Chapter 14*

LIFE TESTING

14.1 INTRODUCTION

In this chapter, we consider a population of items having lifetimes that are assumed to be independent random variables with a common distribution that is specified up to an unknown parameter. The problem of interest will be to use whatever data are available to estimate this parameter.

In Section 14.2, we introduce the concept of the hazard (or failure) rate function — a useful engineering concept that can be utilized to specify lifetime distributions. In Section 14.3, we suppose that the underlying life distribution is exponential and show how to obtain estimates (point, interval, and Bayesian) of its mean under a variety of sampling plans. In Section 14.4, we develop a test of the hypothesis that two exponentially distributed populations have a common mean. In Section 14.5, we consider two approaches to estimating the parameters of a Weibull distribution.

14.2 HAZARD RATE FUNCTIONS

Consider a positive continuous random variable X, that we interpret as being the lifetime of some item, having distribution function F and density f. The *hazard rate* (sometimes called the *failure rate*) function $\lambda(t)$ of F is defined by

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

To interpret $\lambda(t)$, suppose that the item has survived for t hours and we desire the probability that it will not survive for an additional time dt. That is, consider

* Optional chapter.

$P\{X \in (t, t+dt) \mid X > t\}$. Now

$$
\begin{aligned}
P\{X \in (t, t+dt) | X > t\} &= \frac{P\{X \in (t, t+dt), X > t\}}{P\{X > t\}} \\
&= \frac{P\{X \in (t, t+dt)\}}{P\{X > t\}} \\
&\approx \frac{f(t)}{1 - F(t)} dt
\end{aligned}
$$

That is, $\lambda(t)$ represents the conditional probability intensity that an item of age t will fail in the next moment.

Suppose now that the lifetime distribution is exponential. Then, by the memoryless property of the exponential distribution it follows that the distribution of remaining life for a t-year-old item is the same as for a new item. Hence $\lambda(t)$ should be constant, which is verified as follows:

$$
\begin{aligned}
\lambda(t) &= \frac{f(t)}{1 - F(t)} \\
&= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \\
&= \lambda
\end{aligned}
$$

Thus, the failure rate function for the exponential distribution is constant. The parameter λ is often referred to as the *rate* of the distribution.

We now show that the failure rate function $\lambda(t), t \geq 0$, uniquely determines the distribution F. To show this, note that by definition

$$
\begin{aligned}
\lambda(s) &= \frac{f(s)}{1 - F(s)} \\
&= \frac{\frac{d}{ds} F(s)}{1 - F(s)} \\
&= \frac{d}{ds}\{-\log[1 - F(s)]\}
\end{aligned}
$$

Integrating both sides of this equation from 0 to t yields

$$
\begin{aligned}
\int_0^t \lambda(s)\, ds &= -\log[1 - F(t)] + \log[1 - F(0)] \\
&= -\log[1 - F(t)] \quad \text{since } F(0) = 0
\end{aligned}
$$

which implies that

$$1 - F(t) = \exp\left\{-\int_0^t \lambda(s)\, ds\right\} \tag{14.2.1}$$

Hence a distribution function of a positive continuous random variable can be specified by giving its hazard rate function. For instance, if a random variable has a linear hazard rate function — that is, if

$$\lambda(t) = a + bt$$

then its distribution function is given by

$$F(t) = 1 - e^{-at - bt^2/2}$$

and differentiation yields that its density is

$$f(t) = (a + bt)e^{-(at + bt^2/2)}, \qquad t \geq 0$$

When $a = 0$, the foregoing is known as the *Rayleigh density function*.

EXAMPLE 14.2a One often hears that the death rate of a person who smokes is, at each age, twice that of a nonsmoker. What does this mean? Does it mean that a nonsmoker has twice the probability of surviving a given number of years as does a smoker of the same age?

SOLUTION If $\lambda_s(t)$ denotes the hazard rate of a smoker of age t and $\lambda_n(t)$ that of a nonsmoker of age t, then the foregoing is equivalent to the statement that

$$\lambda_s(t) = 2\lambda_n(t)$$

The probability that an A-year-old nonsmoker will survive until age $B, A < B$, is

$$\begin{aligned}
&P\{A\text{-year-old nonsmoker reaches age } B\} \\
&\quad = P\{\text{nonsmoker's lifetime} > B \mid \text{nonsmoker's lifetime} > A\} \\
&\quad = \frac{1 - F_{\text{non}}(B)}{1 - F_{\text{non}}(A)} \\
&\quad = \frac{\exp\left\{-\int_0^B \lambda_n(t)dt\right\}}{\exp\left\{-\int_0^A \lambda_n(t)dt\right\}} \qquad \text{from Equation 14.2.1} \\
&\quad = \exp\left\{-\int_A^B \lambda_n(t)dt\right\}
\end{aligned}$$

whereas the corresponding probability for a smoker is, by the same reasoning,

$$\begin{aligned} P\{A\text{-year-old smoker reaches age } B\} &= \exp\left\{-\int_A^B \lambda_s(t)\,dt\right\} \\ &= \exp\left\{-2\int_A^B \lambda_n(t)\,dt\right\} \\ &= \left[\exp\left\{-\int_A^B \lambda_n(t)\,dt\right\}\right]^2 \end{aligned}$$

In other words, of two individuals of the same age, one of whom is a smoker and the other a nonsmoker, the probability that the smoker survives to any given age is the *square* (not one-half) of the corresponding probability for a nonsmoker. For instance, if $\lambda_n(t) = 1/20$, $50 \le t \le 60$, then the probability that a 50-year-old nonsmoker reaches age 60 is $e^{-1/2} = .607$, whereas the corresponding probability for a smoker is $e^{-1} = .368$. ■

REMARK ON TERMINOLOGY

We will say that X has failure rate function $\lambda(t)$ when more precisely we mean that the distribution function of X has failure rate function $\lambda(t)$.

14.3 THE EXPONENTIAL DISTRIBUTION IN LIFE TESTING

14.3.1 Simultaneous Testing — Stopping at the rth Failure

Suppose that we are testing items whose life distribution is exponential with unknown mean θ. We put n independent items simultaneously on test and stop the experiment when there have been a total of r, $r \le n$, failures. The problem is to then use the observed data to estimate the mean θ.

The observed data will be the following:

$$\textit{Data:} \qquad x_1 \le x_2 \le \cdots \le x_r, \qquad i_1, i_2, \ldots, i_r \tag{14.3.1}$$

with the interpretation that the jth item to fail was item i_j and it failed at time x_j. Thus, if we let X_i, $i = 1, \ldots, n$ denote the lifetime of component i, then the data will be as given in Equation 14.3.1 if

$$X_{i_1} = x_1, X_{i_2} = x_2, \ldots, X_{i_r} = x_r$$
$$\text{other } n - r \text{ of the } X_j \text{ are all greater than } x_r$$

Now the probability density of X_{i_j} is

$$f_{X_{i_j}}(x_j) = \frac{1}{\theta} e^{-x_j/\theta}, \quad j = 1, \ldots, r$$

and so, by independence, the joint probability density of $X_{i_j}, j = 1, \ldots, r$ is

$$f_{X_{i_1},\ldots,X_{i_r}}(x_1, \ldots, x_r) = \prod_{j=1}^{r} \frac{1}{\theta} e^{-x_j/\theta}$$

Also, the probability that the other $n - r$ of the X's are all greater than x_r is, again using independence,

$$P\{X_j > x_r \text{ for } j \neq i_1 \text{ or } i_2 \ldots \text{or } i_r\} = (e^{-x_r/\theta})^{n-r}$$

Hence, we see that the *likelihood* of the observed data — call it $L(x_1, \ldots, x_r, i_1, \ldots, i_r)$ — is, for $x_1 \leq x_2 \leq \cdots \leq x_r$,

$$\begin{aligned} &L(x_1, \ldots, x_r, i_1, \ldots, i_r) \\ &\quad = f_{X_{i_1}, X_{i_2},\ldots,X_{i_r}}(x_1, \ldots, x_r) P\{X_j > x_r, j \neq i_1, \ldots, i_r\} \\ &\quad = \frac{1}{\theta} e^{-x_1/\theta} \cdots \frac{1}{\theta} e^{-x_r/\theta} (e^{-x_r/\theta})^{n-r} \\ &\quad = \frac{1}{\theta^r} \exp\left\{ -\frac{\sum_{i=1}^{r} x_i}{\theta} - \frac{(n-r)x_r}{\theta} \right\} \end{aligned} \tag{14.3.2}$$

REMARK

The likelihood in Equation 14.3.2 not only specifies that the first r failures occur at times $x_1 \leq x_2 \leq \cdots \leq x_r$ but also that the r items to fail were, in order, $i_1, i_2, \ldots, i_r$. If we only desired the density function of the first r failure times, then since there are $n(n-1)\cdots(n-(r-1)) = n!/(n-r)!$ possible (ordered) choices of the first r items to fail, it follows that the joint density is, for $x_1 \leq x_2 \leq \cdots \leq x_r$,

$$f(x_1, x_2, \ldots, x_r) = \frac{n!}{(n-r)!} \frac{1}{\theta^r} \exp\left\{ -\frac{\sum_{i=1}^{r} x_i}{\theta} - \frac{(n-r)}{\theta} x_r \right\}$$

To obtain the maximum likelihood estimator of θ, we take the logarithm of both sides of Equation 14.3.2. This yields

$$\log L(x_1, \ldots, x_r, i_1, \ldots, i_r) = -r \log \theta - \frac{\sum_{i=1}^{r} x_i}{\theta} - \frac{(n-r)x_r}{\theta}$$

and so

$$\frac{\partial}{\partial \theta} \log L(x_1, \ldots, x_r, i_1, \ldots, i_r) = -\frac{r}{\theta} + \frac{\sum_{i=1}^{r} x_i}{\theta^2} + \frac{(n-r)x_r}{\theta^2}$$

Equating to 0 and solving yields that $\hat{\theta}$, the maximum likelihood estimate, is given by

$$\hat{\theta} = \frac{\sum_{i=1}^{r} x_i + (n-r)x_r}{r}$$

Hence, if we let $X_{(i)}$ denote the time at which the ith failure occurs ($X_{(i)}$ is called the ith *order statistic*), then the maximum likelihood estimator of θ is

$$\begin{aligned} \hat{\theta} &= \frac{\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}}{r} \\ &= \frac{\tau}{r} \end{aligned} \tag{14.3.3}$$

where τ, defined to equal the numerator in Equation 14.3.3, is called the *total-time-on-test statistic*. We call it this since the ith item to fail functions for a time $X_{(i)}$ (and then fails), $i = 1, \ldots, r$, whereas the other $n - r$ items function throughout the test (which lasts for a time $X_{(r)}$). Hence the sum of the times that all the items are on test is equal to τ.

To obtain a confidence interval for θ, we will determine the distribution of τ, the total time on test. Recalling that $X_{(i)}$ is the time of the ith failure, $i = 1, \ldots, r$, we will start by rewriting the expression for τ. To write an expression for τ, rather than summing the total time on test of each of the items, let us ask how much additional time on test was generated between each successive failure. That is, let us denote by Y_i, $i = 1, \ldots, r$, the additional time on test generated between the $(i-1)$st and ith failure. Now up to the first $X_{(1)}$ time units (as all n items are functioning throughout this interval), the total time on test is

$$Y_1 = nX_{(1)}$$

Between the first and second failures, there are a total of $n - 1$ functioning items, and so

$$Y_2 = (n-1)(X_{(2)} - X_{(1)})$$

In general, we have

$$
\begin{aligned}
Y_1 &= nX_{(1)} \\
Y_2 &= (n-1)(X_{(2)} - X_{(1)}) \\
&\vdots \\
Y_j &= (n-j+1)(X_{(j)} - X_{(j-1)}) \\
&\vdots \\
Y_r &= (n-r+1)(X_{(r)} - X_{(r-1)})
\end{aligned}
$$

and

$$\tau = \sum_{j=1}^{r} Y_j$$

The importance of the foregoing representation for τ follows from the fact that the distributions of the Y_j's are easily obtained as follows. Since $X_{(1)}$, the time of the first failure, is the minimum of n independent exponential lifetimes, each having rate $1/\theta$, it follows from Proposition 5.6.1 that it is itself exponentially distributed with rate n/θ. That is, $X_{(1)}$ is exponential with mean θ/n, and so $nX_{(1)}$ is exponential with mean θ. Also, at the moment when the first failure occurs, the remaining $n-1$ functioning items are, by the memoryless property of the exponential, as good as new and so each will have an additional life that is exponential with mean θ; hence, the additional time until one of them fails is exponential with rate $(n-1)/\theta$. That is, independent of $X_{(1)}, X_{(2)} - X_{(1)}$ is exponential with mean $\theta/(n-1)$ and so $Y_2 = (n-1)(X_{(2)} - X_{(1)})$ is exponential with mean θ. Indeed, continuing this argument leads us to the following conclusion:

$$
\begin{aligned}
&Y_1, \ldots, Y_r \text{ are independent exponential} \\
&\text{random variables each having mean } \theta
\end{aligned}
\tag{14.3.4}
$$

Hence, since the sum of independent and identically distributed exponential random variables has a gamma distribution (Corollary 5.7.2), we see that

$$\tau \sim \text{gamma}(r, 1/\theta)$$

That is, τ has a gamma distribution with parameters r and $1/\theta$. Equivalently, by recalling that a gamma random variable with parameters $(r, 1/\theta)$ is equivalent to $\theta/2$ times a chi-square random variable with $2r$ degrees of freedom (see Section 5.8.1), we obtain that

$$\frac{2\tau}{\theta} \sim \chi^2_{2r} \tag{14.3.5}$$

That is, $2\tau/\theta$ has a chi-square distribution with $2r$ degrees of freedom. Hence,

$$P\{\chi^2_{1-\alpha/2,2r} < 2\tau/\theta < \chi^2_{\alpha/2,2r}\} = 1 - \alpha$$

and so a $100(1 - \alpha)$ percent confidence interval for θ is

$$\theta \in \left(\frac{2\tau}{\chi^2_{\alpha/2,2r}}, \frac{2\tau}{\chi^2_{1-\alpha/2,2r}}\right) \tag{14.3.6}$$

One-sided confidence intervals can be similarly obtained.

EXAMPLE 14.3a A sample of 50 transistors is simultaneously put on a test that is to be ended when the 15th failure occurs. If the total time on test of all transistors is equal to 525 hours, determine a 95 percent confidence interval for the mean lifetime of a transistor. Assume that the underlying distribution is exponential.

SOLUTION From Program 5.8.1b,

$$\chi^2_{.025,30} = 46.98, \quad \chi^2_{.975,30} = 16.89$$

and so, using Equation 14.3.6, we can assert with 95 percent confidence that

$$\theta \in (22.35, 62.17) \quad \blacksquare$$

In testing a hypothesis about θ, we can use Equation 14.3.6 to determine the p-value of the test data. For instance, suppose we are interested in the one-sided test of

$$H_0 : \theta \geq \theta_0$$

versus the alternative

$$H_1 : \theta < \theta_0$$

This can be tested by first computing the value of the test statistic $2\tau/\theta_0$ — call this value v — and then computing the probability that a chi-square random variable with $2r$ degrees of freedom would be as small as v. This probability is the p-value in the sense that it represents the (maximal) probability that such a small value of $2\tau/\theta_0$ would have been observed if H_0 were true. The hypothesis should then be rejected at all significance levels at least as large as this p-value.

EXAMPLE 14.3b A producer of batteries claims that the lifetimes of the items it manufactures are exponentially distributed with a mean life of at least 150 hours. To test this claim, 100 batteries are simultaneously put on a test that is slated to end when the 20th failure occurs. If, at the end of the experiment, the total test time of all the 100 batteries is equal to 1,800, should the manufacturer's claim be accepted?

SOLUTION Since $2\tau/\theta_0 = 3{,}600/150 = 24$, the p-value is

$$\begin{aligned} p\text{-value} &= P\{\chi^2_{40} \leq 24\} \\ &= .021 \quad \text{from Program 5.8.1a} \end{aligned}$$

Hence, the manufacturer's claim should be rejected at the 5 percent level of significance (indeed at any significance level at least as large as .021). ■

It follows from Equation 14.3.5 that the accuracy of the estimator τ/r depends only on r and not on n, the number of items put on test. The importance of n resides in the fact that by choosing it large enough we can ensure that the test is, with high probability, of short duration. In fact, the moments of $X_{(r)}$, the time at which the test ends, are easily obtained. Since, with $X_{(0)} \equiv 0$,

$$X_{(j)} - X_{(j-1)} = \frac{Y_j}{n-j+1}, \quad j = 1, \ldots, r$$

it follows upon summing that

$$X_{(r)} = \sum_{j=1}^{r} \frac{Y_j}{n-j+1}$$

Hence, from Equation 14.3.4, $X_{(r)}$ is the sum of r independent exponentials having respective means $\theta/n, \theta/(n-1), \ldots, \theta/(n-r+1)$. Using this, we see that

$$\begin{aligned} E[X_{(r)}] &= \sum_{j=1}^{r} \frac{\theta}{n-j+1} = \theta \sum_{j=n-r+1}^{n} \frac{1}{j} \qquad (14.3.7) \\ \text{Var}(X_{(r)}) &= \sum_{j=1}^{r} \left(\frac{\theta}{n-j+1}\right)^2 = \theta^2 \sum_{j=n-r+1}^{n} \frac{1}{j^2} \end{aligned}$$

where the second equality uses the fact that the variance of an exponential is equal to the square of its mean. For large n, we can approximate the preceding sums as follows:

$$\sum_{j=n-r+1}^{n} \frac{1}{j} \approx \int_{n-r+1}^{n} \frac{dx}{x} = \log\left(\frac{n}{n-r+1}\right)$$

$$\sum_{j=n-r+1}^{n} \frac{1}{j^2} \approx \int_{n-r+1}^{n} \frac{dx}{x^2} = \frac{1}{n-r+1} - \frac{1}{n} = \frac{r-1}{n(n-r+1)}$$

Thus, for instance, if in Example 14.3b the true mean life was 120 hours, then the expectation and variance of the length of the test are approximately given by

$$E[X_{(20)}] \approx 120 \log\left(\frac{100}{81}\right) = 25.29$$
$$\operatorname{Var}(X_{(20)}) \approx (120)^2 \frac{19}{100(81)} = 33.78$$

14.3.2 Sequential Testing

Suppose now that we have an infinite supply of items, each of whose lifetime is exponential with an unknown mean θ, which are to be tested sequentially, in that the first item is put on test and on its failure the second is put on test, and so on. That is, as soon as an item fails, it is immediately replaced on life test by the next item. We suppose that at some fixed time T the text ends.

The observed data will consist of the following:

$$\textit{Data}: \quad r, x_1, x_2, \ldots, x_r$$

with the interpretation that there has been a total of r failures with the ith item on test having functioned for a time x_i. Now the foregoing will be the observed data if

$$X_i = x_i, \quad i = 1, \ldots, r, \quad \sum_{i=1}^{r} x_i < T \tag{14.3.8}$$
$$X_{r+1} > T - \sum_{i=1}^{r} x_i$$

where X_i is the functional lifetime of the ith item to be put in use. This follows since in order for there to be r failures, the rth failure must occur before time T — and so $\sum_{i=1}^{r} X_i < T$ — and the functional life of the $(r+1)$st item must exceed $T - \sum_{i=1}^{r} X_i$ (see Figure 14.1).

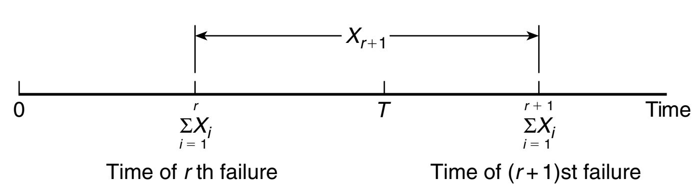


FIGURE 14.1 *r failures by time T.*

From Equation 14.3.8, we obtain that the likelihood of the data $r, x_1, \ldots, x_r$ is as follows:

$$\begin{aligned}
&f(r, x_1, \ldots, x_r|\theta) \\
&\quad = f_{X_1,\ldots,X_r}(x_1, \ldots, x_r)P\left\{X_{r+1} > T - \sum_{i=1}^{r} x_i\right\}, \qquad \sum_{i=1}^{r} x_i < T \\
&\quad = \frac{1}{\theta^r} e^{-\Sigma_{i=1}^r x_i/\theta} e^{-(T-\Sigma_{i=1}^r x_i)/\theta} \\
&\quad = \frac{1}{\theta^r} e^{-T/\theta}
\end{aligned}$$

Therefore,

$$\log f(r, x_1, \ldots, x_r|\theta) = -r\log\theta - \frac{T}{\theta}$$

and so

$$\frac{\partial}{\partial\theta}\log f(r, x_1, \ldots, x_r|\theta) = -\frac{r}{\theta} + \frac{T}{\theta^2}$$

On equating to 0 and solving, we obtain that the maximum likelihood estimate for θ is

$$\hat{\theta} = \frac{T}{r}$$

Since T is the total time on test of all items, it follows once again that the maximum likelihood estimate of the unknown exponential mean is equal to the total time on test divided by the number of observed failures in this time.

If we let $N(T)$ denote the number of failures by time T, then the maximum likelihood estimator of θ is $T/N(T)$. Suppose now that the observed value of $N(T)$ is $N(T) = r$. To determine a $100(1-\alpha)$ percent confidence interval estimate for θ, we will first determine the values θ_L and θ_U, which are such that

$$P_{\theta_U}\{N(T) \geq r\} = \frac{\alpha}{2}, \quad P_{\theta_L}\{N(T) \leq r\} = \frac{\alpha}{2}$$

where by $P_\theta(A)$ we mean that we are computing the probability of the event A under the supposition that θ is the true mean. The $100(1-\alpha)$ percent confidence interval estimate for θ is

$$\theta \in (\theta_L, \theta_U)$$

To understand why those values of θ for which either $\theta < \theta_L$ or $\theta > \theta_U$ are not included in the confidence interval, note that $P_\theta\{N(T) \geq r\}$ decreases and $P_\theta\{N(T) \leq r\}$

increases in θ (why?). Hence,

$$\text{if } \theta < \theta_L, \quad \text{then } P_\theta\{N(T) \le r\} < P_{\theta_L}\{N(T) \le r\} = \frac{\alpha}{2}$$
$$\text{if } \theta > \theta_U, \quad \text{then } P_\theta\{N(T) \ge r\} < P_{\theta_U}\{N(T) \ge r\} = \frac{\alpha}{2}$$

It remains to determine θ_L and θ_U. To do so, note first that the event that $N(T) \ge r$ is equivalent to the statement that the rth failure occurs before or at time T. That is,

$$N(T) \ge r \Leftrightarrow X_1 + \cdots + X_r \le T$$

and so

$$\begin{aligned}
P_\theta\{N(T) \ge r\} &= P_\theta\{X_1 + \cdots + X_r \le T\} \\
&= P\{\Gamma(r, 1/\theta) \le T\} \\
&= P\left\{\frac{\theta}{2}\chi^2_{2r} \le T\right\} \\
&= P\left\{\chi^2_{2r} \le 2T/\theta\right\}
\end{aligned}$$

Hence, upon evaluating the foregoing at $\theta = \theta_U$, and using the fact that $P\{\chi^2_{2r} \le \chi^2_{1-\alpha/2,2r}\} = \alpha/2$, we obtain that

$$\frac{\alpha}{2} = P\left\{\chi^2_{2r} \le \frac{2T}{\theta_U}\right\}$$

and that

$$\frac{2T}{\theta_U} = \chi^2_{1-\alpha/2,2r}$$

or

$$\theta_U = 2T/\chi^2_{1-\alpha/2,2r}$$

Similarly, we can show that

$$\theta_L = 2T/\chi^2_{\alpha/2,2r}$$

and thus the $100(1-\alpha)$ percent confidence interval estimate for θ is

$$\theta \in (2T/\chi^2_{\alpha/2,2r}, \quad 2T/\chi^2_{1-\alpha/2,2r})$$

EXAMPLE 14.3c If a one-at-a-time sequential test yields 10 failures in the fixed time of $T = 500$ hours, then the maximum likelihood estimate of θ is $500/10 = 50$ hours. A 95 percent confidence interval estimate of θ is

$$0 \in (1{,}000/\chi^2_{.025,20},\ 1{,}000/\chi^2_{.975,20})$$

Running Program 5.8.1b yields that

$$\chi^2_{.025,20} = 34.17, \quad \chi^2_{.975,20} = 9.66$$

and so, with 95 percent confidence,

$$\theta \in (29.27, 103.52) \quad ■$$

If we wanted to test the hypothesis

$$H_0 : \theta = \theta_0$$

versus the alternative

$$H_1 : \theta \neq \theta_0$$

then we would first determine the value of $N(T)$. If $N(T) = r$, then the hypothesis would be rejected provided either

$$P_{\theta_0}\{N(T) \leq r\} \leq \frac{\alpha}{2} \quad \text{or} \quad P_{\theta_0}\{N(T) \geq r\} \leq \frac{\alpha}{2}$$

In other words, H_0 would be rejected at all significance levels greater than or equal to the p-value given by

$$\begin{aligned}
p\text{-value} &= 2\ \min(P_{\theta_0}\{N(T) \geq r\}, P_{\theta_0}\{N(T) \leq r\}) \\
p\text{-value} &= 2\ \min(P_{\theta_0}\{N(T) \geq r\}, 1 - P_{\theta_0}\{N(T) \geq r+1\}) \\
&= 2\ \min\left(P\left\{\chi^2_{2r} \leq \frac{2T}{\theta_0}\right\}, 1 - P\left\{\chi^2_{2(r+1)} \leq \frac{2T}{\theta_0}\right\}\right)
\end{aligned}$$

The p-value for a one-sided test is similarly obtained.

The chi-square probabilities in the foregoing can be computed by making use of Program 5.8.1a.

EXAMPLE 14.3d A company claims that the mean lifetimes of the semiconductors it produces is at least 25 hours. To substantiate this claim, an independent testing service has decided to sequentially test, one at a time, the company's semiconductors for 600 hours. If 30 semiconductors failed during this period, what can we say about the validity of the company's claim? Test at the 10 percent level.

SOLUTION This is a one-sided test of

$$H_0 : \theta \geq 25 \quad \text{versus} \quad H_1 : \theta < 25$$

The relevant probability for determining the p-value is the probability that there would have been as many as 30 failures if the mean life were 25. That is,

$$\begin{aligned} p\text{-value} &= P_{25}\{N(600) \geq 30\} \\ &= P\{\chi^2_{60} \leq 1{,}200/25\} \\ &= .132 \quad \text{from Program 5.8.1a} \end{aligned}$$

Thus, H_0 would be accepted when the significance level is .10. ■

14.3.3 Simultaneous Testing — Stopping by a Fixed Time

Suppose again that we are testing items whose life distributions are independent exponential random variables with a common unknown mean θ. As in Section 14.3.1, the n items are simultaneously put on test, but now we suppose that the test is to stop either at some fixed time T or whenever all n items have failed — whichever occurs first. The problem is to use the observed data to estimate θ.

The observed data will be as follows:

$$Data: \quad i_1, i_2, \ldots, i_r, \quad x_1, x_2, \ldots, x_r$$

with the interpretation that the preceding results when the r items numbered $i_1, \ldots, i_r$ are observed to fail at respective times $x_1, \ldots, x_r$ and the other $n - r$ items have not failed by time T.

Since an item will not have failed by time T if and only if its lifetime is greater than T, we see that the likelihood of the foregoing data is

$$\begin{aligned} f(i_1, \ldots, i_r, x_1, \ldots, x_r) &= f_{X_{i_1}, \ldots, X_{i_r}}(x_1, \ldots, x_r) P\{X_j > T, j \neq i_1, \ldots, i_r\} \\ &= \frac{1}{\theta} e^{-x_1/\theta} \cdots \frac{1}{\theta} e^{-x_r/\theta} (e^{-T/\theta})^{n-r} \\ &= \frac{1}{\theta^r} \exp\left\{ -\frac{\sum_{i=1}^{r} x_i}{\theta} - \frac{(n-r)T}{\theta} \right\} \end{aligned}$$

To obtain the maximum likelihood estimates, take logs to obtain

$$\log f(i_1, \ldots, i_r, x_1, \ldots, x_r) = -r \log \theta - \frac{\sum_1^r x_i}{\theta} - \frac{(n-r)T}{\theta}$$

Hence,

$$\frac{\partial}{\partial\theta}\log f(i_1,\ldots,i_r,x_1,\ldots,x_r) = -\frac{r}{\theta} + \frac{\sum_1^r x_i + (n-r)T}{\theta^2}$$

Equating to 0 and solving yields that $\hat{\theta}$, the maximum likelihood estimate, is given by

$$\hat{\theta} = \frac{\sum_{i=1}^{r} x_i + (n-r)T}{r}$$

Hence, if we let R denote the number of items that fail by time T and let $X_{(i)}$ be the ith smallest of the failure times, $i = 1,\ldots,R$, then the maximum likelihood estimator of θ is

$$\hat{\theta} = \frac{\sum_{i=1}^{R} X_{(i)} + (n-R)T}{R}$$

Let τ denote the sum of the times that all items are on life test. Then, because the R items that fail are on test for times $X_{(1)},\ldots,X_{(R)}$ whereas the $n-R$ nonfailed items are all on test for time T, it follows that

$$\tau = \sum_{i=1}^{R} X_{(i)} + (n-R)T$$

and thus we can write the maximum likelihood estimator as

$$\hat{\theta} = \frac{\tau}{R}$$

In words, the maximum likelihood estimator of the mean life is (as in the life testing procedures of Sections 14.3.1 and 14.3.2) equal to the total time on test divided by the number of items observed to fail.

REMARK

As the reader may possibly have surmised, it turns out that for all possible life testing schemes for the exponential distribution, the maximum likelihood estimator of the unknown mean θ will always be equal to the total time on test divided by the number of observed failures. To see why this is true, consider *any* testing situation and suppose that the outcome of the data is that r items are observed to fail after having been on test for times $x_1,\ldots,x_r$, respectively, and that s items have not yet failed when the test ends — at which time they had been on test for respective times $y_1, \dots, y_s$. The likelihood of this outcome will be

$$
\begin{aligned}
\text{likelihood} &= K\frac{1}{\theta}e^{-x_1/\theta} \dots \frac{1}{\theta}e^{-x_r/\theta}e^{-y_1/\theta} \dots e^{-y_s/\theta} \\
&= \frac{K}{\theta^r}\exp\left\{\frac{-\left(\sum_{i=1}^{r} x_i + \sum_{i=1}^{s} y_i\right)}{\theta}\right\}
\end{aligned} \tag{14.3.9}
$$

where K, which is a function of the testing scheme and the data, does not depend on θ. (For instance, K may relate to a testing procedure in which the decision as to when to stop depends not only on the observed data but is allowed to be random.) It follows from the foregoing that the maximum likelihood estimate of θ will be

$$
\hat{\theta} = \frac{\sum_{i=1}^{r} x_i + \sum_{i=1}^{s} y_i}{r} \tag{14.3.10}
$$

But $\sum_{i=1}^{r} x_i + \sum_{i=1}^{s} y_i$ is just the total-time-on-test statistic and so the maximum likelihood estimator of θ is indeed the total time on test divided by the number of observed failures in that time.

The distribution of τ/R is rather complicated for the life testing scheme described in this section* and thus we will not be able to easily derive a confidence interval estimator for θ. Indeed, we will not further pursue this problem but rather will consider the Bayesian approach to estimating θ.

14.3.4 The Bayesian Approach

Suppose that items having independent and identically distributed exponential lifetimes with an unknown mean θ are put on life test. Then, as noted in the remark given in Section 14.3.3, the likelihood of the data can be expressed as

$$
f(\text{data}|\theta) = \frac{K}{\theta^r}e^{-t/\theta}
$$

where t is the total time on test — that is, the sum of the time on test of all items used — and r is the number of observed failures for the given data.

Let $\lambda = 1/\theta$ denote the rate of the exponential distribution. In the Bayesian approach, it is more convenient to work with the rate λ rather than its reciprocal. From the

* For instance, for the scheme considered, τ and R are not only both random but are also dependent.

foregoing we see that

$$f(\text{data}|\lambda) = K\lambda^r e^{-\lambda t}$$

If we suppose prior to testing, that λ is distributed according to the prior density $g(\lambda)$, then the posterior density of λ given the observed data is as follows:

$$\begin{aligned} f(\lambda|\text{data}) &= \frac{f(\text{data}|\lambda)g(\lambda)}{\int f(\text{data}|\lambda)g(\lambda)\,d\lambda} \\ &= \frac{\lambda^r e^{-\lambda t} g(\lambda)}{\int \lambda^r e^{-\lambda t} g(\lambda)\,d\lambda} \end{aligned} \tag{14.3.11}$$

The preceding posterior density becomes particularly convenient to work with when g is a gamma density function with parameters, say, (b, a) — that is, when

$$g(\lambda) = \frac{ae^{-a\lambda}(a\lambda)^{b-1}}{\Gamma(b)}, \quad \lambda > 0$$

for some nonnegative constants a and b. Indeed for this choice of g we have from Equation 14.3.11 that

$$\begin{aligned} f(\lambda|\text{data}) &= Ce^{-(a+t)\lambda}\lambda^{r+b-1} \\ &= Ke^{-(a+t)\lambda}[(a+t)\lambda]^{b+r-1} \end{aligned}$$

where C and K do not depend on λ. Because we recognize the preceding as the gamma density with parameters $(b + r, a + t)$, we can rewrite it as

$$f(\lambda|\text{data}) = \frac{(a+t)e^{-(a+t)\lambda}[(a+t)\lambda]^{b+r-1}}{\Gamma(b+r)}, \quad \lambda > 0$$

In other words, if the prior distribution of λ is gamma with parameters (b, a), then no matter what the testing scheme, the (posterior) conditional distribution of λ given the data is gamma with parameters $(b + R, a + \tau)$, where τ and R represent respectively the total-time-on-test statistic and the number of observed failures. Because the mean of a gamma random variable with parameters (b, a) is equal to b/a (see Section 5.7), we can conclude that $E[\lambda|\text{data}]$, the Bayes estimator of λ, is

$$E[\lambda|\text{data}] = \frac{b+R}{a+\tau}$$

EXAMPLE 14.3e Suppose that 20 items having an exponential life distribution with an unknown rate λ are put on life test at various times. When the test is ended, there have been 10 observed failures — their lifetimes being (in hours) 5, 7, 6.2, 8.1, 7.9, 15, 18, 3.9, 4.6, 5.8. The 10 items that did not fail had, at the time the test was terminated, been on test for times (in hours) 3, 3.2, 4.1, 1.8, 1.6, 2.7, 1.2, 5.4, 10.3, 1.5. If prior to the testing it was felt that λ could be viewed as being a gamma random variable with parameters (2, 20), what is the Bayes estimator of λ?

SOLUTION Since

$$\tau = 116.1, \quad R = 10$$

it follows that the Bayes estimate of λ is

$$E[\lambda|\text{data}] = \frac{12}{136.1} = .088 \quad \blacksquare$$

REMARK

As we have seen, the choice of a gamma prior distribution for the rate of an exponential distribution makes the resulting computations quite simple. Whereas, from an applied viewpoint, this is not a sufficient rationale, such a choice is often made with one justification being that the flexibility in fixing the two parameters of the gamma prior usually enables one to reasonably approximate their true prior feelings.

14.4 A TWO-SAMPLE PROBLEM

A company has set up two separate plants to produce vacuum tubes. The company supposes that tubes produced at Plant I function for an exponentially distributed time with an unknown mean θ_1 whereas those produced at Plant II function for an exponentially distributed time with unknown mean θ_2. To test the hypothesis that there is no difference between the two plants (at least in regard to the lifetimes of the tubes they produce), the company samples n tubes from Plant I and m from Plant II and then utilizes these tubes to determine their lifetimes. How can they thus determine whether the two plants are indeed identical?

If we let $X_1, \ldots, X_n$ denote the lifetimes of the n tubes produced at Plant I and $Y_1, \ldots, Y_m$ denote the lifetimes of the m tubes produced at Plant II, then the problem is to test the hypothesis that $\theta_1 = \theta_2$ when the X_i, $i = 1, \ldots, n$ are a random sample from an exponential distribution with mean θ_1 and the Y_i, $i = 1, \ldots, m$ are a random sample from an exponential distribution with mean θ_2. Moreover, the two samples are supposed to be independent.

To develop a test of the hypothesis that $\theta_1 = \theta_2$, let us begin by noting that $\sum_{i=1}^{n} X_i$ and $\sum_{i=1}^{m} Y_i$ (being the sum of independent and identically distributed exponentials) are independent gamma random variables with respective parameters $(n, 1/\theta_1)$ and $(m, 1/\theta_2)$.

Hence, by the equivalence of the gamma and chi-square distribution it follows that

$$\frac{2}{\theta_1}\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

$$\frac{2}{\theta_2}\sum_{i=1}^{m} Y_i \sim \chi^2_{2m}$$

Hence, it follows from the definition of the F-distribution that

$$\frac{\frac{2}{2n\theta_1}\sum_{i=1}^{n} X_i}{\frac{2}{2m\theta_2}\sum_{i=1}^{m} Y_i} \sim F_{n,m}$$

That is, if $\overline{X}$ and $\overline{Y}$ are the two sample means, respectively, then

$$\frac{\theta_2\overline{X}}{\theta_1\overline{Y}} \quad \text{has an } F\text{-distribution with } n \text{ and } m \text{ degrees of freedom}$$

Hence, when the hypothesis $\theta_1 = \theta_2$ is true, we see that $\overline{X}/\overline{Y}$ has an F-distribution with n and m degrees of freedom. This suggests the following test of the hypothesis that $\theta_1 = \theta_2$.

Test: $H_0 : \theta_1 = \theta_2$ vs. alternative $H_1 : \theta_1 \neq \theta_2$
Step 1: Choose a significance level α.
Step 2: Determine the value of the test statistic $\overline{X}/\overline{Y}$ — say its value is v.
Step 3: Compute $P\{F \leq v\}$ where $F \sim F_{n,m}$. If this probability is either less than $\alpha/2$ (which occurs when $\overline{X}$ is significantly less than $\overline{Y}$) or greater than $1-\alpha/2$ (which occurs when $\overline{X}$ is significantly greater than $\overline{Y}$), then the hypothesis is rejected.

In other words, the p-value of the test data is given by

$$p\text{-value} = 2\ \min(P\{F \leq v\}, 1 - P\{F \leq v\})$$

EXAMPLE 14.4a Test the hypothesis, at the 5 percent level of significance, that the lifetimes of items produced at two given plants have the same exponential life distribution if a sample of size 10 from the first plant has a total lifetime of 420 hours whereas a sample of 15 from the second plant has a total lifetime of 510 hours.

SOLUTION The value of the test statistic $\overline{X}/\overline{Y}$ is $42/34 = 1.2353$. To compute the probability that an F-random variable with parameters 10, 15 is less than this value, we run Program 5.8.3a to obtain that

$$P\{F_{10,15} < 1.2353\} = .6554$$

Because the p-value is equal to $2(1 - .6554) = .6892$, we cannot reject H_0. ■

14.5 THE WEIBULL DISTRIBUTION IN LIFE TESTING

Whereas the exponential distribution arises as the life distribution when the hazard rate function $\lambda(t)$ is assumed to be constant over time, there are many situations in which it is more realistic to suppose that $\lambda(t)$ either increases or decreases over time. One example of such a hazard rate function is given by

$$\lambda(t) = \alpha\beta t^{\beta-1}, \quad t > 0 \tag{14.5.1}$$

where α and β are positive constants. The distribution whose hazard rate function is given by Equation 14.5.1 is called the *Weibull* distribution with parameters (α, β). Note that $\lambda(t)$ increases when $\beta > 1$, decreases when $\beta < 1$, and is constant (reducing to the exponential) when $\beta = 1$.

The Weibull distribution function is obtained from Equation 14.5.1 as follows:

$$\begin{aligned} F(t) &= 1 - \exp\left\{-\int_0^t \lambda(s)\, ds\right\}, \quad t > 0 \\ &= 1 - \exp\{-\alpha t^\beta\} \end{aligned}$$

Differentiating yields its density function:

$$f(t) = \alpha\beta t^{\beta-1} \exp\{-\alpha t^\beta\}, \quad t > 0 \tag{14.5.2}$$

This density is plotted for a variety of values of α and β in Figure 14.2.

Suppose now that $X_1, \ldots, X_n$ are independent Weibull random variables each having parameters (α, β), which are assumed unknown. To estimate α and β, we can employ the maximum likelihood approach. Equation 14.5.2 yields the likelihood, given by

$$f(x_1, \ldots, x_n) = \alpha^n \beta^n x_1^{\beta-1} \cdots x_n^{\beta-1} \exp\left\{-\alpha \sum_{i=1}^n x_i^\beta\right\}$$

Hence,

$$\log f(x_1, \ldots, x_n) = n \log \alpha + n \log \beta + (\beta - 1) \sum_{i=1}^n \log x_i - \alpha \sum_{i=1}^n x_i^\beta$$

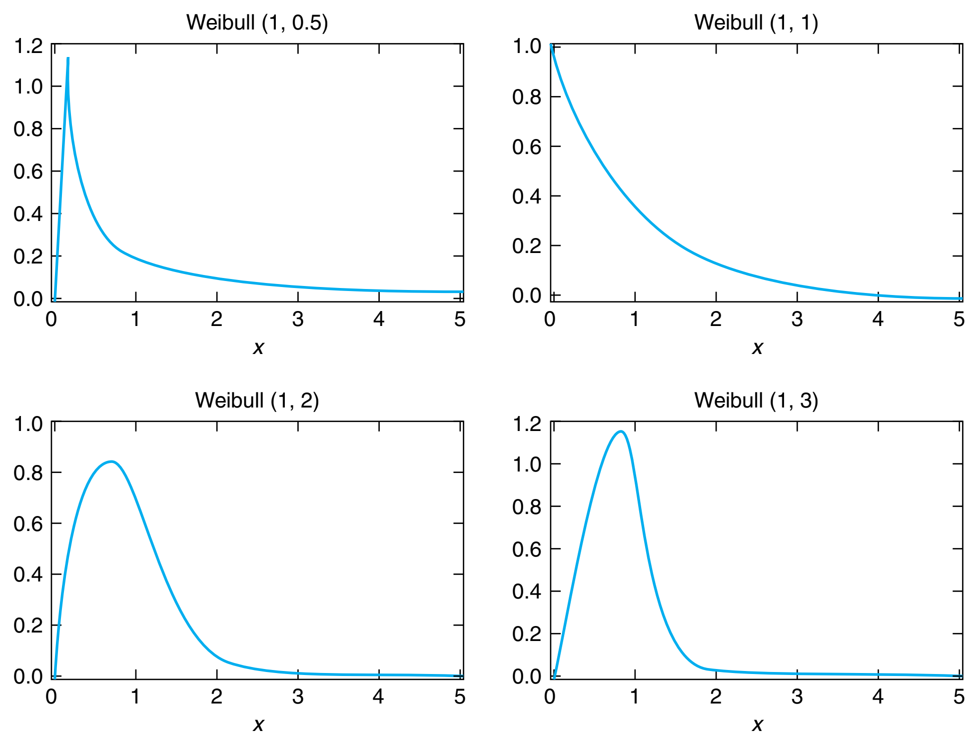


FIGURE 14.2 *Weibull density functions.*

and

$$\frac{\partial}{\partial \alpha} \log f(x_1, \ldots, x_n) = \frac{n}{\alpha} - \sum_{i=1}^{n} x_i^{\beta}$$

$$\frac{\partial}{\partial \beta} \log f(x_1, \ldots, x_n) = \frac{n}{\beta} + \sum_{i=1}^{n} \log x_i - \alpha \sum_{i=1}^{n} x_i^{\beta} \log x_i$$

Equating to zero shows that the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$ are the solutions of

$$\frac{n}{\hat{\alpha}} = \sum_{i=1}^{n} x_i^{\hat{\beta}}$$

$$\frac{n}{\hat{\beta}} + \sum_{i=1}^{n} \log x_i = \hat{\alpha} \sum_{i=1}^{n} x_i^{\hat{\beta}} \log x_i$$

or, equivalently,

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} x_i^{\hat{\beta}}}$$

$$n + \hat{\beta} \log\left(\prod_{i=1}^{n} x_i\right) = \frac{n\hat{\beta} \sum_{i=1}^{n} x_i^{\hat{\beta}} \log x_i}{\sum_{i=1}^{n} x_i^{\hat{\beta}}}$$

This latter equation can then be solved numerically for $\hat{\beta}$, which will then also determine $\hat{\alpha}$. However, rather than pursuing this approach any further, let us consider a second approach, which is not only computationally easier but appears, as indicated by a simulation study, to yield more accurate estimates.

14.5.1 Parameter Estimation by Least Squares

Let $X_1, \ldots, X_n$ be a sample from the distribution

$$F(x) = 1 - e^{-\alpha x^{\beta}}, \quad x \geq 0$$

Note that

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

or

$$\log\left(\frac{1}{1 - F(x)}\right) = \alpha x^{\beta}$$

and so

$$\log\log\left(\frac{1}{1 - F(x)}\right) = \beta \log x + \log \alpha \tag{14.5.3}$$

Now let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ denote the ordered sample values — that is, for $i = 1, \ldots, n$,

$$X_{(i)} = i\text{th smallest of } X_1, \ldots, X_n$$

and suppose that the data results in $X_{(i)} = x_{(i)}$. If we were able to approximate the quantities $\log\log(1/[1-F(x_{(i)})])$ — say, by the values $y_1, \ldots, y_n$ — then from Equation 14.5.3, we could conclude that

$$y_i \approx \beta \log x_{(i)} + \log \alpha, \quad i = 1, \ldots, n \tag{14.5.4}$$

We could then choose α and β to minimize the sum of the squared errors — that is, α and β are chosen to

$$\underset{\alpha,\beta}{\text{minimize}} \sum_{i=1}^{n} (y_i - \beta \log x_{(i)} - \log \alpha)^2$$

Indeed, using Proposition 9.2.1 we obtain that the preceding minimum is attained when $\alpha = \hat{\alpha}, \beta = \hat{\beta}$ where

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i \log x_{(i)} - n\, \overline{\log x}\bar{y}}{\sum_{i=1}^{n} (\log x_{(i)})^2 - n(\overline{\log x})^2}$$

$$\log \hat{\alpha} = \bar{y} - \beta\, \overline{\log x}$$

where

$$\overline{\log x} = \sum_{i=1}^{n} (\log x_{(i)}) \Big/ n, \quad \bar{y} = \sum_{i=1}^{n} y_i \Big/ n$$

To utilize the foregoing, we need to be able to determine values y_i that approximate $\log\log(1/[1 - F(x_{(i)})]) = \log[-\log(1 - F(x_{(i)}))], i = 1, \ldots, n$. We now present two different methods for doing this.

Method 1: This method uses the fact that

$$E[F(X_{(i)})] = \frac{i}{(n+1)} \tag{14.5.5}$$

and then approximates $F(x_{(i)})$ by $E[F(X_{(i)})]$. Thus, this method calls for using

$$\begin{aligned} y_i &= \log\{-\log(1 - E[F(X_{(i)})])\} \\ &= \log\left\{-\log\left(1 - \frac{i}{(n+1)}\right)\right\} \\ &= \log\left\{-\log\left(\frac{n+1-i}{n+1}\right)\right\} \end{aligned} \tag{14.5.6}$$

Method 2: This method uses the fact that

$$E[-\log(1 - F(X_{(i)}))] = \frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-i+1} \tag{14.5.7}$$

and then approximates $-\log(1 - F(x_{(i)}))$ by the foregoing. Thus, this second method calls for setting

$$y_i = \log\left[\frac{1}{n} + \frac{1}{(n-1)} + \cdots + \frac{1}{(n-i+1)}\right] \tag{14.5.8}$$

REMARKS

(a) It is not, at present, clear which method provides superior estimates of the parameters of the Weibull distribution, and extensive simulation studies will be necessary to determine this.

(b) Proofs of equalities 14.5.5 and 14.5.7 [which hold whenever $X_{(i)}$ is the ith smallest of a sample of size n from any continuous distribution F] are outlined in Problems 28–30.

Problems

1. A random variable whose distribution function is given by

$$F(t) = 1 - \exp\{-\alpha t^{\beta}\}, \quad t \geq 0$$

is said to have a Weibull distribution with parameters α, β. Compute its failure rate function.

2. If X and Y are independent random variables having failure rate functions $\lambda_x(t)$ and $\lambda_y(t)$, show that the failure rate function of $Z = \min(X, Y)$ is

$$\lambda_z(t) = \lambda_x(t) + \lambda_y(t)$$

3. The lung cancer rate of a t-year-old male smoker, $\lambda(t)$, is such that

$$\lambda(t) = .027 + .025\left(\frac{t-40}{10}\right)^4, \quad t \geq 40$$

Assuming that a 40-year-old male smoker survives all other hazards, what is the probability that he survives to **(a)** age 50, **(b)** age 60, without contracting lung cancer? In the foregoing we are assuming that he remains a smoker throughout his life.

4. Suppose the life distribution of an item has failure rate function $\lambda(t) = t^3, 0 < t < \infty$.
   (a) What is the probability that the item survives to age 2?
   (b) What is the probability that the item's life is between .4 and 1.4?
   (c) What is the mean life of the item?
   (d) What is the probability a 1-year-old item will survive to age 2?

5. A continuous life distribution is said to be an IFR (increasing failure rate) distribution if its failure rate function $\lambda(t)$ is nondecreasing in t.
   (a) Show that the gamma distribution with density

$$f(t) = \lambda^2 t e^{-\lambda t}, \quad t > 0$$

   is IFR.
   (b) Show, more generally, that the gamma distribution with parameters α, λ is IFR whenever $\alpha \geq 1$.
   Hint: Write

$$\lambda(t) = \left[\frac{\int_t^\infty \lambda e^{-\lambda s} (\lambda s)^{\alpha-1} \, ds}{\lambda e^{-\lambda t} (\lambda t)^{\alpha-1}} \right]^{-1}$$

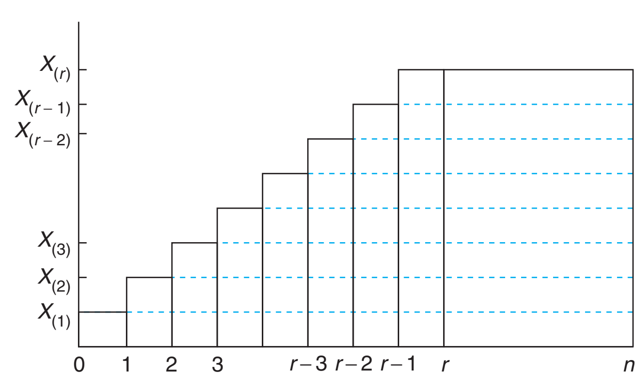
6. Show that the uniform distribution on (a, b) is an IFR distribution.

7. For the model of Section 14.3.1, explain how the following figure can be used to show that

$$\tau = \sum_{j=1}^{r} Y_j$$

where

$$Y_j = (n - j + 1)(X_{(j)} - X_{(j-1)})$$



(*Hint*: Argue that both τ and $\sum_{j=1}^{r} Y_j$ equal the total area of the figure shown.)

8. When 30 transistors were simultaneously put on a life test that was to be terminated when the 10th failure occurred, the observed failure times were (in hours) 4.1, 7.3, 13.2, 18.8, 24.5, 30.8, 38.1, 45.5, 53, 62.2. Assume an exponential life distribution.

(a) What is the maximum likelihood estimate of the mean life of a transistor?
(b) Compute a 95 percent two-sided confidence interval for the mean life of a transistor.
(c) Determine a value c that we can assert, with 95 percent confidence, is less than the mean transistor life.
(d) Test at the $\alpha = .10$ level of significance the hypothesis that the mean lifetime is 7.5 hours versus the alternative that it is not 7.5 hours.

9. Consider a test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ for the model of Section 14.3.1. Suppose that the observed value of $2\tau/\theta_0$ is v. Show that the hypothesis should be rejected at significance level α whenever α is less than the p-value given by

$$p\text{-value} = 2\min(P\{\chi^2_{2r} < v\}, 1 - P\{\chi^2_{2r} < v\})$$

where χ^2_{2r} is a chi-square random variable with $2r$ degrees of freedom.

10. Suppose 30 items are put on test that is scheduled to stop when the 8th failure occurs. If the failure times are, in hours, .35, .73, .99, 1.40, 1.45, 1.83, 2.20, 2.72, test, at the 5 percent level of significance, the hypothesis that the mean life is equal to 10 hours. Assume that the underlying distribution is exponential.

11. Suppose that 20 items are to be put on test that is to be terminated when the 10th failure occurs. If the lifetime distribution is exponential with mean 10 hours, compute the following quantities.

(a) The mean length of the testing period.
(b) The variance of the testing period.

12. Vacuum tubes produced at a certain plant are assumed to have an underlying exponential life distribution having an unknown mean θ. To estimate θ it has been decided to put a certain number n of tubes on test and to stop the test at the 10th failure. If the plant officials want the mean length of the testing period to be 3 hours when the value of θ is $\theta = 20$, approximately how large should n be?

13. A one-at-a-time sequential life testing scheme is scheduled to run for 300 hours. A total of 16 items fail within that time. Assuming an exponential life distribution with unknown mean θ (measured in hours):

(a) Determine the maximum likelihood estimate of θ.

(b) Test at the 5 percent level of significance the hypothesis that $\theta = 20$ versus the alternative that $\theta \neq 20$.

(c) Determine a 95 percent confidence interval for θ.

14. Using the fact that a Poisson process results when the times between successive events are independent and identically distributed exponential random variables, show that

$$P\{X \geq n\} = F_{\chi^2_{2n}}(x)$$

when X is a Poisson random variable with mean $x/2$ and $F_{\chi^2_{2n}}$ is the chi-square distribution function with $2n$ degrees of freedom. (*Hint*: Use the results of Section 14.3.2.)

15. From a sample of items having an exponential life distribution with unknown mean θ, items are tested in sequence. The testing continues until either the rth failure occurs or after a time T elapses.

(a) Determine the likelihood function.

(b) Verify that the maximum likelihood estimator of θ is equal to the total time on test of all items divided by the number of observed failures.

16. Verify that the maximum likelihood estimate corresponding to Equation 14.3.9 is given by Equation 14.3.10.

17. A testing laboratory has facilities to simultaneously life test 5 components. The lab tested a sample of 10 components from a common exponential distribution by initially putting 5 on test and then replacing any failed component by one still waiting to be tested. The test was designed to end either at 200 hours or when all 10 components had failed. If there were a total of 9 failures occurring at times 15, 28.2, 46, 62.2, 76, 86, 128, 153, 197, what is the maximum likelihood estimate of the mean life of a component?

18. Suppose that the remission time, in weeks, of leukemia patients that have undergone a certain type of chemotherapy treatment is an exponential random variable having an unknown mean θ. A group of 20 such patients is being monitored and, at present, their remission times are (in weeks) 1.2, 1.8*, 2.2, 4.1, 5.6, 8.4, 11.8*, 13.4*, 16.2, 21.7, 29*, 41, 42*, 42.4*, 49.3, 60.5, 61*, 94, 98, 99.2* where an asterisk next to the data means that the patient's remission is continuing, whereas a data point without an asterisk means that the remission ended at that time. What is the maximum likelihood estimate of θ?

19. In Problem 17, suppose that prior to the testing phase and based on past experience one felt that the value of $\lambda = 1/\theta$ could be thought of as the outcome of a gamma random variable with parameters 1, 100. What is the Bayes estimate of λ?

20. What is the Bayes estimate of $\lambda = 1/\theta$ in Problem 18 if the prior distribution on λ is exponential with mean 1/30?

21. The following data represent failure times, in minutes, for two types of electrical insulation subject to a certain voltage stress.

Type I	212, 88.5, 122.3, 116.4, 125, 132, 66
Type II	34.6, 54, 162, 49, 78, 121, 128

Test the hypothesis that the two sets of data come from the same exponential distribution.

22. Suppose that the life distributions of two types of transistors are both exponential. To test the equality of means of these two distributions, n_1 type 1 transistors are simultaneously put on a life test that is scheduled to end when there have been a total of r_1 failures. Similarly, n_2 type 2 transistors are simultaneously put on a life test that is to end when there have been r_2 failures.

(a) Using results from Section 14.3.1, show how the hypothesis that the means are equal can be tested by using a test statistic that, when the means are equal, has an F-distribution with $2r_1$ and $2r_2$ degrees of freedom.

(b) Suppose $n_1 = 20$, $r_1 = 10$ and $n_2 = 10$, $r_2 = 7$ with the following data resulting.

Type 1 failures at times:

10.4, 23.2, 31.4, 45, 61.1, 69.6, 81.3, 95.2, 112, 129.4

Type 2 failures at times:

6.1, 13.8, 21.2, 31.6, 46.4, 66.7, 92.4

What is the smallest significance level α for which the hypothesis of equal means would be rejected? (That is, what is the p-value of the test data?)

23. If X is a Weibull random variable with parameters (α, β), show that

$$E[X] = \alpha^{-1/\beta}\Gamma(1 + 1/\beta)$$

where $\Gamma(y)$ is the gamma function defined by

$$\Gamma(y) = \int_0^\infty e^{-x}x^{y-1}\,dx$$

Hint: Write

$$E[X] = \int_0^\infty t\alpha\beta t^{\beta-1}\exp\{-\alpha t^\beta\}\,dt$$

and make the change of variables

$$x = \alpha t^{\beta}, \quad dx = \alpha\beta t^{\beta-1}\, dt$$

24. Show that if X is a Weibull random variable with parameters (α, β), then

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

25. If the following are the sample data from a Weibull population having unknown parameters α and β, determine the least square estimates of these quantities, using either of the methods presented.

Data: 15.4, 16.8, 6.2, 10.6, 21.4, 18.2, 1.6, 12.5, 19.4, 17

26. Show that if X is a Weibull random variable with parameters (α, β), then αX^{β} is an exponential random variable with mean 1.

27. If U is uniformly distributed on (0, 1) — that is, U is a random number — show that $[-(1/\alpha)\log U]^{1/\beta}$ is a Weibull random variable with parameters (α, β).

The next three problems are concerned with verifying Equations 14.5.5 and 14.5.7.

28. If X is a continuous random variable having distribution function F, show that

(a) $F(X)$ is uniformly distributed on (0, 1);
(b) $1 - F(X)$ is uniformly distributed on (0, 1).

29. Let $X_{(i)}$ denote ith smallest of a sample of size n from a continuous distribution function F. Also, let $U_{(i)}$ denote the ith smallest from a sample of size n from a uniform (0, 1) distribution.

(a) Argue that the density function of $U_{(i)}$ is given by

$$f_{U_{(i)}}(t) = \frac{n!}{(n-i)!(i-1)!}t^{i-1}(1-t)^{n-i}, \quad 0 < t < 1$$

[*Hint*: In order for the ith smallest of n uniform (0, 1) random variables to equal t, how many must be less than t and how many must be greater? Also, in how many ways can a set of n elements be broken into three subsets of respective sizes $i-1$, 1, and $n-i$?]

(b) Use part (a) to show that $E[U(i)] = i/(n+1)$. [*Hint*: To evaluate the resulting integral, use the fact that the density in part (a) must integrate to 1.]

(c) Use part (b) and Problem 28a to conclude that $E[F(X_{(i)})] = i/(n+1)$.

30. If U is uniformly distributed on (0, 1), show that $-\log U$ has an exponential distribution with mean 1. Now use Equation 14.3.7 and the results of the previous problems to establish Equation 14.5.7.