, is $\theta_1 = \mu + z_p\sigma$ and $\theta_2 = \log(\sigma)$, where z_p is the p quantile of the standard normal distribution. The best value of p to use depends on the amount of censoring. In particular, if the sample contains failure times with no censoring, choose $p = .5$ with $z_p = 0$ because then $\widehat{\theta}_1$ (the maximum likelihood estimate of the mean) and $\widehat{\theta}_2$ (the maximum likelihood estimate of the log standard deviation) would be statistically independent (this is a well-known result from statistical theory). Exercise 8.20 explores this issue more thoroughly.

4.13 GENERATING PSEUDORANDOM OBSERVATIONS FROM A SPECIFIED DISTRIBUTION

Simulation (or Monte Carlo simulation) methods are becoming increasingly important for many applications of statistics and, indeed, quantitative analysis in general. In particular, it is possible to determine, through simulation, numerical quantities that are difficult or impossible to compute by purely analytical means. This book uses a simulation approach in a number of methods, examples, and exercises. A pseudorandom number generator is the basic building block of any simulation application. This section will show some simple methods for generating pseudorandom numbers from specified probability distributions. The bibliographic notes at the end of this chapter give references for more technical details and more advanced methods of generating pseudorandom numbers from specified distributions.

4.13.1 Uniform Pseudorandom Number Generator

Most computers, data analysis software, and spreadsheets provide a pseudorandom number generator for the uniform distribution on (0, 1) [denoted by UNIF(0, 1)]. This distribution has its probability distributed uniformly from (0, 1). The cdf and pdf of the UNIF(0, 1) distribution are $F_U(u) = u$ and $f_U(u) = 1, \quad 0 < u < 1$. Pseudorandom numbers from the UNIF(0, 1) distribution can be used easily to generate random numbers from other distributions, both discrete and continuous.

4.13.2 Pseudorandom Observations from Continuous Distributions

Suppose $U_1, \ldots, U_n$ is a pseudorandom sample from a UNIF(0, 1). Then if $t_p = F_T^{-1}(p)$ is the quantile function for the distribution of the random variable T from which a sample of pseudorandom numbers is desired, $T_1 = F_T^{-1}(U_1), \ldots, T_n = F_T^{-1}(U_n)$ is a pseudorandom sample from F_T. For example, to generate a pseudorandom sample from the Weibull distribution for specified parameters η and β, first obtain the UNIF(0, 1) pseudorandom sample $U_1, \ldots, U_n$ and then compute $T_1 = \eta[-\log(1 - U_1)]^{1/\beta}, \ldots, T_n = \eta[-\log(1 - U_n)]^{1/\beta}$. Similarly, for the lognormal distribution the pseudorandom sample can be obtained from $T_1 = \exp[\mu + \Phi_{\text{nor}}^{-1}(U_1)\sigma], \ldots, T_n = \exp[\mu + \Phi_{\text{nor}}^{-1}(U_n)\sigma]$.

4.13.3 Efficient Generation of Censored Pseudorandom Samples

This section shows how to generate pseudorandom *censored* samples from a specified cdf $F(t; \boldsymbol{\theta})$. Such samples are useful for implementing simulations like those used throughout the book and for bootstrap methods like those described in Chapter 9.

General Approach

Let $U_{(i)}$ denote the ith order statistic from a random sample of size n from a UNIF(0, 1) distribution. Using the properties of order statistics, the conditional distribution of