Chapter 5

Statistical Methods in Survival Analysis

The previous chapter introduced probability models that are frequently used in survival analysis. This chapter introduces the associated statistical methods.

The focus in this chapter is the use of maximum likelihood for parameter estimation and inference. Likelihood theory is illustrated in the first section. The matrix of the expected values of the opposite of the second partial derivatives of the log likelihood function is known as the *Fisher information matrix* and its statistical analog, the *observed information matrix*, is useful for determining confidence intervals for parameters. Asymptotic properties of the likelihood function, which are associated with large sample sizes, are reviewed in the second section. One distinctive feature of lifetime data is the presence of *censoring*, which occurs when only an upper or lower bound on the lifetime is known. Statistical methods for handling censored data values are introduced in the third section. The focus is on right censoring, where only a lower bound on the failure time is known. These methods are applied to the exponential distribution and the Weibull distribution in the next two sections. Finally, the last section indicates how to fit the proportional hazards model to a data set consisting of lifetimes with associated covariates.

5.1 Likelihood Theory

There are always merits in obtaining raw data (that is, exact individual failure times), as opposed to grouped data (counts of the number of failures over prescribed time intervals). Given raw data, we can always construct grouped data, but the converse is typically not true; therefore, we limit discussion in this chapter to the raw data case.

The random variable T has denoted a random lifetime in previous chapter. So it is natural to use $T_1, T_2, \ldots, T_n$ to denote a *random sample* of n such lifetimes, where n is the number of items on test. When specific values are given for realizations of such lifetimes, which is typically the case from this point forward, they are denoted by $t_1, t_2, \ldots, t_n$. In other words, $t_1, t_2, \ldots, t_n$ are the experimental values of the mutually independent and identically distributed random variables $T_1, T_2, \ldots, T_n$. The associated ordered observations, or order statistics, are denoted by $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$.

The Greek letter θ is often used to denote a generic unknown parameter. We will refer to $\hat{\theta}$ in the abstract as a *point estimator*; when $\hat{\theta}$ assumes a specific numeric value, it will be referred to as a *point estimate*. The probability distribution of a statistic is referred to as a *sampling distribution*.

Assume that there is a single unknown parameter θ in the probability model for T. Assume further that the data values $t_1, t_2, \ldots, t_n$ are mutually independent and identically distributed random variables. The joint probability density function of the data values is the product of the marginal probability density functions of the individual observations:

$$L(t_1, t_2, \ldots, t_n, \theta) = \prod_{i=1}^{n} f(t_i; \theta).$$

This function is the *likelihood function*. In order to simplify the notation, the likelihood function is often written as simply

$$L(\theta) = \prod_{i=1}^{n} f(t_i).$$

The maximum likelihood estimator of θ, which is denoted by $\hat{\theta}$, is the value of θ that maximizes $L(\theta)$.

The next example reviews the associated notions of the log likelihood function, score vector, maximum likelihood estimator, Fisher information matrix, and observed information matrix for a two-parameter lifetime model. We assume for now that there are no censored observations in the data set; all of the failure times are observed.

Example 5.1 Let $t_1, t_2, \ldots, t_n$ be a random sample from an *inverse Gaussian* (Wald) population having unknown positive parameters λ and μ, where μ is the population mean. The probability density function of the inverse Gaussian distribution is

$$f(t) = \sqrt{\frac{\lambda}{2\pi}}\, t^{-3/2} e^{-\lambda(t-\mu)^2/(2\mu^2 t)} \qquad t > 0.$$

Find the likelihood function, log likelihood function, score vector, maximum likelihood estimator, Fisher information matrix, and observed information matrix.

The likelihood function is

$$\begin{aligned} L(\boldsymbol{t}, \lambda, \mu) &= \prod_{i=1}^{n} \sqrt{\frac{\lambda}{2\pi}}\, t_i^{-3/2} e^{-\lambda(t_i-\mu)^2/(2\mu^2 t_i)} \\ &= \lambda^{n/2} (2\pi)^{-n/2} \left[\prod_{i=1}^{n} t_i \right]^{-3/2} e^{-\lambda/(2\mu^2)\sum_{i=1}^{n} (t_i-\mu)^2/t_i}, \end{aligned}$$

where $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$. The likelihood function and any monotonic transformation of the likelihood function are maximized at the same value. Since the calculus and algebra is often easier when working with the logarithm of the likelihood function, we do so in this setting. The log likelihood function is

$$\ln L(\boldsymbol{t}, \lambda, \mu) = \frac{n}{2} \ln \lambda - \frac{n}{2} \ln(2\pi) - \frac{3}{2} \sum_{i=1}^{n} \ln t_i - \frac{\lambda}{2\mu^2} \sum_{i=1}^{n} \frac{(t_i-\mu)^2}{t_i}.$$

The two-component score vector, $\boldsymbol{U}(\lambda, \mu)$, consists of the partial derivatives with respect to the two unknown parameters:

$$\frac{\partial \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \lambda} = \frac{n}{2\lambda} - \frac{1}{2\mu^2} \sum_{i=1}^{n} \frac{(t_i-\mu)^2}{t_i}$$

and

$$\frac{\partial \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \mu} = \frac{\lambda}{\mu^3}\left[\sum_{i=1}^{n} t_i - n\mu\right].$$

When the second equation is equated to zero, the maximum likelihood estimator $\hat{\mu}$ is determined. Then using $\hat{\mu}$ as an argument in the first equation and solving for $\hat{\lambda}$ results in the maximum likelihood estimators

$$\hat{\lambda} = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_i} - \frac{n}{\sum_{i=1}^{n} t_i}\right]^{-1} \qquad \text{and} \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} t_i.$$

The second partial derivatives of the log likelihood function are

$$\frac{\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \lambda^2} = -\frac{n}{2\lambda^2}$$

$$\frac{\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \lambda \partial \mu} = \frac{1}{\mu^3}\sum_{i=1}^{n} t_i - \frac{n}{\mu^2}$$

$$\frac{\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \mu^2} = -\frac{3\lambda}{\mu^4}\sum_{i=1}^{n} t_i + \frac{2n\lambda}{\mu^3}.$$

Since $E[T] = \mu$ for the inverse Gaussian distribution, the Fisher information matrix consists of the expected values of the negatives of these derivatives:

$$I(\lambda, \mu) = \begin{pmatrix} E\left[\dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \lambda^2}\right] & E\left[\dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \lambda \partial \mu}\right] \\ E\left[\dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \mu \partial \lambda}\right] & E\left[\dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \mu^2}\right] \end{pmatrix} = \begin{pmatrix} \dfrac{n}{2\lambda^2} & 0 \\ 0 & \dfrac{n\lambda}{\mu^3} \end{pmatrix}.$$

The Fisher information matrix is the variance–covariance matrix of the score vector. The off-diagonal elements being zero for the inverse Gaussian distribution implies that the elements of the score vector are uncorrelated. Although this example has simple closed-form expressions for the Fisher information matrix, it is more often the case that the elements of the Fisher information matrix are not closed form. The observed information matrix can be calculated for all distributions; it uses the maximum likelihood estimates:

$$O(\hat{\lambda}, \hat{\mu}) = \begin{pmatrix} \dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \lambda^2} & \dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \lambda \partial \mu} \\ \dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \mu \partial \lambda} & \dfrac{-\partial^2 \ln L(\boldsymbol{t}, \lambda, \mu)}{\partial \mu^2} \end{pmatrix}_{\lambda = \hat{\lambda},\, \mu = \hat{\mu}} = \begin{pmatrix} \dfrac{n}{2\hat{\lambda}^2} & 0 \\ 0 & \dfrac{n\hat{\lambda}}{\hat{\mu}^3} \end{pmatrix}.$$

In some cases, it is possible to find the exact distribution of a pivotal quantity which results in exact statistical inference (that is, constructing exact confidence intervals and performing exact hypothesis tests). It is more often the case that exact statistical inference is not possible, and asymptotic properties associated with the likelihood function must be relied on for approximate inference. The next section reviews some asymptotic properties that arise in likelihood theory. When a large data set of lifetimes is available, these properties often lead to approximate statistical methods of inference.

5.2 Asymptotic Properties

When the number of items on test n is large, there are some asymptotic results concerning the likelihood function that are useful for constructing confidence intervals and performing hypothesis tests associated with a vector of p unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$. As indicated in the example in the last section, the $p \times 1$ score vector $\boldsymbol{U}(\boldsymbol{\theta})$ has elements

$$U_i(\boldsymbol{\theta}) = \frac{\partial \ln L(\boldsymbol{t}, \boldsymbol{\theta})}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \sum_{j=1}^{n} \ln f(t_j, \boldsymbol{\theta})$$

for $i = 1, 2, \ldots, p$. Therefore, each element of the score vector is a sum of mutually independent random variables, and, when n is large, the elements of $\boldsymbol{U}(\boldsymbol{\theta})$ are asymptotically normally distributed by the central limit theorem. More specifically, the score vector $\boldsymbol{U}(\boldsymbol{\theta})$ is asymptotically normal with population mean $\mathbf{0}$ and variance–covariance matrix $I(\boldsymbol{\theta})$, where $I(\boldsymbol{\theta})$ is the Fisher information matrix. This means that when the true value for the parameter vector is $\boldsymbol{\theta}_0$ then

$$\boldsymbol{U}'(\boldsymbol{\theta}_0) I(\boldsymbol{\theta}_0)^{-1} \boldsymbol{U}(\boldsymbol{\theta}_0)$$

is asymptotically chi-square with p degrees of freedom. This can be used to determine confidence intervals and perform hypothesis tests with respect to $\boldsymbol{\theta}$.

The maximum likelihood estimator for the parameter vector $\hat{\boldsymbol{\theta}}$ can also be used for confidence intervals and hypothesis testing. Since $\hat{\boldsymbol{\theta}}$ is asymptotically normal with population mean $\boldsymbol{\theta}$ and variance–covariance matrix $I^{-1}(\boldsymbol{\theta})$, when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' I(\boldsymbol{\theta}_0) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$$

is also asymptotically chi-square with p degrees of freedom. Two statistics that are asymptotically equivalent to this statistic that can be used to estimate the value of the chi-square random variable are

$$\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' I\left(\hat{\boldsymbol{\theta}}\right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$$

and

$$\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' O\left(\hat{\boldsymbol{\theta}}\right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right).$$

A third asymptotic result involves the likelihood ratio statistic

$$-2\left[\ln L(\boldsymbol{\theta}) - \ln L(\hat{\boldsymbol{\theta}})\right] = -2 \ln \left[\frac{L(\boldsymbol{\theta})}{L(\hat{\boldsymbol{\theta}})}\right],$$

which is asymptotically chi-square with p degrees of freedom. The conditions necessary for these asymptotic properties to apply are cited at the end of the chapter.

These three asymptotic results are summarized in the result below, where the a above the $\sim$ is shorthand for "asymptotically distributed."

Theorem 5.1 Let $t_1, t_2, \ldots, t_n$ be mutually independent and identically distributed lifetimes from a population distribution with p unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$. Then

$$\boldsymbol{U}'(\boldsymbol{\theta}_0) I(\boldsymbol{\theta}_0)^{-1} \boldsymbol{U}(\boldsymbol{\theta}_0) \overset{a}{\sim} \chi^2(p), \quad \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' O\left(\hat{\boldsymbol{\theta}}\right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \overset{a}{\sim} \chi^2(p), \quad -2 \ln \left[\frac{L(\boldsymbol{\theta})}{L(\hat{\boldsymbol{\theta}})}\right] \overset{a}{\sim} \chi^2(p).$$

Example 5.2 Let $t_1, t_2, \ldots, t_n$ be a random sample from a population with probability density function

$$f(t) = \frac{1}{\sqrt{2\pi t^3}}\, e^{-(t-\mu)^2/(2\mu^2 t)} \qquad\qquad t > 0,$$

where μ is a positive unknown parameter, which is the population mean. This population distribution is a special case of the two-parameter inverse Gaussian distribution. Use one of the asymptotic results from Theorem 5.1 to construct an asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for μ.

The first step is to find the maximum likelihood estimator of μ. The likelihood function is

$$\begin{aligned} L(\boldsymbol{t}, \mu) &= \prod_{i=1}^{n} \left(2\pi t_i^3\right)^{-1/2} e^{-(t_i-\mu)^2/\left(2\mu^2 t_i\right)} \\ &= (2\pi)^{-n/2} \left[\prod_{i=1}^{n} t_i\right]^{-3/2} e^{-\sum_{i=1}^{n}(t_i-\mu)^2/\left(2\mu^2 t_i\right)}. \end{aligned}$$

The log likelihood function is

$$\ln L(\boldsymbol{t}, \mu) = -\frac{n}{2}\ln(2\pi) - \frac{3}{2}\sum_{i=1}^{n} \ln t_i - \frac{1}{2\mu^2}\sum_{i=1}^{n} \frac{(t_i-\mu)^2}{t_i}.$$

The score is the derivative of the log likelihood function with respect to μ, which, after simplification, is

$$\frac{\partial \ln L(\boldsymbol{t},\mu)}{\partial \mu} = \frac{1}{\mu^3}\left[\sum_{i=1}^{n} t_i - n\mu\right].$$

When this equation is equated to zero, the maximum likelihood estimator for μ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} t_i,$$

which is the sample mean. The second partial derivative of the log likelihood function is

$$\frac{\partial^2 \ln L(\boldsymbol{t}, \mu)}{\partial \mu^2} = -\frac{3}{\mu^4}\sum_{i=1}^{n} t_i + \frac{2n}{\mu^3},$$

which is negative at the maximum likelihood estimator, so the maximum likelihood estimator maximizes the log likelihood function. The next step is to find the 1×1 Fisher information matrix. Using the second partial derivative of the log likelihood function, the Fisher information matrix is

$$I(\mu) = E\left[-\frac{\partial^2 \ln L(\boldsymbol{t}, \mu)}{\partial \mu^2}\right] = E\left[\frac{3}{\mu^4}\sum_{i=1}^{n} t_i - \frac{2n}{\mu^3}\right] = \frac{3n\mu}{\mu^4} - \frac{2n}{\mu^3} = \frac{n}{\mu^3}$$

because $E[X] = \mu$ for this population distribution. The 1×1 observed information is

$$O(\hat{\mu}) = \left[-\frac{\partial^2 \ln L(\boldsymbol{t}, \mu)}{\partial \mu^2}\right]_{\mu=\hat{\mu}} = \frac{n}{\hat{\mu}^3}.$$

In order to construct an asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for μ, recall that $\hat{\mu}$ is asymptotically normal with population mean μ and variance–covariance matrix $I^{-1}(\mu)$. In other words,

$$\hat{\mu} \overset{a}{\sim} N\left(\mu, I^{-1}(\mu)\right).$$

For large values of n, we replace the Fisher information matrix with the observed information matrix:

$$\hat{\mu} \overset{a}{\sim} N\left(\mu, O^{-1}(\hat{\mu})\right)$$

or

$$\hat{\mu} \overset{a}{\sim} N\left(\mu, \frac{\hat{\mu}^3}{n}\right).$$

This random variable can be standardized by subtracting its population mean and dividing by its population standard deviation:

$$\frac{\hat{\mu}-\mu}{\sqrt{\hat{\mu}^3/n}} \overset{a}{\sim} N(0,1).$$

So the probability that this random variable falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$ for large n is

$$\lim_{n\to\infty} P\left(-z_{\alpha/2} < \frac{\hat{\mu}-\mu}{\sqrt{\hat{\mu}^3/n}} < z_{\alpha/2}\right) = 1-\alpha.$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Rearranging the inequality

$$-z_{\alpha/2} < \frac{\hat{\mu}-\mu}{\sqrt{\hat{\mu}^3/n}} < z_{\alpha/2}$$

so that μ is in the center of the inequality yields the asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval

$$\hat{\mu} - z_{\alpha/2}\frac{\hat{\mu}^{3/2}}{\sqrt{n}} < \mu < \hat{\mu} + z_{\alpha/2}\frac{\hat{\mu}^{3/2}}{\sqrt{n}},$$

where $\hat{\mu}$ is the sample mean of the observed data values. The actual coverage of confidence intervals developed in this fashion typically approaches $1-\alpha$ as the number of items on test n increases.

All of the statistical methods developed thus far have assumed that we are able to observe all n of the items on test fail. The associated lifetimes are denoted by $t_1, t_2, \ldots, t_n$. Although this is ideal and might be the case in some settings, a short testing time or items with long lifetimes might result in some items that survive the test. The lifetimes of the items which do not fail during the test are known as *right-censored* observations. The lifetimes of these items are not observed, but are known to exceed the time at which the test is concluded. If a decision concerning the acceptability of the items must be made with some of the items still operating at the end of the test, then a statistical model must be formulated to account for the unobserved lifetimes of these items. The next section introduces the important topic of censoring, which is pervasive in survival analysis.

5.3 Censoring

Censoring occurs in lifetime data sets when only an upper or lower bound on the lifetime is known. Censoring occurs frequently in lifetime data sets because it is often impossible or impractical to observe the lifetimes of all the items on test. A data set for which all failure times are known is called a *complete data set*. Figure 5.1 illustrates a complete data set of $n = 5$ items placed on test simultaneously at time $t = 0$, where the $\times$'s denote failure times. Consider the two endpoints of each of the horizontal segments. It is critical to provide an unambiguous definition of the time origin (for example, the time a product is purchased or the time a cancer is diagnosed). Likewise, failure must be defined in an unambiguous fashion. This is easier to define for a light bulb or a fuse than for a ball bearing or a sock. Outside of a reliability setting, a data set of lifetimes is often generically referred to as *time-to-event* data, corresponding to the time between the time origin and the event of interest. A *censored observation* occurs when only a bound is known on the time of failure. If a data set contains one or more censored observations, it is called a *censored data set*.

The most common type of censoring is known as *right censoring*. In a right-censored data set, one or more items have only a lower bound known on their lifetime. The term *sample size* is now vague. From this point forward, we use n to denote the *number of items on test* and use r to denote the *number of observed failures*. In an industrial life testing situation, for example, $n = 12$ cell phones are put on a continuous, rigorous life test on January 1, and $r = 3$ of the cell phones have failed by December 31. These failed cell phones are discarded upon failure. The remaining $n - r = 12 - 3 = 9$ cell phones that are still operating on December 31 have lifetimes that exceed 365 days, and are therefore right-censored observations. Right censoring is not limited to just reliability applications. In a medical study in which T is the survival time after the diagnosis of a particular type of cancer, for example, a patient can either (a) still be alive at the end of a study, (b) die of a cause other than the particular type of cancer, constituting a right-censored observation, or (c) lose contact with the study (for example, if they leave town), constituting a right-censored observation.

Three special cases of right censoring are common in survival analysis. The first is Type II or *order statistic* censoring. As shown in Figure 5.2, this corresponds to terminating a study upon one of the ordered failures. The diagram corresponds to a set of $n = 5$ items placed on a test simultaneously at time $t = 0$. The test is terminated when $r = 3$ failures are observed. Time advances from left to right in Figure 5.2 and the failure of the first item (corresponding to the third ordered observed failure) terminates the test. The lifetimes of the third and fourth items are right censored. Observed failure times are indicated by an $\times$ and right-censoring times are indicated by a $\circ$. In Type II censoring, the time to complete the test is random.

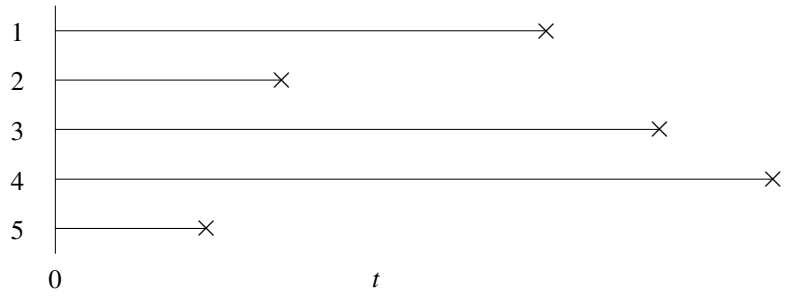

Figure 5.1: Complete data set with $n = 5$.

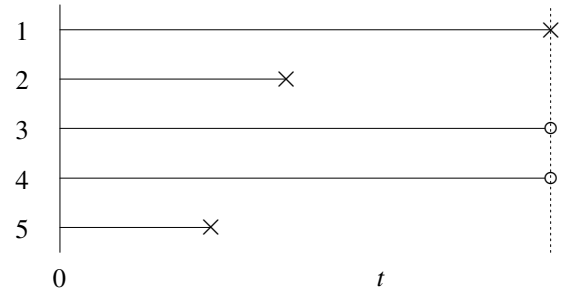

Figure 5.2: Type II right-censored data set with $n = 5$ and $r = 3$.

The second special case is Type I or *time* censoring. As shown in Figure 5.3, this corresponds to terminating the study at a particular time. The diagram shows a set of $n = 5$ items placed on a test simultaneously at $t = 0$ that is terminated at the time indicated by the dotted vertical line. For the realization illustrated in Figure 5.3, there are $r = 4$ observed failures. In Type I censoring, the number of failures r is random.

Finally, *random censoring* occurs when individual items are withdrawn from the test at any time during the study. Figure 5.4 illustrates a realization of a randomly right-censored life test with $n = 5$ items on test and $r = 2$ observed failures. It is usually assumed that the failure times and the censoring times are mutually independent random variables and that the probability distribution of the censoring times does not involve any unknown parameters from the failure time distribution. In other words, in a randomly censored data set, items cannot be more or less likely to be censored because they are at unusually high or low risk of failure.

Although other types of censoring exist, such as *left censoring* and *interval censoring*, the focus of this chapter will be on right censoring because it is the most common type of censoring. In the case of right censoring, the ratio r/n is the fraction of items which are observed to fail. When r/n is close to one, the data set is referred to as a *lightly censored* data set; when r/n is close to zero, the data set is referred to as a *heavily censored* data set. In the reliability setting, many data sets are heavily censored because the items have long lifetimes. In the biomedical setting, certain cancers have long remission times, resulting in heavily censored data sets.

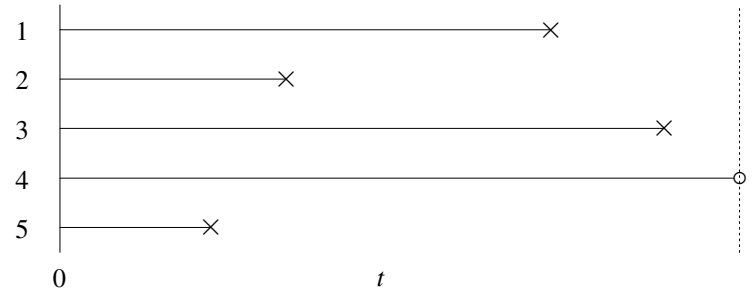

Figure 5.3: Type I right-censored data set with $n = 5$ and $r = 4$.

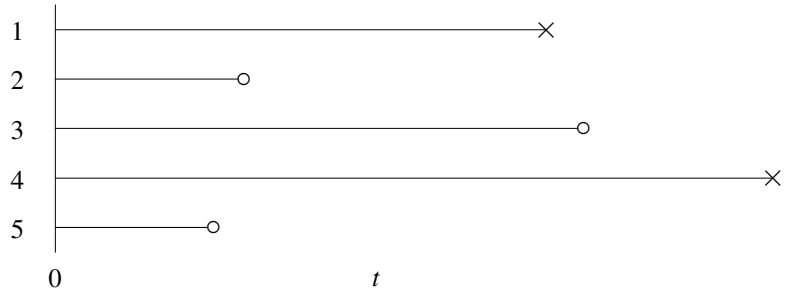

Figure 5.4: Randomly right-censored data set with $n = 5$ and $r = 2$.

Of the following three approaches to handling the problem of censoring, only one is both valid and practical. The first approach is to ignore all the censored values and to perform analysis only on those items that were observed to fail. Although this simplifies the mathematics involved, it is not a valid approach. If, for example, this approach is used on a right-censored data set, the analyst is discarding the right-censored values, and these are typically the items that have survived the longest. In this case, the analyst arrives at an overly pessimistic result concerning the lifetime distribution because the best items (that is, the right-censored observations) have been excluded from the analysis. A second approach is to wait for all the right-censored observations to fail. Although this approach is valid statistically, it is not practical. In an industrial setting, waiting for the last light bulb to burn out or the last machine to fail may take so long that the product being tested will not get to market in time. In a medical setting, waiting for the last patient to die from a particular disease may take decades. For these reasons, the proper approach is to handle censored observations probabilistically, including the censored values in the likelihood function.

The likelihood function for a censored data set can be written in several different equivalent forms. Let $t_1, t_2, \ldots, t_n$ be mutually independent observations denoting lifetimes sampled randomly from a population. The corresponding right-censoring times are denoted by $c_1, c_2, \ldots, c_n$. The t_i and c_i values are assumed to be independent, for $i = 1, 2, \ldots, n$. In the case of Type I right censoring, $c_1 = c_2 = \cdots = c_n = c$. The set U contains the indexes of the items that are observed to fail during the test (that is, the uncensored observations):

$$U = \{i \,|\, t_i \leq c_i\}.$$

The set C contains the indexes of the items whose failure time exceeds the corresponding censoring time (that is, those that are right censored):

$$C = \{\, i \,|\, t_i > c_i\}.$$

This notation, along with an important notion known as *alignment*, are illustrated in the next example.

Example 5.3 Consider the case of $n = 5$ items placed on test as indicated in Figure 5.5. Find the sets U and C.

Observe that the right-censored data set depicted in Figure 5.5, unlike the previous right-censored data sets with $n = 5$ items on test, does not have all of the items starting on test at time $t = 0$. This is quite common in practice. A software engineer, for example,

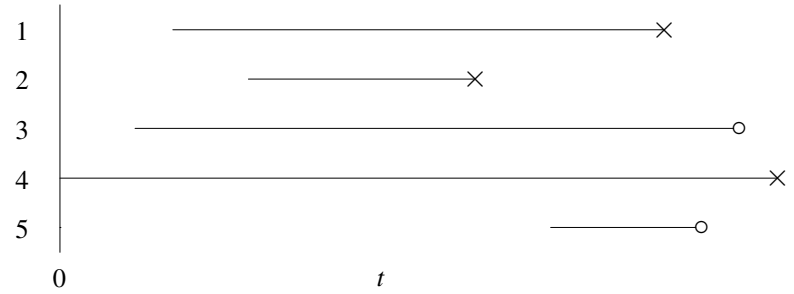

Figure 5.5: Randomly right-censored data set.

cannot get all customers to purchase a computer program at the same time; a medical researcher evaluating the time between first and second heart attacks cannot get all of the patients in the study to have their first heart attack at the same time; a casualty actuary cannot get all customers to purchase motorcycle insurance at the same time. In all cases, it is necessary to shift each data value back to a common origin. As long as there are not any changes to the items over the time window of observation, aligning the data values in this fashion is appropriate. Figure 5.6 displays the aligned data set. In this particular case, the first, second, and fourth items were observed to fail, and the failure times for the third and fifth items were right-censored. Therefore, the sets U and C are

$$U = \{1, 2, 4\} \qquad \text{and} \qquad C = \{3, 5\}.$$

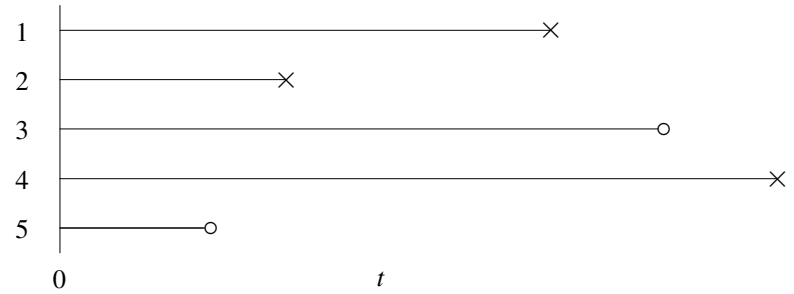

Figure 5.6: Aligned randomly right-censored data set.

The usual form for right-censored lifetime data is given by the pairs (x_i, δ_i), where $x_i = \min\{t_i, c_i\}$ and δ_i is a censoring indicator variable:

$$\delta_i = \begin{cases} 0 & t_i > c_i \\ 1 & t_i \leq c_i \end{cases}$$

for $i = 1, 2, \ldots, n$. The (x_i, δ_i) pairs can be reconstructed from the (t_i, c_i) pairs and vice versa. Hence, δ_i is 1 if the failure of item i is observed and 0 if the failure of item i is right censored, and

x_i is the failure time (when $\delta_i = 1$) or the censoring time (when $\delta_i = 0$). For the vector of unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$, ignoring a constant factor, the likelihood function is

$$L(\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta})^{\delta_i} S(x_i, \boldsymbol{\theta})^{1-\delta_i} = \prod_{i \in U} f(t_i, \boldsymbol{\theta}) \prod_{i \in C} S(c_i, \boldsymbol{\theta})$$

where $S(c_i, \boldsymbol{\theta})$ is the survivor function of the population distribution with parameters $\boldsymbol{\theta}$ evaluated at censoring time c_i, $i \in C$. The reason that the survivor function is the appropriate term in the likelihood function for a right-censored observation is that $S(c_i, \boldsymbol{\theta})$ is the probability that item i survives to c_i. The log likelihood function is

$$\ln L(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i \in U} \ln f(t_i, \boldsymbol{\theta}) + \sum_{i \in C} \ln S(c_i, \boldsymbol{\theta}),$$

or

$$\ln L(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i \in U} \ln f(x_i, \boldsymbol{\theta}) + \sum_{i \in C} \ln S(x_i, \boldsymbol{\theta}).$$

Since the probability density function is the product of the hazard function and the survivor function, the log likelihood function can be simplified to

$$\ln L(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i \in U} \ln h(x_i, \boldsymbol{\theta}) + \sum_{i \in U} \ln S(x_i, \boldsymbol{\theta}) + \sum_{i \in C} \ln S(x_i, \boldsymbol{\theta})$$

or

$$\ln L(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i \in U} \ln h(x_i, \boldsymbol{\theta}) + \sum_{i=1}^{n} \ln S(x_i, \boldsymbol{\theta}),$$

where the second summation now includes all n items on test. Finally, to write the log likelihood in terms of the hazard and cumulative hazard functions only,

$$\ln L(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i \in U} \ln h(x_i, \boldsymbol{\theta}) - \sum_{i=1}^{n} H(x_i, \boldsymbol{\theta}),$$

since $H(t) = -\ln S(t)$. The choice of which of these three expressions for the log likelihood to use for a particular distribution depends on the particular forms of $S(t)$, $f(t)$, $h(t)$, and $H(t)$. In other words, one of the distribution representations may possess a mathematical form that is advantageous over the others.

The next example will use the last version of the log likelihood function to find a maximum likelihood estimator and an asymptotically exact confidence interval for an unknown parameter.

Example 5.4 Consider a life test with n items on test with random right censoring and $r \geq 1$ observed failures. Assume that previous tests on these same items informs us that lifetimes of the items are drawn from a Rayleigh population with positive unknown parameter λ. Find the maximum likelihood estimator and construct an asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for λ.

The survivor function for the Rayleigh distribution is

$$S(t) = e^{-(\lambda t)^2} \qquad t \geq 0.$$

The associated cumulative hazard function and hazard function are

$$H(t) = -\ln S(t) = (\lambda t)^2 \qquad t \geq 0$$

and

$$h(t) = H'(t) = 2\lambda^2 t \qquad t \geq 0.$$

In the case of random right censoring, the log likelihood function is

$$\begin{aligned}\ln L(\boldsymbol{x}, \lambda) &= \sum_{i \in U} \ln h(x_i, \lambda) - \sum_{i=1}^{n} H(x_i, \lambda) \\ &= \sum_{i \in U} \ln\left(2\lambda^2 x_i\right) - \sum_{i=1}^{n} (\lambda x_i)^2 \\ &= r \ln 2 + 2r \ln \lambda + \sum_{i \in U} \ln x_i - \lambda^2 \sum_{i=1}^{n} x_i^2,\end{aligned}$$

where r is the number of observed failures. The single-element score vector can be found by differentiating the log likelihood function with respect to λ:

$$\frac{\partial \ln L(\boldsymbol{x}, \lambda)}{\partial \lambda} = \frac{2r}{\lambda} - 2\lambda \sum_{i=1}^{n} x_i^2.$$

Equating the score to zero and solving for λ yields the maximum likelihood estimator

$$\hat{\lambda} = \sqrt{\frac{r}{\sum_{i=1}^{n} x_i^2}}.$$

The second derivative of the log likelihood function is

$$\frac{\partial^2 \ln L(\boldsymbol{x}, \lambda)}{\partial \lambda^2} = -\frac{2r}{\lambda^2} - 2\sum_{i=1}^{n} x_i^2.$$

As an aside, the 1×1 Fisher information matrix

$$I(\lambda) = E\left[-\frac{\partial^2 \ln L(\boldsymbol{x}, \lambda)}{\partial \lambda^2}\right] = E\left[\frac{2r}{\lambda^2} + 2\sum_{i=1}^{n} x_i^2\right]$$

cannot be calculated without knowing the probability distribution of the censoring times. The observed information matrix, however, can be calculated as

$$O\left(\hat{\lambda}\right) = \left[-\frac{\partial^2 \ln L(\boldsymbol{x}, \lambda)}{\partial \lambda^2}\right]_{\lambda = \hat{\lambda}} = \frac{2r}{\hat{\lambda}^2} + 2\sum_{i=1}^{n} x_i^2 = 4\sum_{i=1}^{n} x_i^2.$$

For large values of n, we know that

$$\hat{\lambda} \overset{a}{\sim} N\left(\lambda, O^{-1}\left(\hat{\lambda}\right)\right)$$

or

$$\hat{\lambda} \overset{a}{\sim} N\left(\lambda, \left(4\sum_{i=1}^{n} x_i^2\right)^{-1}\right).$$

Standardizing by subtracting the population mean and dividing by the population standard deviation of $\hat{\lambda}$ gives

$$\frac{\hat{\lambda} - \lambda}{\left(4\sum_{i=1}^{n} x_i^2\right)^{-1/2}} \overset{a}{\sim} N(0, 1),$$

which implies that

$$\lim_{n\to\infty} P\left(-z_{\alpha/2} < \frac{\hat{\lambda}-\lambda}{\left(4\sum_{i=1}^{n} x_i^2\right)^{-1/2}} < z_{\alpha/2}\right) = 1-\alpha.$$

Performing the algebra required to isolate λ in the center of the inequality results in an asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for λ:

$$\hat{\lambda} - z_{\alpha/2}\left(4\sum_{i=1}^{n} x_i^2\right)^{-1/2} < \lambda < \hat{\lambda} + z_{\alpha/2}\left(4\sum_{i=1}^{n} x_i^2\right)^{-1/2}.$$

This confidence interval narrows as $\sum_{i=1}^{n} x_i^2$ increases. So placing a large number of items on test with a lightly censored data set with r/n close to one will result in a narrow confidence interval for λ.

To provide a numerical illustration, assume that the $n = 5$ items on a randomly right-censored life test with $r = 3$ observed failures illustrated in Figure 5.6 are

$$1.3,\ 0.6,\ 1.6^*,\ 1.9,\ 0.4^*,$$

where the superscript $*$ denotes a right-censored observation. For this data set,

$$\sum_{i=1}^{n} x_i^2 = 1.3^2 + 0.6^2 + 1.6^2 + 1.9^2 + 0.4^2 = 1.69 + 0.36 + 2.56 + 3.61 + 0.16 = 8.38.$$

The maximum likelihood estimate of λ is

$$\hat{\lambda} = \sqrt{\frac{r}{\sum_{i=1}^{n} x_i^2}} = \sqrt{\frac{3}{8.38}} = 0.598.$$

An asymptotically exact two-sided 95% confidence interval for λ is

$$0.598 - 1.96\,(4\cdot 8.38)^{-1/2} < \lambda < 0.598 + 1.96\,(4\cdot 8.38)^{-1/2}$$

or

$$0.260 < \lambda < 0.937.$$

To summarize the material introduced so far in this chapter, point estimators are statistics calculated from a data set to estimate an unknown parameter. Confidence intervals reflect the precision of a point estimator. The most common technique for determining a point estimator for an unknown parameter is maximum likelihood estimation, which involves finding the parameter value(s) that make the observed data values the most likely. The maximum likelihood estimators are usually found by using calculus to maximize the log likelihood function. Most population lifetime distributions do not have exact confidence intervals for unknown parameters, so the asymptotic properties of the likelihood function can be used to generate approximate confidence intervals for unknown parameters. Finally, many data sets in reliability are censored, which means that only a bound is known on the lifetime for one or more of the data values. The most common censoring mechanism is known as right censoring, where only a lower bound on the lifetime is known. The number of items on test is denoted by n and the number of observed failures is denoted by r.

The next section applies the techniques developed so far in this chapter to the exponential distribution.

5.4 Exponential Distribution

The exponential distribution is popular due to its tractability for parameter estimation and inference. The exponential distribution can be parameterized by either its population rate λ or its population mean $\mu = 1/\lambda$. Using the rate to parameterize the distribution, the survivor, density, hazard, and cumulative hazard functions are

$$S(t,\lambda) = e^{-\lambda t} \qquad f(t,\lambda) = \lambda e^{-\lambda t} \qquad h(t,\lambda) = \lambda \qquad H(t,\lambda) = \lambda t$$

for $t \geq 0$. Note that the unknown parameter λ has been added as an argument in these lifetime distribution representations because it is now also an argument in the likelihood function and is estimated from data.

All the analysis in this and subsequent sections assumes that a *random sample* of n items from a population has been placed on a test and subjected to typical environmental conditions. Equivalently, $t_1, t_2, \ldots, t_n$ are independent and identically distributed random lifetimes from a particular population distribution (exponential in this section). As with all statistical inference, care must be taken to ensure that a random sample of lifetimes is collected. Consequently, random numbers should be used to determine which n items to place on test. In a reliability setting, laboratory conditions should adequately mimic field conditions. Only representative items should be placed on test because items manufactured using a previous design may have a different failure pattern than those with the current design. This is more difficult in a biomedical setting because of inherent differences between patients.

Four classes of data sets (complete, Type II right censored, Type I right censored, and randomly right censored) are considered in separate subsections. In all cases, n is the number of items placed on test and r is the number of observed failures.

5.4.1 Complete Data Sets

A complete data set is typically the easiest to analyze because extensive analytical work exists for finding point and interval estimators for parameters. Also, by testing each item to failure, we have equal confidence in the fitted model in both the left-hand and right-hand tails of the distribution. A heavily right-censored data set, on the other hand, might fit well in the left-hand tail of the distribution where failures were observed, but we have less confidence in the right-hand tail of the distribution where there were few or no failures.

A complete data set consists of failure times $t_1, t_2, \ldots, t_n$. Although lowercase letters are used to denote the failure times here to be consistent with the notation for censoring times, the failure times are nonnegative random variables. The likelihood function can be written as a product of the probability density functions evaluated at the failure times:

$$L(\lambda) = \prod_{i=1}^{n} f(t_i, \lambda).$$

Note that the $\boldsymbol{t}$ argument has been left out of the likelihood expression for compactness. Using the last expression for the log likelihood function (adapted for a complete data set) from Section 5.3,

$$\ln L(\lambda) = \sum_{i=1}^{n} \big[\ln h(t_i, \lambda) - H(t_i, \lambda)\big].$$

For the exponential distribution, this is

$$\ln L(\lambda) = \sum_{i=1}^{n} \big[\ln \lambda - \lambda t_i\big] = n \ln \lambda - \lambda \sum_{i=1}^{n} t_i.$$

To determine the maximum likelihood estimator for λ, the single-element score vector

$$U(\lambda) = \frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i,$$

also known as the *score statistic*, is equated to zero and solved for λ, yielding

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i},$$

where the denominator is often referred to as the *total time on test*. Not surprisingly, the maximum likelihood estimator $\hat{\lambda}$ is the reciprocal of the sample mean.

Theorem 5.2 Let $t_1, t_2, \ldots, t_n$ be the observed values of n mutually independent and identically distributed exponential(λ) random variables. The maximum likelihood estimator of λ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}.$$

Example 5.5 A complete data set of $n = 23$ ball bearing failure times associated with testing the endurance of deep-groove ball bearings has been extensively studied. The failure times measured in 10^6 revolutions, ordered for readability, are

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84	51.96
54.12	55.56	67.80	68.64	68.64	68.88	84.12	93.12	98.64
		105.12	105.84	127.92	128.04	173.40.		

Notice that there is a single tied value of 68.64 million revolutions. Fit the exponential distribution to the $n = 23$ ball bearing failure times.

For this particular data set, the total time on test is $\sum_{i=1}^{n} t_i = 1661.16$ million revolutions, yielding the maximum likelihood estimate

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{23}{1661.16} = 0.01385$$

failure per 10^6 revolutions. The number of significant digits reported in the point estimate matches the number of digits in the data set. The value of the log likelihood function at the maximum likelihood estimate is $\ln L(\hat{\lambda}) = -121.435$, which will be used later in this chapter to compare the exponential and Weibull fits to this data set.

Figure 5.7 displays a graph of the *empirical survivor function*, which takes a downward step of $1/n = 1/23$ at each data value, along with the fitted exponential survivor function $S(t) = e^{-\hat{\lambda}t}$. Empirical and fitted distributions are traditionally compared by plotting the two the survivor functions on the same set of axes because the probability density function and hazard function suffer from the drawback of requiring the data to be divided into cells to plot the empirical distribution. It is apparent from this figure that the exponential distribution is a very poor fit. This particular data set was chosen for this example to illustrate one of the shortcomings of using the exponential distribution to model any data set without assessing the adequacy of the fit. Extreme caution must

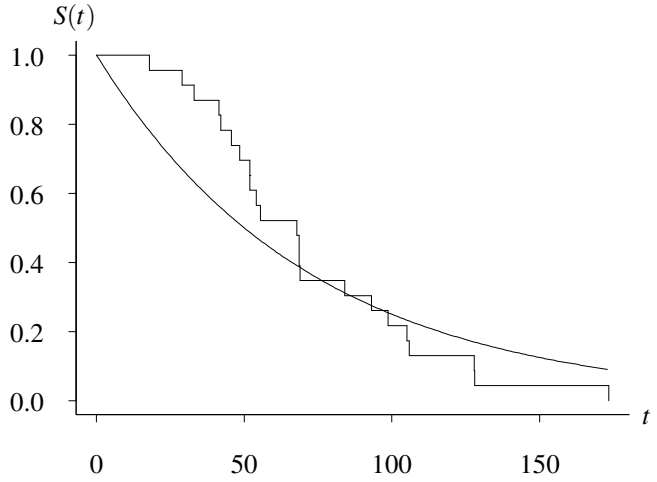

Figure 5.7: Empirical and exponential fitted survivor functions for the ball bearing data set.

be exercised when using the exponential distribution since, as indicated in Figure 5.7, the exponential distribution is not an adequate probability model for this data set.

There are two clues that the exponential distribution would perform poorly in this setting. First, we neglected to plot a histogram of the ball bearing failure times prior to fitting the exponential distribution. The histogram in Figure 5.8 indicates a nonzero mode to the population probability density function, implying that the exponential distribution is probably not going to be an adequate model. Second, knowing the physics of failure can be helpful in this case. Ball bearings typically fail by wearing out. When a ball bearing's diameter falls outside of a prescribed range, it is considered to be failed. This indicates that the hazard function for a ball bearing will probably increase over time, so a distribution with a monotone increasing hazard function from the IFR class

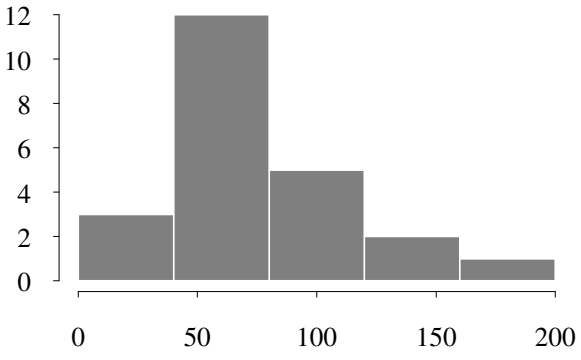

Figure 5.8: Histogram of the ball bearing failure times.

would be a better choice than the exponential distribution. As shown in the next section, the Weibull distribution provides a much better approximation to this particular data set. Since the exponential distribution can be fitted to any data set that has at least one observed failure, the adequacy of the model must always be assessed. The point and interval estimators associated with the exponential distribution are legitimate only if the data set is a random sample drawn from an exponential population. That is almost certainly *not* the case for this particular data set.

Information matrices. To find the information matrix associated with a complete data set from an exponential(λ) population, the derivative of the score statistic is required:

$$\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2}.$$

Taking the expected value of the negative of this quantity yields the 1×1 Fisher information matrix

$$I(\lambda) = E\left[\frac{-\partial^2 \ln L(\lambda)}{\partial \lambda^2}\right] = E\left[\frac{n}{\lambda^2}\right] = \frac{n}{\lambda^2}.$$

If the maximum likelihood estimator $\hat{\lambda}$ is used as an argument in the negative of the second partial derivative of the log likelihood function, the 1×1 observed information matrix is obtained:

$$O(\hat{\lambda}) = \left[\frac{-\partial^2 \ln L(\lambda)}{\partial \lambda^2}\right]_{\lambda=\hat{\lambda}} = \frac{n}{\hat{\lambda}^2} = \frac{\left(\sum_{i=1}^n t_i\right)^2}{n}.$$

Confidence interval for λ. Asymptotic confidence intervals for λ based on the likelihood ratio statistic or the observed information matrix are unnecessary for a complete data set because the sampling distribution of $\sum_{i=1}^n t_i$ is tractable. In particular, from Theorem 4.5,

$$2\lambda \sum_{i=1}^n t_i = \frac{2n\lambda}{\hat{\lambda}}$$

has the chi-square distribution with $2n$ degrees of freedom. Therefore, with probability $1-\alpha$,

$$\chi^2_{2n,\,1-\alpha/2} < \frac{2n\lambda}{\hat{\lambda}} < \chi^2_{2n,\,\alpha/2},$$

where $\chi^2_{2n,\,p}$ is the $(1-p)$th fractile of the chi-square distribution with $2n$ degrees of freedom. Performing the algebra required to isolate λ in the middle of the inequality yields an exact two-sided $100(1-\alpha)\%$ confidence interval for λ.

Theorem 5.3 Let $t_1, t_2, \ldots, t_n$ be the observed values of n mutually independent and identically distributed exponential(λ) random variables. Let $\hat{\lambda}$ denote the maximum likelihood estimator of λ. An exact two-sided $100(1-\alpha)\%$ confidence interval for λ is

$$\frac{\hat{\lambda}\chi^2_{2n,\,1-\alpha/2}}{2n} < \lambda < \frac{\hat{\lambda}\chi^2_{2n,\,\alpha/2}}{2n}.$$

A well-known example of a randomly right-censored data set is drawn from the biostatistical literature. The focus here is on determining an estimate of the remission rate of a complete data set of remission times for patients in a control group.

Example 5.6 A clinical trial is conducted to determine the effect of an experimental drug named 6–mercaptopurine (6–MP) on leukemia remission times. A sample of $n = 21$ leukemia patients is treated with 6–MP, and the remission times are recorded. There are $r = 9$ individuals for whom the remission time is observed, and the remission times for the remaining 12 individuals are randomly censored on the right. Letting an asterisk denote a right-censored observation, the remission times (in weeks) are

$$\begin{array}{cccccccccccc} 6 & 6 & 6 & 6^* & 7 & 9^* & 10 & 10^* & 11^* & 13 & 16 \\ 17^* & 19^* & 20^* & 22 & 23 & 25^* & 32^* & 32^* & 34^* & 35^*. \end{array}$$

In addition, 21 other leukemia patients are not given the drug, and they serve as a control group. For this group there is no censoring and the remission times are

$$\begin{array}{cccccccccccc} 1 & 1 & 2 & 2 & 3 & 4 & 4 & 5 & 5 & 8 & 8 \\ 8 & 8 & 11 & 11 & 12 & 12 & 15 & 17 & 22 & 23. \end{array}$$

This data set illustrates the simplest possible use of a covariate for modeling: a single binary covariate indicating the group to which each data value belongs. Fit the exponential distribution to the $n = 21$ remission times in the control group of the 6–MP clinical trial. Give a point estimator and a 95% confidence interval for λ.

Having learned our lesson from the previous example, we begin by drawing a histogram of the remission times, which is displayed in Figure 5.9. The shape of the histogram reveals significant random sampling variability which can be attributed to the small ($n = 21$) number patients in the control group. Modeling the remission times with a probability distribution that has a mode of zero seems reasonable based on the shape of the histogram, so we will proceed with fitting the exponential distribution.

The total time on test is

$$\sum_{i=1}^{21} t_i = 182$$

weeks. The maximum likelihood estimate is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{21}{182} = 0.12$$

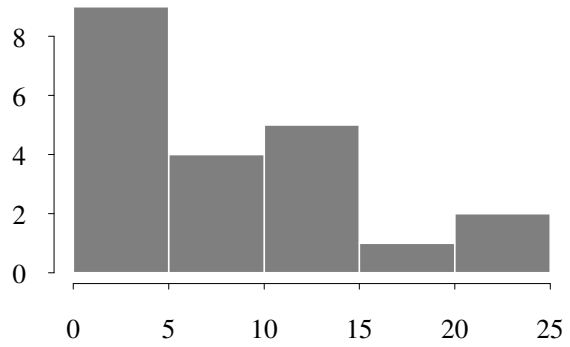

Figure 5.9: Histogram of the leukemia remission times.

remission per week. Figure 5.10 shows the empirical survivor function, which takes a downward step of $1/n = 1/21$ at each data point, along with the survivor function for the fitted exponential distribution. In spite of the discrete nature of the data, the excessive number of ties, and the fact that the number of patients in the control group is rather small, the exponential distribution does a reasonable job of approximating the empirical survivor function.

The observed information matrix is

$$O(\hat{\lambda}) = \left[\frac{-\partial^2 \ln L(\lambda)}{\partial \lambda^2}\right]_{\lambda=\hat{\lambda}} = \frac{\left(\sum_{i=1}^{n} t_i\right)^2}{n} = \frac{182^2}{21} = 1577.$$

Since the data set is complete, an exact two-sided 95% confidence interval for the failure rate of the distribution can be determined. Since $\chi^2_{42,0.975} = 26.0$ and $\chi^2_{42,0.025} = 61.8$, the formula for the confidence interval

$$\frac{\hat{\lambda}\chi^2_{2n,1-\alpha/2}}{2n} < \lambda < \frac{\hat{\lambda}\chi^2_{2n,\alpha/2}}{2n}$$

becomes

$$\frac{(0.12)(26.0)}{42} < \lambda < \frac{(0.12)(61.8)}{42}$$

or

$$0.071 < \lambda < 0.17.$$

The involvement of the non-symmetric chi-square distribution in this confidence interval means that the interval is not symmetric about the maximum likelihood estimate. For this and subsequent examples, intermediate calculations involving numeric quantities, such as critical values or total time on test values, are performed to as much precision as possible, then final values are reported using only significant digits.

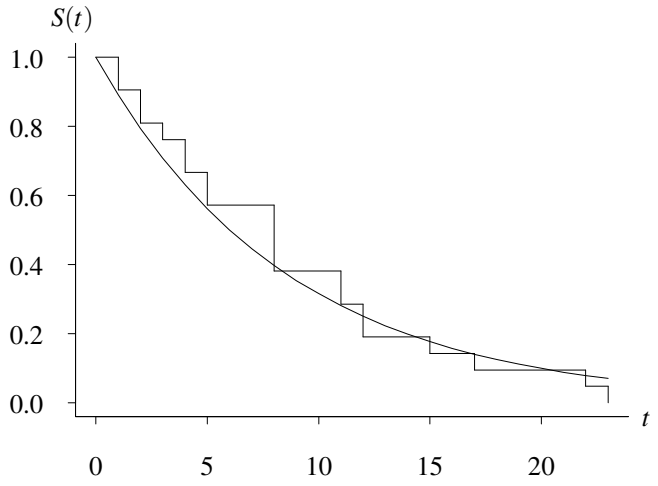

Figure 5.10: Empirical and exponential fitted survivor functions for the 6–MP control group.

The R code given below calculates the maximum likelihood estimator $\hat{\lambda}$, calculates the endpoints of the exact two-sided confidence interval for λ, and conducts the Kolmogorov–Smirnov goodness-of-fit test. $p = 0.55$.

```
x  = c(1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15,
       17, 22, 23)
n  = length(x)
l  = n / sum(x)
lo = l * qchisq(0.025, 2 * n) / (2 * n)
hi = l * qchisq(0.975, 2 * n) / (2 * n)
p  = ks.test(x, "pexp", l, exact = FALSE)$p.value
print(c(lo, l, hi, p))
```

Since the p-value for the Kolmogorov–Smirnov test is $p = 0.55$, there is not sufficient evidence in the data to reject the null hypothesis that the data values were drawn from an exponential population. This conclusion is consistent with the empirical and exponential fitted survivor functions in Figure 5.10. The exponential distribution provides a reasonable approximation to the leukemia remission times.

The importance of assessing model adequacy applies to all fitted distributions—not just the exponential distribution. Furthermore, if a modeler knows the failure physics (for example, fatigue crack growth) underlying a process, then an appropriate probability model that is consistent with the failure physics should be chosen.

So far we have fitted the exponential distribution to two complete data sets: the ball bearing failure times from Example 5.5 and the 6–MP remission times for the control group from Example 5.6. We visually assessed the two fits in Figures 5.7 and 5.10 by comparing the empirical survivor function, which takes a downward step of $1/n$ at each data value, with the fitted survivor function $S(t) = e^{-\hat{\lambda}t}$ and concluded that the exponential distribution did a very poor job of approximating the ball bearing failure times and a (barely) adequate job of approximating the remission times of the patients in the control group of the 6–MP clinical trial. This visual assessment was subjective and was followed by a formal goodness-of-fit test in order to draw these conclusions for the 6–MP remission times.

Confidence intervals for measures other than λ. It is possible to find point and interval estimators for measures other than λ by using the invariance property for maximum likelihood estimators and by rearranging the confidence interval formula. Define

$$L = \frac{\hat{\lambda}\chi^2_{2n,\,1-\alpha/2}}{2n} \qquad \text{and} \qquad U = \frac{\hat{\lambda}\chi^2_{2n,\,\alpha/2}}{2n}$$

as the lower and upper bounds on the exact two-sided $100(1-\alpha)\%$ confidence interval for λ. If the measure of interest is $\mu = 1/\lambda$, for example, then the point estimator is the sample mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} t_i$. Rearranging the confidence interval

$$L < \lambda < U$$

by taking reciprocals yields the exact two-sided $100(1-\alpha)\%$ confidence interval for μ:

$$\frac{1}{U} < \mu < \frac{1}{L}.$$

As a second example, consider the probability of survival to a fixed time t, $S(t) = e^{-\lambda t}$. By the invariance property of maximum likelihood estimators, the maximum likelihood estimator for the survivor function at time t is

$$\hat{S}(t) = e^{-\hat{\lambda} t}.$$

A confidence interval for $S(t)$, on the other hand, can be found by rearranging the confidence interval

$$L < \lambda < U$$

in the following fashion:

$$-U < -\lambda < -L$$
$$e^{-Ut} < e^{-\lambda t} < e^{-Lt}$$
$$e^{-Ut} < S(t) < e^{-Lt}.$$

These formulas for point and interval estimates for quantities other than λ are illustrated next for a complete data set that is assumed to be drawn from an exponential population.

Example 5.7 Assuming that the exponential distribution is an appropriate model for the remission times in the control group of the 6–MP clinical trial, find point estimators and exact two-sided 95% confidence intervals for the mean remission time and the probability that a patient in the control group has a remission time that exceeds 10 weeks.

The point estimators in this case are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{182}{21} = 8.7$$

weeks and

$$\hat{S}(t) = e^{-\hat{\lambda} t},$$

which is $\hat{S}(10) = e^{-(0.12)(10)} = 0.32$. The values of L and U for the exact two-sided 95% confidence interval for λ from the previous example are $L = 0.071$ and $U = 0.17$. Finding a confidence interval for the population mean requires taking reciprocals of these limits:

$$\frac{1}{0.17} < \mu < \frac{1}{0.071}$$
$$5.9 < \mu < 14.$$

An exact two-sided 95% confidence interval for $S(100)$ using the formula

$$e^{-Ut} < S(t) < e^{-Lt}$$

results in

$$e^{-(0.17)(10)} < S(100) < e^{-(0.071)(10)}$$

or

$$0.18 < S(10) < 0.49.$$

Although the manipulation of the confidence interval for λ is performed here in the case of a complete data set, these techniques may also be applied to any of the right-censoring mechanisms to be described in the next three subsections.

5.4.2 Type II Censored Data Sets

A life test of n items that is terminated when r failures have occurred produces a Type II right-censored data set. The previous subsection is a special case of Type II censoring when $r = n$. As before, assume that the failure times are $t_1, t_2, \ldots, t_n$, the test is terminated upon the rth ordered failure, the censoring times are $c_1 = c_2 = \cdots = c_n = t_{(r)}$ for all items, and $x_i = \min\{t_i, c_i\}$ for $i = 1, 2, \ldots, n$.

Since $h(t, \lambda) = \lambda$ and $H(t, \lambda) = \lambda t$ for $t \geq 0$, the log likelihood function is

$$\ln L(\lambda) = \sum_{i \in U} \ln h(x_i, \lambda) - \sum_{i=1}^{n} H(x_i, \lambda) = r \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

because there are r observed failures. The expression

$$\sum_{i=1}^{n} x_i = \sum_{i \in U} t_i + \sum_{i \in C} c_i = \sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)},$$

where $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ are the order statistics of the observed failure times, is the total time on test. It represents the total accumulated time that the n items accrue while on test.

To determine the maximum likelihood estimator, the log likelihood function is differentiated with respect to λ,

$$U(\lambda) = \frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{r}{\lambda} - \sum_{i=1}^{n} x_i$$

and is equated to zero, yielding the maximum likelihood estimator.

Theorem 5.4 Let $t_1, t_2, \ldots, t_n$ be the observed values of n mutually independent and identically distributed exponential(λ) random variables. The associated test is terminated at time $t_{(r)}$ (Type II right censoring) for $r \geq 1$. The censoring times are $c_1 = c_2 = \cdots = c_n = t_{(r)}$ for all items, and $x_i = \min\{t_i, c_i\}$ for $i = 1, 2, \ldots, n$. The maximum likelihood estimator of λ is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} x_i}.$$

So the maximum likelihood estimator of the failure rate is the ratio of the number of observed failures to the total time on test. The second partial derivative of the log likelihood function is

$$\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} = -\frac{r}{\lambda^2},$$

so the information matrix is

$$I(\lambda) = E\left[\frac{-\partial^2 \ln L(\lambda)}{\partial \lambda^2}\right] = \frac{r}{\lambda^2},$$

and the observed information matrix is

$$O(\hat{\lambda}) = \left[\frac{-\partial^2 \ln L(\lambda)}{\partial \lambda^2}\right]_{\lambda=\hat{\lambda}} = \frac{r}{\hat{\lambda}^2} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{r}.$$

Exact confidence intervals and hypothesis tests concerning λ can be derived by using the result

$$2\lambda \sum_{i=1}^{n} x_i = \frac{2r\lambda}{\hat{\lambda}} \sim \chi^2(2r),$$

where $\chi^2(2r)$ is the chi-square distribution with $2r$ degrees of freedom. This result can be proved in a similar fashion to Theorem 4.5 of the exponential distribution from Section 4.2. Using this fact, an exact two-sided confidence interval for λ can be constructed in a similar fashion to that for a complete data set. It can be stated with probability $1-\alpha$ that

$$\chi^2_{2r,\,1-\alpha/2} < \frac{2r\lambda}{\hat{\lambda}} < \chi^2_{2r,\,\alpha/2}.$$

Rearranging terms yields an exact two-sided $100(1-\alpha)\%$ confidence interval for the failure rate λ.

Theorem 5.5 Let $t_1, t_2, \ldots, t_n$ be the observed values of n mutually independent and identically distributed exponential(λ) random variables. The associated test is terminated at time $t_{(r)}$ (Type II right censoring) for $r \geq 1$. The censoring times are $c_1 = c_2 = \cdots = c_n = t_{(r)}$ for all items, and $x_i = \min\{t_i, c_i\}$ for $i = 1, 2, \ldots, n$. An exact two-sided $100(1-\alpha)\%$ confidence interval for the failure rate λ is

$$\frac{\hat{\lambda}\chi^2_{2r,\,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r,\,\alpha/2}}{2r}.$$

Example 5.8 A Type II right-censored data set of $n = 15$ automotive a/c switches has been collected. The test was terminated when the fifth failure occurred. The $r = 5$ ordered observed failure times measured in number of cycles are

$$t_{(1)} = 1410, \qquad t_{(2)} = 1872, \qquad t_{(3)} = 3138, \qquad t_{(4)} = 4218, \qquad t_{(5)} = 6971.$$

The remaining 10 automotive a/c switches are right-censored at 6971 cycles. Any parametric model that is fitted to this data set is only considered valid from 0 to 6971 cycles unless there is some evidence (perhaps from previous test results) that indicates that the parametric model is valid beyond 6971 cycles. Fit the exponential distribution to this data set and give point and interval estimates for the failure rate and the mean time to failure.

A diagram that can be helpful in visualizing lifetime data sets is given in Figure 5.11. The top five horizontal lines ending with $\times$ denote the $r = 5$ observed failures and the bottom $n - r = 15 - 5 = 10$ horizontal lines ending with $\circ$ denote the right-censored observations at 6971 cycles. Each of the right-censored observations will have an unseen $\times$ somewhere to the right of the censoring time indicated by the $\circ$. It is a worthwhile thought experiment to imagine where those $\times$s might occur for each right-censored observation in this particular data set. Once you have visualized the approximate positions of the ten right-censored failure times, try to guess the approximate population mean of the 15 failure times.

For this particular data set, the total time on test is $\sum_{i=1}^{n} x_i = 87{,}319$ cycles, yielding a maximum likelihood estimate

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} x_i} = \frac{5}{87{,}319} = 0.00005726$$

failure per cycle. Equivalently, the maximum likelihood estimate of the population mean time to failure is

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{r} = \frac{87{,}319}{5} = 17{,}464$$

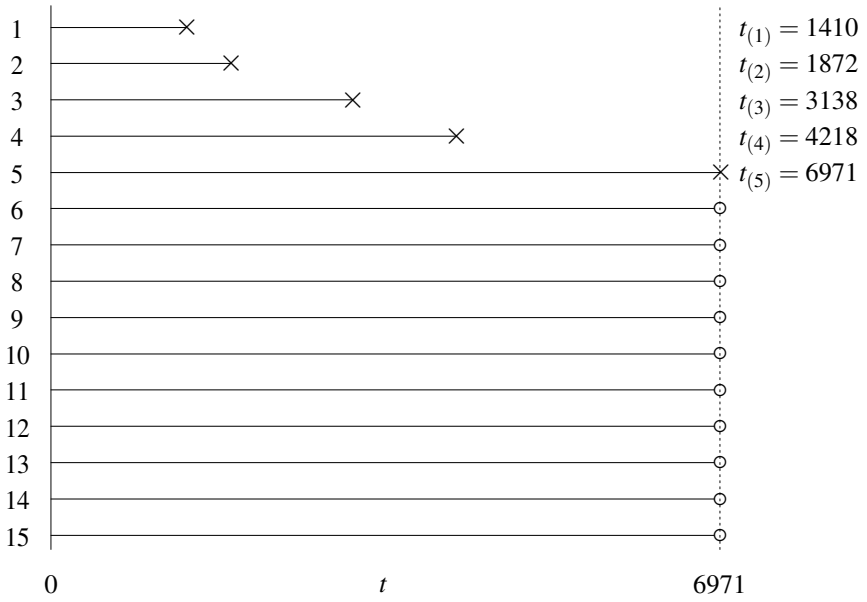

Figure 5.11: Automotive switches failure and censoring times with $n = 5$ and $r = 3$.

cycles. Notice that the estimated mean time to failure exceeds the largest observed failure time, $t_{(5)} = 6971$. As long as there is evidence, perhaps from previous testing on identical or similar automotive switches, to support the exponential failure time distribution, this estimate of the population mean time to failure is meaningful. Figure 5.12

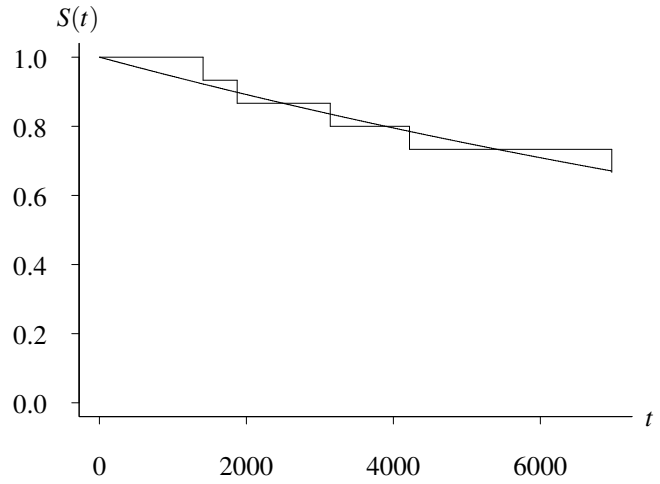

Figure 5.12: Empirical and exponential fitted survivor functions for the a/c switch data set.

shows the empirical survivor function (which takes downward steps of $1/n = 1/15$ at each of the five observed failure times) and the associated fitted exponential survivor function. In this case, the exponential distribution appears to adequately model the lifetimes through 6971 cycles. The confidence intervals given below are only exact when the data values are drawn from an exponential population, so assessing the fit is a crucial part of data analysis. Assessing the adequacy of the fit is more difficult for a right-censored data set because it is impossible to determine what the lifetime distribution looks like after the last observed failure time (6971 cycles for this data set) unless previous test results support the exponential model.

The observed information matrix based on using the failure rate as the unknown parameter is

$$O(\hat{\lambda}) = \left[\frac{-\partial^2 \ln L(\lambda)}{\partial \lambda^2}\right]_{\lambda=\hat{\lambda}} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{r} = \frac{(87{,}319)^2}{5} = 1{,}525{,}000{,}000.$$

Since the data set is Type II right censored, an exact two-sided 95% confidence interval for the failure rate of the distribution can be determined. Using the chi-square critical values, $\chi^2_{10,0.975} = 3.247$ and $\chi^2_{10,0.025} = 20.49$, the formula for the confidence interval

$$\frac{\hat{\lambda}\chi^2_{2r,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r,\alpha/2}}{2r}$$

becomes

$$\frac{(0.00005726)(3.247)}{10} < \lambda < \frac{(0.00005726)(20.49)}{10}$$

or

$$0.00001859 < \lambda < 0.0001173.$$

Taking reciprocals, this is equivalent to an exact two-sided 95% confidence interval for the population mean number of cycles to failure of

$$8526 < \mu < 53{,}785.$$

Not surprisingly, with only $r = 5$ observed failures, this is a rather wide confidence interval for μ, and hence there is not as much precision as in the case of the 6–MP control group data, in which there were $n = r = 21$ observed remission times. The R code below calculates the point estimates for λ and μ and the associated exact two-sided 95% confidence intervals.

```
n      = 15
r      = 5
x      = c(1410, 1872, 3138, 4218, 6971, rep(6971, n - r))
lam    = r / sum(x)
mu     = 1 / lam
lam.lo = lam * qchisq(0.025, 2 * r) / (2 * r)
lam.hi = lam * qchisq(0.975, 2 * r) / (2 * r)
mu.lo  = 1 / lam.hi
mu.hi  = 1 / lam.lo
```

Hypothesis testing, which is the rough equivalent of interval estimation, is also possible in the case of Type II censoring because the sampling distribution of $2\lambda\sum_{i=1}^{n} x_i$ is tractable. Some aspects of hypothesis testing in the setting of Type II censoring, such as the alternative hypothesis, one- and two-tailed tests, and p-values are illustrated in the next example. The example shows how a life test can be used to check a manufacturer's claimed mean time to failure.

Example 5.9 The producer of the automotive switches tested in the previous example claims that the population mean time to failure of their switches is $\mu = 100{,}000$ cycles. Is there enough evidence in the data set of 15 switches placed on test to conclude that the population mean time to failure is less than 100,000 cycles? Assume that the automotive switch lifetimes are exponentially distributed.

The producer's claim is certainly suspect because the maximum likelihood estimator for the population mean time to failure is only $\hat{\mu} = 17{,}464$ from the previous example. The null and alternative hypotheses for the hypothesis test are

$$H_0 : \mu = 100{,}000$$
$$H_1 : \mu < 100{,}000$$

or, equivalently,

$$H_0 : \lambda = 0.00001$$
$$H_1 : \lambda > 0.00001$$

in terms of the failure rate. So the hypothesis test being conducted here is to determine whether there is statistically significant evidence in the data set to conclude that the population mean time to failure of the switches is less then 100,000 cycles. Since small values of $\sum_{i=1}^{n} x_i$ lead to rejecting H_0, the attained level of significance (p-value) is

$$p = P\left(\sum_{i=1}^{n} x_i < 87{,}319 \,\middle|\, \lambda = 0.00001\right).$$

Since $2\lambda\sum_{i=1}^{n} x_i \sim \chi^2(2r)$, the p-value, when H_0 is true, is

$$\begin{aligned} p &= P\left((2)(0.00001)\sum_{i=1}^{n} x_i < (2)(0.00001)(87{,}319)\right) \\ &= P\left(\chi^2(10) < 1.746\right) \\ &= 0.002. \end{aligned}$$

This p-value can be calculated with the following R statements.

```
n = 15
r = 5
x = c(1410, 1872, 3138, 4218, 6971, rep(6971, n - r))
pchisq(2 * 0.00001 * sum(x), 2 * r)
```

Although the number of observed failures is small, there is adequate evidence from this data set to conclude that the population mean number of cycles to failure is less than 100,000 (for example, the null hypothesis can be rejected at significance levels $\alpha = 0.10$, 0.05, and 0.01). We conclude that the manufacturer is probably exaggerating the magnitude of the population mean time to failure based on this hypothesis test.

The fact that the distribution of $2\lambda\sum_{i=1}^{n} x_i = 2r\lambda/\hat{\lambda}$ is independent of n implies that $\hat{\lambda}$ has the same precision in a test of r items tested until all have failed as that for a test of n items tested until r items have failed. So the justification for obtaining a Type II censored data set over a complete data set is time savings. The additional costs associated with this time savings are the additional $n-r$ test stands and the additional $n-r$ items to place on test.

If a limited number of test stands are available for testing, the only way to speed up the test is to perform a test with replacement in which failed items are immediately replaced with new items. This will decrease the expected time to complete the test, which is terminated when r of the items fail. The sequence of failures in this case is a Poisson process with rate $n\lambda$.

Although the inference for Type II censoring is tractable, the unfortunate consequence is that the time to complete the test is a random variable. Constraints on the time to run a life test may make a Type I censored data set more practical.

5.4.3 Type I Censored Data Sets

The analysis for Type I censored data sets is similar to that for the Type II censoring case. The test is terminated at time c. The censoring times for each item on test are the same: $c_1 = c_2 = \cdots = c_n = c$. The number of observed failures, r, is a random variable. The total time on test in this case is

$$\sum_{i=1}^{n} x_i = \sum_{i \in U} t_i + \sum_{i \in C} c_i = \sum_{i=1}^{r} t_{(i)} + (n-r)c.$$

As before, the log likelihood function is

$$\ln L(\lambda) = \sum_{i \in U} \ln h(x_i, \lambda) - \sum_{i=1}^{n} H(x_i, \lambda) = r \ln \lambda - \lambda \sum_{i=1}^{n} x_i,$$

and the score statistic is

$$U(\lambda) = \frac{r}{\lambda} - \sum_{i=1}^{n} x_i.$$

The maximum likelihood estimator for $r > 0$ is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} x_i},$$

the information matrix is

$$I(\lambda) = \frac{r}{\lambda^2},$$

and the observed information matrix is

$$O(\hat{\lambda}) = \frac{r}{\hat{\lambda}^2}.$$

The functional form of the maximum likelihood estimator is identical to the Type II censoring case. For identical values of r, Type I censoring has a larger total time on test $\sum_{i=1}^{n} x_i$ than the corresponding Type II censoring case because a Type I test ends *between* failures r and $r+1$. Thus the expected value of $\hat{\lambda}$ is smaller for Type I censoring than for Type II censoring. One problem that arises with Type I censoring is that the sampling distribution of $\sum_{i=1}^{n} x_i$ is no longer tractable, so an exact confidence interval for λ has not been established. Although many more complicated methods exist, one of the best approximation methods is to assume that $2\lambda\sum_{i=1}^{n} x_i$ has the chi-square distribution with $2r+1$ degrees of freedom. This approximation, illustrated in Figure 5.13, is based on the fact

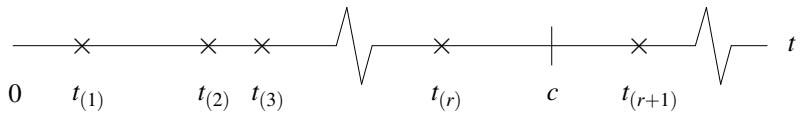

Figure 5.13: Approximation technique for confidence intervals for Type I censoring.

that if $c = t_{(r)}$, then $2\lambda\sum_{i=1}^{n} x_i \sim \chi^2(2r)$, and if $c = t_{(r+1)}$, then $2\lambda\sum_{i=1}^{n} x_i \sim \chi^2(2r+2)$. Since c is between $t_{(r)}$ and $t_{(r+1)}$, $2\lambda\sum_{i=1}^{n} x_i$ will be approximately chi-square with $2r+1$ degrees of freedom. This constitutes a proof, after a little algebra, of the following result.

Theorem 5.6 Let $t_1, t_2, \ldots, t_n$ be the observed values of n mutually independent and identically distributed exponential(λ) random variables. The associated test is terminated at time c (Type I right censoring) for some positive real number c. The censoring times are $c_1 = c_2 = \cdots = c_n = c$ for all items, and $x_i = \min\{t_i, c_i\}$ for $i = 1, 2, \ldots, n$. The maximum likelihood estimator for λ is

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

and an approximate two-sided $100(1-\alpha)\%$ confidence interval for the failure rate λ is

$$\frac{\hat{\lambda}\chi^2_{2r+1,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r+1,\alpha/2}}{2r}.$$

Example 5.10 A life test of $n = 100$ light bulbs is run for $c = 5000$ hours. Failed items are not replaced upon failure in this Type I right censored data set. If the total time on test is $\sum_{i=1}^{n} x_i = 384{,}968$ hours, and $r = 32$ failures are observed, find a point and interval estimator for the failure rate.

It is impossible to check to see whether the exponential distribution is an appropriate model for the light bulb failure times from the problem statement because the actual failure times are not given. Assuming that the exponential model is appropriate, the maximum likelihood estimate for the failure rate is

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} x_i} = \frac{32}{384{,}968} = 0.0000831$$

failure per hour, or, equivalently, the maximum likelihood estimate for the population mean time to failure is its reciprocal, 12,030 hours. To obtain an approximate 95% confidence interval for the failure rate, the chi-square critical values for $2r+1 = 65$ degrees of freedom must be determined. These critical values are $\chi^2_{65,0.975} = 44.60$ and $\chi^2_{65,0.025} = 89.18$. The approximate two-sided $100(1-\alpha)\%$ confidence interval for λ

$$\frac{\hat{\lambda}\chi^2_{2r+1,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r+1,\alpha/2}}{2r}$$

becomes

$$\frac{(0.0000831)(44.60)}{64} < \lambda < \frac{(0.0000831)(89.18)}{64}$$

or

$$0.0000579 < \lambda < 0.000116.$$

Taking reciprocals, this is equivalent to an approximate 95% confidence interval for the population mean number of cycles to failure of

$$8630 < \mu < 17{,}260.$$

The R statements below compute the point and interval estimates.

```
n      = 100
r      = 32
ttt    = 384968
lam    = r / ttt
mu     = 1 / lam
lam.lo = lam * qchisq(0.025, 2 * r + 1) / (2 * r)
lam.hi = lam * qchisq(0.975, 2 * r + 1) / (2 * r)
mu.lo  = 1 / lam.hi
mu.hi  = 1 / lam.lo
```

5.4.4 Randomly Censored Data Sets

Many of the examples that have the random censoring mechanism for which the failure times $t_1, t_2, \ldots, t_n$ and the censoring times $c_1, c_2, \ldots, c_n$ are independent random variables are from biostatistics. Random censoring occurs frequently in biostatistics because it is not always possible to control the time patients enter and exit the study. The log likelihood function, score statistic, information matrix, and observed information matrix are the same as in the Type I censoring case. The total time on test is now simply

$$\sum_{i=1}^{n} x_i = \sum_{i \in U} t_i + \sum_{i \in C} c_i.$$

The sampling distribution of $\sum_{i=1}^{n} x_i$ is more complicated in this case, so asymptotic properties must be relied on to determine approximate confidence intervals for λ. In the example that follows, three different approximation procedures for determining a confidence interval for λ are illustrated.

The first technique is based on an approximation to a result that holds exactly in the Type II censoring case: $2\lambda \sum_{i=1}^{n} x_i \sim \chi^2(2r)$. The second technique is based on the likelihood ratio statistic, where $-2[\ln L(\lambda) - \ln L(\hat{\lambda})]$ is asymptotically chi-square with 1 degree of freedom. The third technique is based on the fact that the maximum likelihood estimator $\hat{\lambda}$ is asymptotically normal with population mean λ and a population variance that is the inverse of the observed information matrix. Since this third technique results in a symmetric confidence interval, it should only be used with large sample sizes.

Example 5.11 Find the maximum likelihood estimate and three approximate 95% confidence intervals for the remission rate λ for the treatment group (those who received the drug 6–MP) in the leukemia study described in Example 5.6.

For this data set, there are $n = 21$ individuals on test and $r = 9$ observed failures. A "failure" for this data set is the end of a remission period. The total time on test for this

data set is $\sum_{i=1}^{n} x_i = 359$ weeks. The log likelihood function is

$$\ln L(\lambda) = r \ln \lambda - \lambda \sum_{i=1}^{n} x_i = 9 \ln \lambda - 359\lambda.$$

As shown by the vertical dashed line in Figure 5.14, this function is maximized at

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} x_i} = \frac{9}{359} = 0.0251$$

remission per week. The maximum likelihood estimate of the expected remission time is $\hat{\mu} = 359/9 = 39.9$ weeks. The value of the log likelihood function at the maximum likelihood estimate is $\ln L(\hat{\lambda}) = -42.17$, as indicated by the horizontal dashed line in Figure 5.14. The observed information matrix is

$$O(\hat{\lambda}) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{r} = \frac{(359)^2}{9} = 14{,}320.$$

The three approximation techniques for determining a confidence interval for λ are outlined next. Under the assumption that $2\lambda \sum_{i=1}^{n} x_i$ is approximately chi-square with $2r$ degrees of freedom (this is satisfied exactly in the Type II censoring case), an approximate two-sided $100(1-\alpha)\%$ confidence interval for λ is

$$\frac{\hat{\lambda}\chi^2_{2r,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r,\alpha/2}}{2r},$$

which, for the 6–MP treatment group remission times with $\alpha = 0.05$, is

$$\frac{(9)(8.23)}{(359)(18)} < \lambda < \frac{(9)(31.53)}{(359)(18)}$$

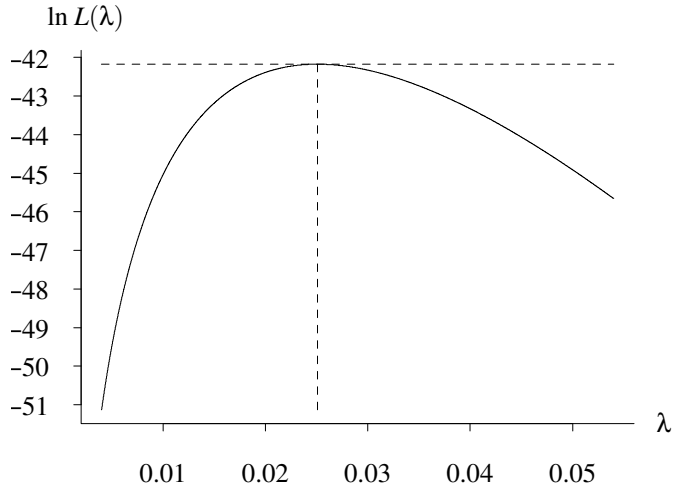

Figure 5.14: Log likelihood function for the 6–MP treatment group.

because $\chi^2_{18,0.975} = 8.23$ and $\chi^2_{18,0.025} = 31.53$, or

$$0.0115 < \lambda < 0.0439.$$

The second approximate confidence interval for λ is based on the likelihood ratio statistic, $-2[\ln L(\lambda) - \ln L(\hat{\lambda})]$, which is asymptotically chi-square with 1 degree of freedom. Thus, with probability $1-\alpha$, the inequality

$$-2\left[\ln L(\lambda) - \ln L\left(\hat{\lambda}\right)\right] < \chi^2_{1,\alpha}$$

is approximately satisfied. For the 6–MP remission times in the treatment group and $\alpha = 0.05$, this can be rearranged as

$$\ln L(\lambda) > \ln L\left(\hat{\lambda}\right) - \frac{3.84}{2}$$

because $\chi^2_{1,0.05} = 3.84$, or

$$\ln L(\lambda) > -42.17 - \frac{3.84}{2}.$$

As shown by the horizontal dashed lines in Figure 5.15, this corresponds to all values of λ for which the log likelihood function is within $3.84/2 = 1.92$ units of its largest value. The inequality reduces to

$$9\ln\lambda - 359\lambda > -42.17 - 1.92,$$

which can be solved numerically to determine the endpoints. Many computer languages have an equation solver that can determine the two λ values satisfying

$$9\ln\lambda - 359\lambda = -42.17 - 1.92 = -44.09.$$

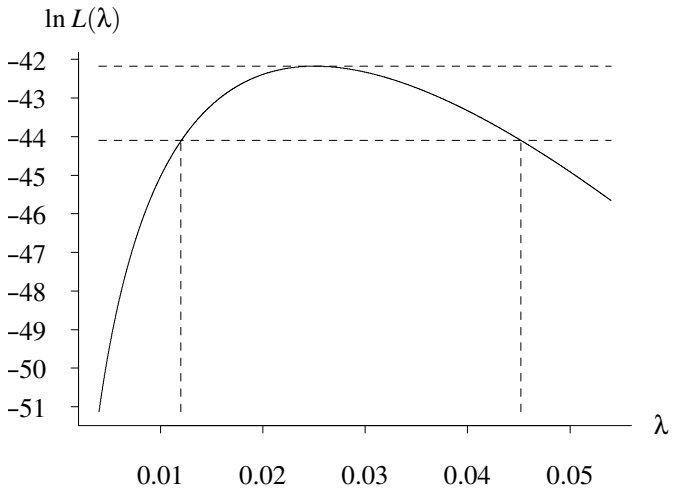

Figure 5.15: Log likelihood function and 95% confidence limits for λ for the 6–MP treatment group.

In this particular example, the approximate two-sided confidence interval for λ is

$$0.0120 < \lambda < 0.0452,$$

which is shifted slightly to the right of the previous confidence interval. The lower and upper bounds for this confidence interval are indicated by the vertical dashed lines in Figure 5.15.

The final confidence interval for λ is based on the fact that the sampling distribution of $\hat{\lambda}$ is asymptotically normal with population mean λ and population variance $I(\lambda)^{-1}$. Replacing $I(\lambda)$ by the observed information matrix $O(\hat{\lambda})$, with approximate probability $1-\alpha$,

$$-z_{\alpha/2} < \frac{\hat{\lambda}-\lambda}{O(\hat{\lambda})^{-1/2}} < z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ fractile of the standard normal distribution. This is equivalent to

$$\hat{\lambda} - z_{\alpha/2} O(\hat{\lambda})^{-1/2} < \lambda < \hat{\lambda} + z_{\alpha/2} O(\hat{\lambda})^{-1/2}.$$

For the 6–MP treatment group remission times, an approximate two-sided 95% confidence interval for λ is

$$\frac{9}{359} - (1.96)(14,320)^{-1/2} < \lambda < \frac{9}{359} + (1.96)(14,320)^{-1/2}$$

or

$$0.0087 < \lambda < 0.0414,$$

which has smaller bounds than the previous two interval estimators.

To summarize the conclusions of this long example, the maximum likelihood estimate of the failure rate is

$$\hat{\lambda} = 0.0251$$

remission per week, which corresponds to an estimated mean remission time of 39.9 weeks. The three approximate two-sided 95% confidence intervals for λ are given in the second column of Table 5.1. Taking reciprocals, the third column contains the associated approximate two-sided 95% confidence intervals for the population mean remission time μ. The confidence intervals for λ associated with the first two techniques are not symmetric about the maximum likelihood estimator because they are based on the non-symmetric chi-square distribution. Since there are only $n = 21$ patients in the clinical trial and only $r = 9$ observed remission times, we have more faith in the actual coverage of the first two confidence interval techniques. This conclusion would need to be confirmed by a Monte Carlo simulation experiment.

Basis for confidence interval	Confidence interval for λ	Confidence interval for μ
Type II censoring approximate result	$0.0115 < \lambda < 0.0439$	$22.8 < \mu < 87.2$
Likelihood ratio statistic	$0.0120 < \lambda < 0.0452$	$22.1 < \mu < 83.0$
Asymptotic normality of the MLE	$0.0087 < \lambda < 0.0414$	$24.1 < \mu < 115.1$

Table 5.1: Approximate 95% confidence intervals for λ and μ for the 6–MP treatment group.

To summarize, the maximum likelihood estimator for the failure rate λ in the random censoring case is the same as in the complete, Type II and Type I censoring cases:

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} x_i}.$$

Three approximate confidence intervals for λ are based on (*a*) an exact result from Type II censoring, (*b*) the asymptotic distribution of the likelihood ratio statistic, and (*c*) the asymptotic normality of the maximum likelihood estimator. The confidence interval based on the asymptotic normality of the maximum likelihood estimator is symmetric and is therefore recommended only in the case of a large number of items on test.

5.5 Weibull Distribution

The Weibull distribution is typically more appropriate for modeling the lifetimes of items with a strictly increasing or decreasing hazard function, such as mechanical items. Rather than looking at each censoring mechanism (for example, no censoring, Type II censoring, Type I censoring) individually, we proceed directly to the general case of random censoring.

Maximum likelihood estimators. As before, let $t_1, t_2, \ldots, t_n$ be the failure times, $c_1, c_2, \ldots, c_n$ be the associated censoring times, and $x_i = \min\{t_i, c_i\}$ for $i = 1, 2, \ldots, n$. The Weibull distribution has hazard and cumulative hazard functions

$$h(t, \lambda, \kappa) = \kappa\lambda(\lambda t)^{\kappa-1} \qquad t \geq 0$$

and

$$H(t, \lambda, \kappa) = (\lambda t)^{\kappa} \qquad t \geq 0.$$

When there are r observed failures, the log likelihood function is

$$\begin{aligned}
\ln L(\lambda, \kappa) &= \sum_{i \in U} \ln h(x_i, \lambda, \kappa) - \sum_{i=1}^{n} H(x_i, \lambda, \kappa) \\
&= \sum_{i \in U} \left(\ln \kappa + \kappa \ln \lambda + (\kappa - 1) \ln x_i \right) - \sum_{i=1}^{n} (\lambda x_i)^{\kappa} \\
&= r \ln \kappa + \kappa r \ln \lambda + (\kappa - 1) \sum_{i \in U} \ln x_i - \lambda^{\kappa} \sum_{i=1}^{n} x_i^{\kappa},
\end{aligned}$$

and the 2×1 score vector has elements

$$U_1(\lambda, \kappa) = \frac{\partial \ln L(\lambda, \kappa)}{\partial \lambda} = \frac{\kappa r}{\lambda} - \kappa\lambda^{\kappa-1} \sum_{i=1}^{n} x_i^{\kappa}$$

and

$$U_2(\lambda, \kappa) = \frac{\partial \ln L(\lambda, \kappa)}{\partial \kappa} = \frac{r}{\kappa} + r \ln \lambda + \sum_{i \in U} \ln x_i - \sum_{i=1}^{n} (\lambda x_i)^{\kappa} \ln(\lambda x_i).$$

When these equations are set equal to zero, the simultaneous equations have no closed-form solution for $\hat{\lambda}$ and $\hat{\kappa}$:

$$\frac{\kappa r}{\lambda} - \kappa\lambda^{\kappa-1} \sum_{i=1}^{n} x_i^{\kappa} = 0,$$

$$\frac{r}{\kappa} + r\ln\lambda + \sum_{i\in U}\ln x_i - \sum_{i=1}^{n}(\lambda x_i)^{\kappa}\ln(\lambda x_i) = 0.$$

One piece of good fortune, however, to avoid solving a 2×2 set of nonlinear equations, is that the first equation can be solved for λ in terms of κ as

$$\lambda = \left(\frac{r}{\sum_{i=1}^{n} x_i^{\kappa}}\right)^{1/\kappa}.$$

Notice that λ reduces to the maximum likelihood estimator for the exponential distribution when $\kappa = 1$. Using this expression for λ in terms of κ in the second element of the score vector yields a single, albeit more complicated, expression with κ as the only unknown. After applying some algebra, this equation reduces to

$$g(\kappa) = \frac{r}{\kappa} + \sum_{i\in U}\ln x_i - \frac{r\sum_{i=1}^{n} x_i^{\kappa}\ln x_i}{\sum_{i=1}^{n} x_i^{\kappa}} = 0,$$

which must be solved iteratively. One technique that can be used to solve this equation is the Newton–Raphson procedure, which uses

$$\kappa_{j+1} = \kappa_j - \frac{g(\kappa_j)}{g'(\kappa_j)},$$

where κ_0 is an initial estimator. The iterative procedure can be repeated until the desired accuracy for κ is achieved; that is, $|\kappa_{j+1} - \kappa_j| < \varepsilon$, for some small positive real number ε. When the accuracy is achieved, the maximum likelihood estimator $\hat{\kappa}$ is used to calculate $\hat{\lambda} = \left(r/\sum_{i=1}^{n} x_i^{\hat{\kappa}}\right)^{1/\hat{\kappa}}$. The derivative of $g(\kappa)$ reduces to

$$g'(\kappa) = -\frac{r}{\kappa^2} - \frac{r}{\left(\sum_{i=1}^{n} x_i^{\kappa}\right)^2}\left[\left(\sum_{i=1}^{n} x_i^{\kappa}\right)\left(\sum_{i=1}^{n}(\ln x_i)^2 x_i^{\kappa}\right) - \left(\sum_{i=1}^{n} x_i^{\kappa}\ln x_i\right)^2\right].$$

Determining an initial estimator κ_0 is not trivial. When there are no censored observations, Menon's initial estimator for κ_0 is

$$\kappa_0 = \left\{\frac{6}{(n-1)\pi^2}\left[\sum_{i=1}^{n}(\ln t_i)^2 - \frac{\left(\sum_{i=1}^{n}\ln t_i\right)^2}{n}\right]\right\}^{-1/2}.$$

Least squares estimation can be used in the case of a right-censored data set. The Newton–Raphson procedure can fail to converge to the maximum likelihood estimators. A bisection algorithm or fixed point algorithm often provides more reliable convergence.

Fisher and observed information matrices. The 2×2 Fisher and observed information matrices are based on the following partial derivatives:

$$\frac{-\partial^2 \ln L(\lambda,\kappa)}{\partial\lambda^2} = \frac{\kappa r}{\lambda^2} + \kappa(\kappa-1)\lambda^{\kappa-2}\sum_{i=1}^{n} x_i^{\kappa},$$

$$\frac{-\partial^2 \ln L(\lambda,\kappa)}{\partial\lambda\partial\kappa} = -\frac{r}{\lambda} + \left[\left(\kappa\lambda^{\kappa-1}\right)\left(\sum_{i=1}^{n} x_i^{\kappa}\ln x_i\right) + \left(\sum_{i=1}^{n} x_i^{\kappa}\right)\left(\kappa\lambda^{\kappa-1}\ln\lambda + \lambda^{\kappa-1}\right)\right]$$

$$= -\frac{r}{\lambda} + \lambda^{\kappa-1}\left[\kappa\sum_{i=1}^{n} x_i^{\kappa}\ln x_i + (1+\kappa\ln\lambda)\sum_{i=1}^{n} x_i^{\kappa}\right],$$

$$\frac{-\partial^2 \ln L(\lambda,\kappa)}{\partial\kappa^2} = \frac{r}{\kappa^2} + \sum_{i=1}^{n}(\lambda x_i)^{\kappa}(\ln\lambda x_i)^2.$$

The expected values of these quantities are not tractable, so the Fisher information matrix does not have closed-form elements. The observed information matrix, however, can be determined by using $\hat{\lambda}$ and $\hat{\kappa}$ as arguments in these expressions.

Example 5.12 Example 5.5 showed that the exponential distribution was a poor approximation to the ball bearing data lifetimes. The histogram in Figure 5.8 indicated that a probability distribution with a nonzero mode and an increasing hazard function might provide a better fit. Fit the Weibull distribution to the ball bearing lifetimes and assess the fit.

The maximum likelihood estimates, using the Newton–Raphson technique described previously, are $\hat{\lambda} = 0.0122$ and $\hat{\kappa} = 2.10$. Figure 5.16 shows the empirical survivor function, along with the fitted exponential and Weibull survivor functions. It is clear that the Weibull distribution is superior to the exponential distribution in fitting the ball bearing failure times because it is capable of modeling wear out. The log likelihood function evaluated at the maximum likelihood estimators is $\ln L(\hat{\lambda}, \hat{\kappa}) = -113.691$. The log likelihood function is shown in Figure 5.17.

The observed information matrix is

$$O(\hat{\lambda}, \hat{\kappa}) = \begin{bmatrix} 681{,}000 & 875 \\ 875 & 10.4 \end{bmatrix},$$

revealing a positive correlation between the elements of the score vector. Using the fact that the likelihood ratio statistic, $-2\left[\ln L(\lambda,\kappa) - \ln L(\hat{\lambda}, \hat{\kappa})\right]$, is asymptotically $\chi^2(2)$,

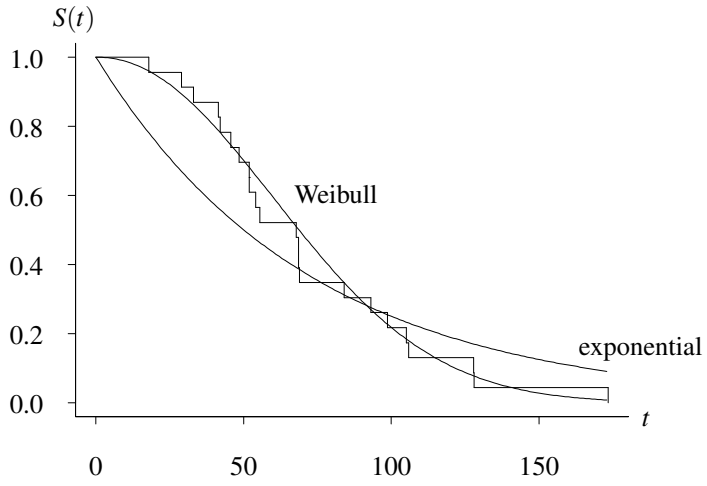

Figure 5.16: Exponential and Weibull fits to the ball bearing data.

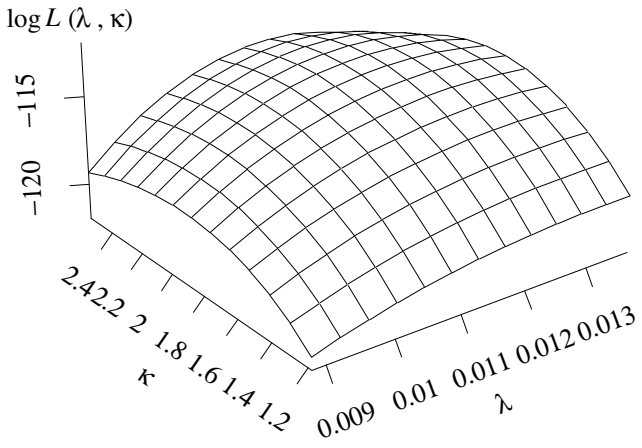

Figure 5.17: Log likelihood function for the ball bearing data.

an approximate 95% confidence region for the parameters is all λ and κ satisfying

$$-2[\ln L(\lambda, \kappa) + 113.691] < 5.99,$$

since $\chi^2_{2,0.05} = 5.99$. The 95% confidence region is shown in Figure 5.18, and, not surprisingly, the line $\kappa = 1$ is not interior to the region. This indicates that the exponential distribution is not an appropriate model for this particular data set. This is yet more statistical evidence that the ball bearings are wearing out. Note that the boundary of this region is a level surface of the log likelihood function shown in Figure 5.17 that is cut $5.99/2$ units below the maximum of the log likelihood function.

The R code to generate the confidence region is given below. The `crplot` function contained in the `conf` package calculates the maximum likelihood estimates $\hat{\lambda}$ and $\hat{\kappa}$ and plots the 95% confidence region. The first argument to `crplot` contains the data values, the second argument contains α, and the third argument contains the name of the population distribution. Setting the `pts` argument to `FALSE` means the points along the boundary are connected by lines; setting the `origin` argument to `TRUE` means the origin is included in the plot; setting the `info` argument to `TRUE` means the maximum likelihood estimates and boundary points in the confidence region can easily be retrieved.

```
library(conf)
bb = c(17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48, 51.84,
       51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12,
       93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40)
crplot(bb, 0.05, "weibull", pts = FALSE, origin = TRUE, info = TRUE)
```

As further evidence that the Weibull distribution is a significantly better model than the exponential, the likelihood ratio statistic can be used to determine whether κ is significant. Evaluating the log likelihood values at the maximum likelihood estimators

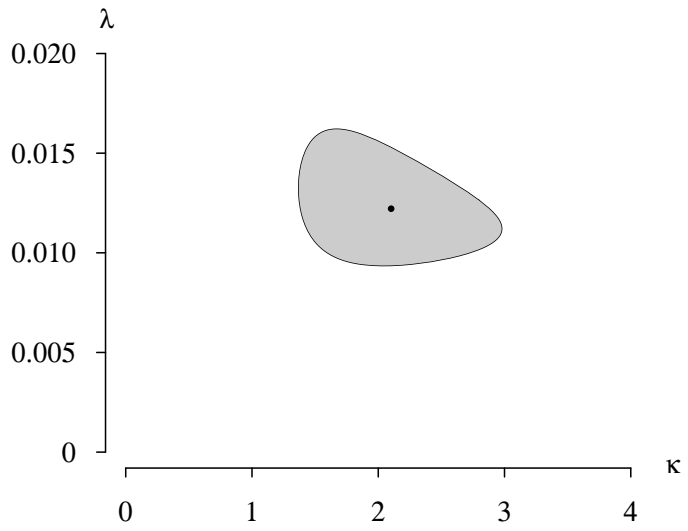

Figure 5.18: Confidence region ($\alpha = 0.05$) for λ and κ for the ball bearing data.

in the Weibull and exponential fits, the likelihood ratio statistic is

$$-2\left[\ln L\left(\hat{\lambda}\right) - \ln L\left(\hat{\lambda}, \hat{\kappa}\right)\right] = -2[-121.435 + 113.691] = 15.488.$$

This value shows that there is a statistically significant difference between κ and 1 when it is compared with the critical value $\chi^2_{1,0.05} = 3.84$.

If we are still uncertain as to whether κ is significantly different from 1, the standard errors of the distribution of the parameter estimators can be computed by determining the inverse of the observed information matrix

$$O^{-1}(\hat{\lambda}, \hat{\kappa}) = \begin{bmatrix} 0.00000165 & -0.000139 \\ -0.000139 & 0.108 \end{bmatrix}.$$

This matrix is an estimate of the variance–covariance matrix for the parameter estimates $\hat{\lambda}$ and $\hat{\kappa}$. The standard errors of the parameter estimates are the square roots of the diagonal elements

$$\hat{\sigma}_{\hat{\lambda}} = 0.00128 \qquad \hat{\sigma}_{\hat{\kappa}} = 0.329.$$

Thus, an asymptotic 95% confidence interval for κ is

$$2.10 - (1.96)(0.329) < \kappa < 2.10 + (1.96)(0.329)$$

or

$$1.46 < \kappa < 2.74,$$

since $z_{0.025} = 1.96$. Since this confidence interval does not contain 1, the parameter κ is statistically significant. Three different techniques have all drawn the same conclusion: the ball bearings are wearing out because there is a statistically significant difference between $\hat{\kappa} = 2.10$ and $\kappa = 1$.

5.6 Proportional Hazards Model

Parameter estimation for the proportional hazards model, which was introduced in Section 4.6, is considered in this section. Since there is now a vector of covariates in addition to a failure or censoring time for each item on test, special notation must be established to accommodate the covariates. The proportional hazards model has the unique feature that the baseline distribution need not be defined in order to estimate the regression coefficients associated with the covariates.

A lifetime model that incorporates a vector of covariates $\boldsymbol{z} = (z_1, z_2, \ldots, z_q)'$ models the impact of the covariates on survival. The reason for including this vector may be to determine which covariates significantly affect survival, to determine the distribution of the lifetime for a particular setting of the covariates, or to fit a more complicated distribution from a small data set, as opposed to fitting separate distributions for each level of the covariates.

The proportional hazards model was defined in Section 4.6 by

$$h(t, \boldsymbol{z}) = \psi(\boldsymbol{z}) h_0(t),$$

for $t \geq 0$, where $h_0(t)$ is a baseline hazard function. The covariates increase the hazard function when $\psi(\boldsymbol{z}) > 1$ or decrease the hazard function when $\psi(\boldsymbol{z}) < 1$. The goal of this section is to develop techniques for estimating the $q \times 1$ vector of regression coefficients $\boldsymbol{\beta}$ from a data set consisting of n items on test and r observed failure times.

The notation used to describe a data set in a lifetime model involving covariates will borrow some notation established earlier in this chapter, but also establish some new notation. As before, n is the number of items on test and r is the number of observed failures. The failure time of the ith item on test, t_i, is either observed or right censored at time c_i, for $i = 1, 2, \ldots, n$. As before, let $x_i = \min\{t_i, c_i\}$ and δ_i be a censoring indicator variable (1 for an observed failure and 0 for a right-censored value), for $i = 1, 2, \ldots, n$. In addition, a $q \times 1$ vector of covariates $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iq})'$ is collected for each item on test, for $i = 1, 2, \ldots, n$. Thus, z_{ij} is the value of covariate j for item i, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, q$. This formulation of the problem can be stated in matrix form as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1q} \\ z_{21} & z_{22} & \cdots & z_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nq} \end{bmatrix}.$$

Each row in the $\boldsymbol{Z}$ matrix consists of the values of the q covariates collected on a particular item. The matrix approach is useful because complicated systems of equations can be expressed compactly and operations on data sets can be performed efficiently by a computer. For parameter estimation, the survivor, density, hazard, and cumulative hazard functions now have the extra arguments $\boldsymbol{z}$ and $\boldsymbol{\beta}$ associated with them:

$$S(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) \qquad f(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) \qquad h(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) \qquad H(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}),$$

for $t \geq 0$, where the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ consists of the p unknown parameters associated with the baseline distribution, which must be estimated along with the regression coefficients $\boldsymbol{\beta}$.

Parameter estimation for the proportional hazards model can be divided into two cases. The first case is when the baseline distribution is known. This case applies when previous test results have indicated that a particular functional form of the baseline distribution is appropriate. The second case is when the baseline distribution is unknown. This is almost certainly the case when looking at a data set of lifetimes and covariates for the first time without any guidance with respect to an appropriate baseline distribution.

5.6.1 Known Baseline Distribution

When the baseline distribution is known, the parameter estimation procedure follows along the same lines as in the previous sections. The hazard function and cumulative hazard function in the proportional hazards model are

$$h(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \psi(\boldsymbol{z}) h_0(t)$$

and

$$H(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \psi(\boldsymbol{z}) H_0(t)$$

for $t \geq 0$, where $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters associated with the baseline distribution. For simplicity and mathematical tractability, only the log linear form of the link function, which is $\psi(\boldsymbol{z}) = e^{\boldsymbol{\beta}'\boldsymbol{z}}$, is considered here. This assumption is not necessary for some of the derivations, so many of the results apply to a wider range of link functions. When the log linear form of the link function is assumed, the hazard function and cumulative hazard function become

$$h(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) = e^{\boldsymbol{\beta}'\boldsymbol{z}} h_0(t)$$

and

$$H(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) = e^{\boldsymbol{\beta}'\boldsymbol{z}} H_0(t)$$

for $t \geq 0$, where $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters associated with the baseline distribution. The log likelihood function is

$$\begin{aligned} \ln L(\boldsymbol{\theta}, \boldsymbol{\beta}) &= \sum_{i \in U} \ln h(x_i, \boldsymbol{z}_i, \boldsymbol{\theta}, \boldsymbol{\beta}) - \sum_{i=1}^{n} H(x_i, \boldsymbol{z}_i, \boldsymbol{\theta}, \boldsymbol{\beta}) \\ &= \sum_{i \in U} \left[\boldsymbol{\beta}' \boldsymbol{z}_i + \ln h_0(x_i) \right] - \sum_{i=1}^{n} e^{\boldsymbol{\beta}'\boldsymbol{z}_i} H_0(x_i). \end{aligned}$$

This expression can be differentiated with respect to all the unknown parameters to arrive at the score vector, which is then equated to zero and solved numerically to arrive at the maximum likelihood estimates.

Two observations with respect to this model formulation are important. First, the maximum likelihood estimates for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ for most of the models in this section cannot be expressed in closed form (as was the case for the exponential distribution in Section 5.4), so numerical methods typically need to be used to find the values of the estimates. Second, the choice of whether to use a model of dependence or to examine each population separately is dependent on the number of unique covariate vectors $\boldsymbol{z}$ and the number of items on test, n. If, for example, n is large and there is only a single binary covariate (that is, only two unique covariate vectors, $z_1 = 0$ and $z_1 = 1$), it is probably wiser to analyze each of the two populations separately by the techniques described earlier.

Although numerical methods are required to find $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ in general, there are closed-form expressions in a very narrow case that satisfies the following conditions.

- The log linear link function $\psi(\boldsymbol{z}) = e^{\boldsymbol{\beta}'\boldsymbol{z}}$ is used to incorporate the vector of covariates $\boldsymbol{z}$ into the lifetime model.
- The baseline distribution is exponential(λ), which means that the baseline hazard function is $h_0(t) = \lambda$ and the baseline cumulative hazard function is $H_0(t) = \lambda t$ for $t \geq 0$.

Under these assumptions, the general form for the hazard function in the proportional hazards model

$$h(t, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \psi(\boldsymbol{z}) h_0(t)$$

reduces to the special case

$$h(t, \boldsymbol{z}, \lambda, \boldsymbol{\beta}) = \lambda e^{\boldsymbol{\beta}'\boldsymbol{z}}.$$

for $t \geq 0$. It is often more convenient notationally to define an additional covariate, $z_0 = 1$, for all n items on test. This allows the baseline parameter $\lambda = e^{\beta_0 z_0}$ to be included in the vector of regression coefficients, rather than being considered separately. The baseline hazard function is effectively absorbed into the link function. In this case, the hazard function can be expressed as

$$h(t, \boldsymbol{z}, \boldsymbol{\beta}) = e^{\boldsymbol{\beta}'\boldsymbol{z}}$$

for $t \geq 0$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)'$ and $\boldsymbol{z} = (z_0, z_1, \ldots, z_q)'$. The corresponding cumulative hazard function is

$$H(t, \boldsymbol{z}, \boldsymbol{\beta}) = te^{\boldsymbol{\beta}'\boldsymbol{z}}$$

for $t \geq 0$. Using this parameterization, the log likelihood function is

$$\begin{aligned}\ln L(\boldsymbol{\beta}) &= \sum_{i \in U} \ln h(x_i, \boldsymbol{z}_i, \boldsymbol{\beta}) - \sum_{i=1}^{n} H(x_i, \boldsymbol{z}_i, \boldsymbol{\beta}) \\ &= \sum_{i \in U} \boldsymbol{\beta}'\boldsymbol{z}_i - \sum_{i=1}^{n} x_i e^{\boldsymbol{\beta}'\boldsymbol{z}_i}.\end{aligned}$$

Differentiating this expression with respect to β_j yields the elements of the score vector

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i \in U} z_{ij} - \sum_{i=1}^{n} x_i z_{ij} e^{\boldsymbol{\beta}'\boldsymbol{z}_i}$$

for $j = 0, 1, \ldots, q$. When the elements of the score vector are equated to zero, the resulting set of $q+1$ nonlinear equations in $\boldsymbol{\beta}$ must be solved numerically in the general case. There is a closed-form solution for this set of simultaneous equations when there is a single binary covariate, often referred to as the two-sample case.

To find the observed information matrix and the Fisher information matrix, a second partial derivative of the log likelihood function is required:

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k} = -\sum_{i=1}^{n} x_i z_{ij} z_{ik} e^{\boldsymbol{\beta}'\boldsymbol{z}_i}$$

for $j = 0, 1, \ldots, q$ and $k = 0, 1, \ldots, q$. The observed information matrix can be determined by using the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ as an argument in this second partial derivative. Thus, the (j, k) element of the observed information matrix is

$$\left[-\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{n} x_i z_{ij} z_{ik} e^{\hat{\boldsymbol{\beta}}'\boldsymbol{z}_i}$$

for $j = 0, 1, \ldots, q$ and $k = 0, 1, \ldots, q$. For computational purposes, this can be expressed in matrix form as

$$O(\hat{\boldsymbol{\beta}}) = \boldsymbol{Z}'\hat{\boldsymbol{B}}\boldsymbol{Z},$$

where $\hat{\boldsymbol{B}}$ is an $n \times n$ diagonal matrix whose elements are $x_1 e^{\hat{\boldsymbol{\beta}}'\boldsymbol{z}_1}, x_2 e^{\hat{\boldsymbol{\beta}}'\boldsymbol{z}_2}, \ldots, x_n e^{\hat{\boldsymbol{\beta}}'\boldsymbol{z}_n}$. The Fisher information matrix is more difficult to calculate because it involves the expected value of the second partial derivative:

$$E\left[-\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}\right] = \sum_{i=1}^{n} z_{ij} z_{ik} e^{\boldsymbol{\beta}'\boldsymbol{z}_i} E[x_i]$$

for $j = 0, 1, \ldots, q$ and $k = 0, 1, \ldots, q$. Determining the value of $E[x_i]$ will be considered separately in the paragraphs that follow for uncensored ($r = n$) and censored ($r < n$) data sets.

For a complete data set, $E[x_i] = E[t_i]$, for $i = 1, 2, \ldots, n$, because there is no censoring. Since the population mean of the exponential distribution is the reciprocal of the failure rate and the ith item on test has failure rate $e^{\boldsymbol{\beta}'\boldsymbol{z}_i}$, $E[x_i] = e^{-\boldsymbol{\beta}'\boldsymbol{z}_i}$. Returning to the Fisher information matrix, the (j, k) element is

$$E\left[-\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}\right] = \sum_{i=1}^{n} z_{ij} z_{ik} e^{\boldsymbol{\beta}'\boldsymbol{z}_i} e^{-\boldsymbol{\beta}'\boldsymbol{z}_i} = \sum_{i=1}^{n} z_{ij} z_{ik}$$

for $j = 0, 1, \ldots, q$ and $k = 0, 1, \ldots, q$. This result for the Fisher information matrix has a particularly tractable matrix representation

$$I(\boldsymbol{\beta}) = \boldsymbol{Z}'\boldsymbol{Z},$$

which is a function of the matrix of covariates only.

For a censored data set, the expression for $E[x_i]$ is a bit more complicated. Since the failure rate for the ith item on test is $e^{\boldsymbol{\beta}'\boldsymbol{z}_i}$,

$$\begin{aligned} E[x_i] &= E\left[\min\{t_i, c_i\}\right] \\ &= \int_0^{c_i} t_i f_{T_i}(t_i) dt_i + c_i P[t_i \geq c_i] \\ &= \int_0^{c_i} t_i e^{\boldsymbol{\beta}'\boldsymbol{z}_i} e^{-e^{\boldsymbol{\beta}'\boldsymbol{z}_i} t_i} dt_i + c_i e^{-e^{\boldsymbol{\beta}'\boldsymbol{z}_i} c_i} \\ &= e^{-\boldsymbol{\beta}'\boldsymbol{z}_i}\left(1 - e^{-e^{\boldsymbol{\beta}'\boldsymbol{z}_i} c_i}\right) \end{aligned}$$

for $i = 1, 2, \ldots, n$, by using integration by parts. This means that the (j, k) element of the Fisher information matrix is

$$E\left[-\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k}\right] = \sum_{i=1}^{n} z_{ij} z_{ik} e^{\boldsymbol{\beta}'\boldsymbol{z}_i} e^{-\boldsymbol{\beta}'\boldsymbol{z}_i}\left[1 - e^{-e^{\boldsymbol{\beta}'\boldsymbol{z}_i} c_i}\right] = \sum_{i=1}^{n} z_{ij} z_{ik}(1 - \gamma_i),$$

where $\gamma_i = e^{-e^{\boldsymbol{\beta}'\boldsymbol{z}_i} c_i}$ is the probability that the ith item on test is censored, for $i = 1, 2, \ldots, n$. The potential censoring time for the ith item on test, c_i, must be known for each item in order to compute the Fisher information matrix, which is not always the case in practice. Letting $\boldsymbol{\Gamma}$ be a diagonal matrix with elements $\gamma_1, \gamma_2, \ldots, \gamma_n$, the Fisher information matrix can be written in matrix form as

$$I(\boldsymbol{\beta}) = \boldsymbol{Z}'(\mathrm{I} - \boldsymbol{\Gamma})\boldsymbol{Z},$$

which is independent of the failure times.

Before ending the discussion on the exponential baseline distribution, the two-sample case, where a binary covariate z_1 is used to differentiate between the control ($z_1 = 0$) and treatment ($z_1 = 1$) cases, is considered. This case is of interest because the maximum likelihood estimates can be expressed in closed form. The notation for the two-sample case is summarized in Table 5.2. As before, $z_0 = 1$ is included in the vector of covariates to account for the baseline distribution. The set of two nonlinear equations for finding the estimates of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ obtained by setting the score vector equal to $\mathbf{0}$ is

$$\begin{aligned} \sum_{i \in U} z_{i0} - \sum_{i=1}^{n} x_i z_{i0} e^{\beta_0 z_{i0} + \beta_1 z_{i1}} &= 0, \\ \sum_{i \in U} z_{i1} - \sum_{i=1}^{n} x_i z_{i1} e^{\beta_0 z_{i0} + \beta_1 z_{i1}} &= 0. \end{aligned}$$

Let $r_0 > 0$ be the number of observed failures in the control group ($z_1 = 0$), and let $r_1 > 0$ be the number of observed failures in the treatment group ($z_1 = 1$). Since $z_0 = 1$ for all items on test, the equations reduce to

$$r_0 + r_1 - \sum_{i=1}^{n} x_i e^{\beta_0 + \beta_1 z_{i1}} = 0,$$
$$r_1 - \sum_{i=1}^{n} x_i z_{i1} e^{\beta_0 + \beta_1 z_{i1}} = 0.$$

These equations can be further simplified by partitioning the summations based on the value of z_1:

$$r_0 + r_1 - \sum_{i|z_{i1}=0} x_i e^{\beta_0} - \sum_{i|z_{i1}=1} x_i e^{\beta_0 + \beta_1} = 0,$$
$$r_1 - \sum_{i|z_{i1}=1} x_i e^{\beta_0 + \beta_1} = 0.$$

Letting $\lambda_0 = e^{\beta_0}$ be the failure rate in the control group ($z_1 = 0$) and letting $\lambda_1 = e^{\beta_0 + \beta_1}$ be the failure rate in the treatment group ($z_1 = 1$), the equations become

$$r_0 + r_1 - \lambda_0 \sum_{i|z_{i1}=0} x_i - \lambda_1 \sum_{i|z_{i1}=1} x_i = 0,$$
$$r_1 - \lambda_1 \sum_{i|z_{i1}=1} x_i = 0.$$

When these equations are solved simultaneously, the maximum likelihood estimates for λ_0 and λ_1 are the same as those for the exponential distribution with two separate populations:

$$\hat{\lambda}_0 = \frac{r_0}{\sum_{i|z_{i1}=0} x_i} \qquad \text{and} \qquad \hat{\lambda}_1 = \frac{r_1}{\sum_{i|z_{i1}=1} x_i}.$$

These estimators are the ratio of the number of observed failures to the total time on test within the two groups.

Example 5.13 The patients in the 6–MP drug experiment described in Example 5.6 are broken down into a control group that did not receive the drug ($z_1 = 0$) and a treatment group that did receive the drug ($z_1 = 1$). The remission times, in weeks, for the 21 patients in the control group are

1 1 2 2 3 4 4 5 5 8 8
8 8 11 11 12 12 15 17 22 23.

	Control Group	Treatment Group
Number of failures	r_0	r_1
Baseline covariate z_0	1	1
Binary covariate z_1	0	1

Table 5.2: Single binary covariate proportional hazards model notation.

The remission times for the 21 patients that received the drug are

$$\begin{array}{cccccccccccc} 6 & 6 & 6 & 6^* & 7 & 9^* & 10 & 10^* & 11^* & 13 & 16 \\ 17^* & 19^* & 20^* & 22 & 23 & 25^* & 32^* & 32^* & 34^* & 35^*. \end{array}$$

There are a total of $n = 42$ patients in the clinical trial, and there are a total of $r = 30$ observed cancer recurrences, $r_0 = 21$ of which are in the control group and $r_1 = 9$ of which are in the treatment group. The values of $\boldsymbol{x}$, $\boldsymbol{\delta}$, and $\boldsymbol{Z}$ are given in Figure 5.19; the control group values have been arbitrarily placed first in the $\boldsymbol{x}$ vector. Note that for this analysis the *order* of the observations in the $\boldsymbol{x}$ vector is irrelevant. For tied values, the censored values have been placed last.

The maximum likelihood estimates for the failure rates for the two populations are

$$\hat{\lambda}_0 = \frac{r_0}{\sum_{i|z_{i1}=0} x_i} = \frac{21}{182} = 0.115 \qquad \text{and} \qquad \hat{\lambda}_1 = \frac{r_1}{\sum_{i|z_{i1}=1} x_i} = \frac{9}{359} = 0.0251$$

or, equivalently, in terms of the estimated mean remission times, the expected remission times of the control and treatment groups are estimated to be $\frac{182}{21} = 8.67$ weeks and $\frac{359}{9} = 39.9$ weeks, respectively. These estimates can be easily converted to the coefficients in the proportional hazards model:

$$\hat{\beta}_0 = \ln\left[\frac{21}{182}\right] = -2.16 \qquad \text{and} \qquad \hat{\beta}_1 = \ln\left[\frac{(9)(182)}{(359)(21)}\right] = -1.53.$$

Confidence intervals can be determined separately for the two populations because the remission times in each are assumed to be exponentially distributed. Using the techniques from Section 5.4, an exact two-sided 95% confidence interval for λ_0 is

$$\frac{(0.115)(26.00)}{42} < \lambda_0 < \frac{(0.115)(61.78)}{42}$$

$$0.0714 < \lambda_0 < 0.170$$

based on the chi-square distribution with 42 degrees of freedom. An approximate two-sided 95% confidence interval for λ_1 is

$$\frac{(0.0251)(8.23)}{18} < \lambda_1 < \frac{(0.0251)(31.53)}{18}$$

$$0.0115 < \lambda_1 < 0.0439$$

based on the chi-square distribution with 18 degrees of freedom. The first confidence interval is exact because the control group contains no censored observations, and the second confidence interval is approximate because the treatment group has randomly censored observations. Since these confidence intervals do not overlap, it can be concluded that 6–MP is effective in increasing remission times. If the side effects from 6–MP are minor, it should be prescribed to all leukemia patients.

Since exact confidence intervals apply only to the two-sample case with an exponential baseline distribution and Type II censoring, asymptotic intervals will also be calculated here to illustrate how they are developed in the general case. The Fisher information matrix cannot be calculated for this data set because the observed remission times do

$$
\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 3 \\ 4 \\ 4 \\ 5 \\ 5 \\ 8 \\ 8 \\ 8 \\ 8 \\ 11 \\ 11 \\ 12 \\ 12 \\ 15 \\ 17 \\ 22 \\ 23 \\ 6 \\ 6 \\ 6 \\ 6 \\ 7 \\ 9 \\ 10 \\ 10 \\ 11 \\ 13 \\ 16 \\ 17 \\ 19 \\ 20 \\ 22 \\ 23 \\ 25 \\ 32 \\ 32 \\ 34 \\ 35 \end{bmatrix}
\qquad
\delta = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\qquad
\boldsymbol{Z} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}
$$

Figure 5.19: Data values for the 6–MP experiment with a single binary covariate.

not have corresponding known censoring times. The observed information matrix, on the other hand, is easily calculated using the matrix formulation

$$O(\hat{\boldsymbol{\beta}}) = \boldsymbol{Z}'\hat{\boldsymbol{B}}\boldsymbol{Z} = \begin{bmatrix} 30 & 9 \\ 9 & 9 \end{bmatrix},$$

where $\hat{\boldsymbol{B}}$ is a 42×42 diagonal matrix with elements $x_1 e^{\hat{\boldsymbol{\beta}}' \boldsymbol{z}_1}$, $x_2 e^{\hat{\boldsymbol{\beta}}' \boldsymbol{z}_2}$, ..., $x_{42} e^{\hat{\boldsymbol{\beta}}' \boldsymbol{z}_{42}}$. Since the determinant of this matrix is $(30)(9) - 9^2 = 189$, it has inverse

$$O^{-1}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 9/189 & -9/189 \\ -9/189 & 30/189 \end{bmatrix},$$

which estimates the variance–covariance matrix of the maximum likelihood estimates. The off-diagonal elements of $O^{-1}(\hat{\boldsymbol{\beta}})$ indicate a negative correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$. The square roots of the diagonal elements yield asymptotic estimates for the standard deviation of the regression parameter estimates. Thus, the asymptotic estimated standard deviation of the estimate for β_0 is

$$\sqrt{\hat{V}\left[\hat{\beta}_0\right]} = \sqrt{\frac{9}{189}} = 0.218,$$

and the asymptotic estimated standard deviation of the estimate for β_1 is

$$\sqrt{\hat{V}\left[\hat{\beta}_1\right]} = \sqrt{\frac{30}{189}} = 0.398.$$

These values can be used in the usual fashion to obtain asymptotically valid confidence intervals and perform hypothesis testing with respect to the regression parameter estimates. Note that $\hat{\beta}_1 = -1.53$ is more than three standard deviation units away from 0, supporting the conclusion that there is a statistically significant difference between the patients who take 6–MP versus those that do not with respect to their remission times. Since the sign of $\hat{\beta}_1$ is negative, the drug prolongs the remission times. More specifically, since the proportional hazards model is being used, a patient taking the 6–MP drug will have a hazard function that is estimated to be $e^{\hat{\beta}_1} = e^{-1.53} = 0.217$ times that of a patient who does not take the drug.

Parameter estimation for single binary covariate is ideal in the sense that the parameter estimates can be expressed in closed form. The next subsection considers the more common situation in which the baseline distribution is unknown.

5.6.2 Unknown Baseline Distribution

In many applications, the baseline distribution is not known. Furthermore, the modeler may not be interested in the baseline distribution, rather only in the influence of the covariates on survival. A technique has been developed for the proportional hazards model that allows the coefficient vector $\boldsymbol{\beta}$ to be estimated without knowledge of the parametric form of the baseline distribution. This type of analysis might be appropriate when the modeler wants to detect which covariates are significant, to determine which covariate is the most significant, or to analyze interactions among covariates. This technique is characteristic of nonparametric methods because it is impossible to misspecify the baseline distribution.

The focus of this estimation technique is on the *indexes* of the components on test, as will be seen in the derivation to follow. Since this procedure is very different from all previous point estimation derivations, an example will be carried through the derivation to illustrate the notation and the method. The purpose in this small example is to determine whether light bulb wattage influences light bulb survival. This introduction to parameter estimation will alternate between the small example and the general case. In this example and the derivation, it is initially assumed that there is no censoring and there are no tied observations.

Example 5.14 A set of $n = 3$ light bulbs are placed on test. The first and second bulbs are 100-watt bulbs and the third bulb is a 60-watt bulb. A single ($q = 1$) covariate z_1 assumes the value 0 for a 60-watt bulb and 1 for a 100-watt bulb. The purpose of the test is to determine if the wattage has any influence on the survival distribution of the bulbs. The baseline distribution is unknown and unspecified, so there is only one parameter in the proportional hazards model, the regression coefficient β_1, that needs to be estimated. This small data set is used for illustrative purposes only, and we would obviously need to collect more than three data points to detect any statistically significant difference between the two wattages. Let $t_1 = 80$, $t_2 = 20$, and $t_3 = 50$ denote the lifetimes of the three bulbs. From the notation developed earlier in this chapter,

$$\boldsymbol{x} = \begin{bmatrix} 80 \\ 20 \\ 50 \end{bmatrix} \qquad \boldsymbol{\delta} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{Z} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

The order statistics are $t_{(1)} = 20$, $t_{(2)} = 50$, and $t_{(3)} = 80$. Figure 5.20 illustrates the definitions made thus far. Recall that the first subscript on z_{ij} is the bulb number and the second subscript is the covariate number. The *risk set* $R(t)$, parameterized by the

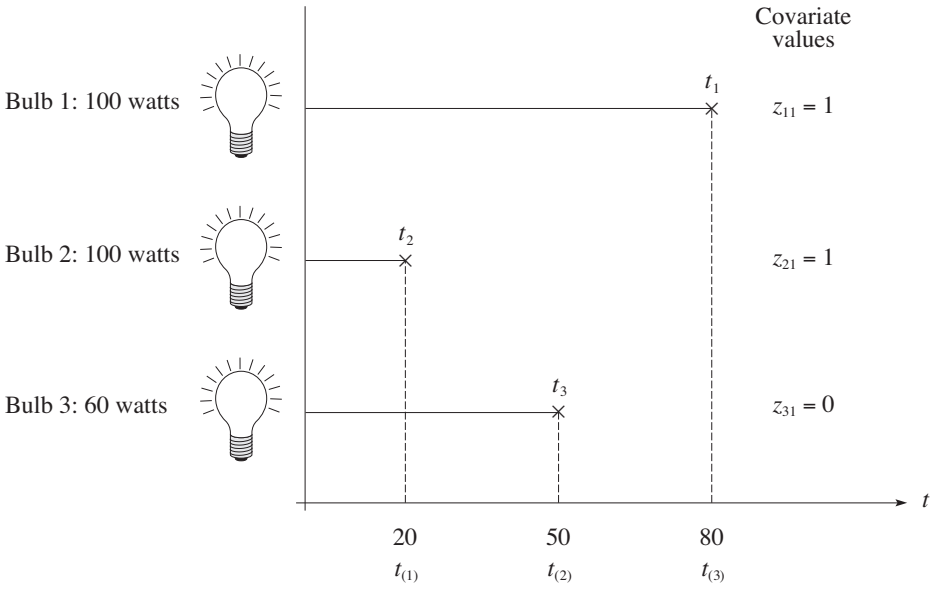

Figure 5.20: Proportional hazards parameter estimation notation.

failure times, is defined as the set of indexes of bulbs at risk just prior to time t. In this case

$$R(t_{(1)}) = R(t_2) = R(20) = \{1, 2, 3\}$$

since all bulbs are at risk just prior to $t_{(1)}$. At time $t_{(2)}$, the risk set is

$$R(t_{(2)}) = R(t_3) = R(50) = \{1, 3\}$$

since bulbs 1 and 3 are at risk just prior to $t_{(2)}$. Finally, at time $t_{(3)}$, the risk set is

$$R(t_{(3)}) = R(t_1) = R(80) = \{1\}$$

since only bulb 1 is still on test just prior to $t_{(3)}$. Similar to the concept of a pointer array from computer science, a *rank vector* $\boldsymbol{r}$ is used here to simplify the notation. The ith element of the rank vector is the index of the item that fails at $t_{(i)}$, for $i = 1, 2, 3$. For this particular data set,

$$\boldsymbol{r} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

because bulb 2 fails first, bulb 3 fails next, and bulb 1 fails last. The failure times for each bulb can therefore be determined from the order statistics and the rank vector.

The notation defined in the example is easily extended from three items on test with a single binary covariate to the general case. Let $t_1, t_2, \ldots, t_n$ be n distinct lifetimes. Each lifetime t_i has an associated $q \times 1$ vector of covariates $\boldsymbol{z}_i$, for $i = 1, 2, \ldots, n$. The ith order statistic is given by $t_{(i)}$, and the risk set $R(t_{(i)})$ is the set of indexes of all items that are at risk just prior to $t_{(i)}$, for $i = 1, 2, \ldots, n$. The ith element of the rank vector $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$ is the index of the item that fails at time $t_{(i)}$, for $i = 1, 2, \ldots, n$. The observed failure times and their associated indexes are equivalent to the observed order statistics and the associated rank vector. Now that the new notation has been defined, the emphasis transitions to determining the probability that a particular permutation of the indexes appears in the rank vector.

Example 5.15 We now return to the light bulb life test described in Example 5.14. The joint probability distribution of the elements of the rank vector, denoted by $f(r_1, r_2, r_3)$, is now considered for the data set containing $n = 3$ observations. In this case, there are $3! = 6$ possible permutations of the ranks of the observations:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \quad \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

If the wattage of the light bulb had no influence on the survival time, then clearly $f(r_1, r_2, r_3) = \frac{1}{6}$ for all six permutations because all three items are drawn from a homogeneous population with respect to survival. Switching to the non-equally-likely case, the probability mass function for the rank vector will be determined by finding the conditional probabilities associated with the ranks. For example, assume that a failure has just occurred at time $t_{(2)} = 50$, and the history up to time 50, which is bulb 2 failed at time $t_{(1)} = 20$, is known. The bulb that fails at time 50 is either bulb 1 or bulb 3. For small Δt, the conditional probability that the bulb failing at time 50 is bulb 1 is

$$P\big(r_2 = 1 \,\big|\, t_{(1)} = 20, t_{(2)} = 50, r_1 = 2\big) = \frac{P(\text{bulb 1 fails at time 50})}{P(\text{one item from } R(t_{(2)}) \text{ fails at time 50})}$$

$$
\begin{aligned}
&= \frac{h(50, z_{11})\Delta t}{h(50, z_{11})\Delta t + h(50, z_{31})\Delta t} \\
&= \frac{h(50, z_{11})}{h(50, z_{11}) + h(50, z_{31})} \\
&= \frac{\psi(z_{11})h_0(50)}{\psi(z_{11})h_0(50) + \psi(z_{31})h_0(50)} \\
&= \frac{\psi(z_{11})}{\psi(z_{11}) + \psi(z_{31})} \\
&= \frac{e^{\beta_1 z_{11}}}{e^{\beta_1 z_{11}} + e^{\beta_1 z_{31}}} \\
&= \frac{e^{\beta_1}}{e^{\beta_1} + 1}
\end{aligned}
$$

because the first bulb is 100 watts ($z_{11} = 1$) and the third bulb is 60 watts ($z_{31} = 0$). Note that the baseline hazard function has dropped out of this expression, so this probability will be the same regardless of the choice of $h_0(t)$. Also, the first two order statistics, $t_{(1)}$ and $t_{(2)}$, were not used in the calculation of this conditional probability. By similar reasoning, the conditional probability that the 60-watt bulb is the second to fail is

$$P\big(r_2 = 3 \,\big|\, t_{(1)} = 20, t_{(2)} = 50, r_1 = 2\big) = \frac{1}{e^{\beta_1} + 1}.$$

In the example, as well as in the general case, the conditional probability expression does not involve the failure times, making it possible to shorten $P(r_j = i \,|\, t_{(1)}, t_{(2)}, \ldots, t_{(j)}, r_1, r_2, \ldots, r_{j-1})$ to just $P(r_j = i \,|\, r_1, r_2, \ldots, r_{j-1})$. The probability that the jth element of the rank vector will be equal to i, given $t_{(j)}$ and the failure history up to $t_{(j)}$, is

$$
\begin{aligned}
P\big(r_j = i \,\big|\, r_1, r_2, \ldots, r_{j-1}\big) &= \frac{h(t_{(j)}, \boldsymbol{z}_i)\Delta t}{\sum_{k \in R(t_{(j)})} h(t_{(j)}, \boldsymbol{z}_k)\Delta t} \\
&= \frac{h(t_{(j)}, \boldsymbol{z}_i)}{\sum_{k \in R(t_{(j)})} h(t_{(j)}, \boldsymbol{z}_k)} \\
&= \frac{\psi(\boldsymbol{z}_i)h_0(t_{(j)})}{\sum_{k \in R(t_{(j)})} \psi(\boldsymbol{z}_k)h_0(t_{(j)})} \\
&= \frac{\psi(\boldsymbol{z}_i)}{\sum_{k \in R(t_{(j)})} \psi(\boldsymbol{z}_k)} \\
&= \frac{e^{\boldsymbol{\beta}' \boldsymbol{z}_i}}{\sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k}}.
\end{aligned}
$$

Example 5.16 We continue with the light bulb life test with $n = 3$ bulbs on test from Examples 5.14 and 5.15. It is now a simple task to use this conditional probability to determine the probability mass function for the indexes. For the three light bulbs, this probability mass function is

$$
\begin{aligned}
f(r_1, r_2, r_3) &= f(r_3 \,|\, r_1, r_2) f(r_1, r_2) \\
&= f(r_3 \,|\, r_1, r_2) f(r_2 \,|\, r_1) f(r_1) \\
&= f(r_1) f(r_2 \,|\, r_1) f(r_3 \,|\, r_1, r_2)
\end{aligned}
$$

over all six permutations of the rank vector. Since the sequence that was observed for the rank vector was $\boldsymbol{r} = (2, 3, 1)'$, this becomes

$$\begin{aligned} f(2,3,1) &= f(2)f(3\,|\,2)f(1\,|\,2,3) \\ &= \frac{\psi(z_{21})}{\psi(z_{11})+\psi(z_{21})+\psi(z_{31})} \cdot \frac{\psi(z_{31})}{\psi(z_{11})+\psi(z_{31})} \cdot \frac{\psi(z_{11})}{\psi(z_{11})} \\ &= \frac{e^{\beta_1 z_{21}}}{e^{\beta_1 z_{11}}+e^{\beta_1 z_{21}}+e^{\beta_1 z_{31}}} \cdot \frac{e^{\beta_1 z_{31}}}{e^{\beta_1 z_{11}}+e^{\beta_1 z_{31}}} \\ &= \frac{e^{\beta_1}}{e^{\beta_1}+e^{\beta_1}+1} \cdot \frac{1}{e^{\beta_1}+1} \\ &= \frac{e^{\beta_1}}{\left(2e^{\beta_1}+1\right)\left(e^{\beta_1}+1\right)}. \end{aligned}$$

Treating this expression as a likelihood function $L(\beta_1)$, the problem reduces to determining the β_1 value that maximizes the log likelihood function

$$\ln L(\beta_1) = \beta_1 - \ln\left(2e^{\beta_1}+1\right) - \ln\left(e^{\beta_1}+1\right).$$

The score statistic is

$$\frac{\partial \ln L(\beta_1)}{\partial \beta_1} = 1 - \frac{2e^{\beta_1}}{2e^{\beta_1}+1} - \frac{e^{\beta_1}}{e^{\beta_1}+1}.$$

Setting the score statistic to zero and solving for the maximum likelihood estimate, $\hat{\beta}_1 = (-\ln 2)/2 = -0.347$. Since $\hat{\beta}_1 < 0$, there is lower risk for the 100-watt bulbs than for 60-watt bulbs. More specifically, the hazard function for 100-watt light bulbs is $e^{\hat{\beta}_1} = \sqrt{2}/2 = e^{-0.347} = 0.707$ times that of the baseline hazard function for 60-watt bulbs, regardless of what baseline distribution is considered. To see if this regression coefficient is statistically significant involves calculating the negative of the derivative of the score:

$$-\frac{\partial^2 \ln L(\beta_1)}{\partial \beta_1^2} = \frac{2e^{\beta_1}}{\left(2e^{\beta_1}+1\right)^2} + \frac{e^{\beta_1}}{\left(e^{\beta_1}+1\right)^2}.$$

When this expression is evaluated at $\beta_1 = \hat{\beta}_1 = -0.347$, the 1×1 observed information matrix is 0.485, so the asymptotic estimate of the variance of $\hat{\beta}_1$ is $1/0.485 = 2.06$, and the asymptotic estimate of the standard deviation of $\hat{\beta}_1$ is $\sqrt{2.06} = 1.44$. Since $\hat{\beta}_1$ is only a fraction of a standard deviation away from 0, z_1 is not statistically significant. This result is not surprising considering the small number of light bulbs placed on the life test. Note that these values are only asymptotically correct and are obviously poor approximations when $n = 3$. The p-value for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ is 0.809, indicating that there is no statistical evidence that wattage influences the longevity of a light bulb for this tiny data set. In addition, only the order of the failure times and not their numerical values were used to find $\hat{\beta}_1$. This means, for example, that the failure time of the third bulb, t_3, could have fallen anywhere on the interval $(20, 80)$, and the estimate would have been the same because the order of the observed failure times was not changed.

The R code below confirms the calculations given above. The `coxph` function, which is part of the `survival` package, is used to calculate the estimated regression coefficient

$\hat{\beta}_1 = -0.347$, which is stored in `b`, the 1×1 observed information matrix, which is stored in `v`, and the p-value for the hypothesis test, which is stored in `p`.

```
library(survival)
failtimes = c(80, 20, 50)
censor    = c(1, 1, 1)
z         = c(1, 1, 0)
bulbs.fit = coxph(Surv(failtimes, censor) ~ z)
b         = bulbs.fit$coef
v         = bulbs.fit$var
p         = 2 * pnorm(b / sqrt(v))
```

The procedure for estimating β_1 can be generalized from the example without any significant difficulties. The probability mass function for the indexes, or the likelihood function for $\boldsymbol{\beta}$, is now

$$\begin{aligned} L(\boldsymbol{\beta}) &= f(r_1) f(r_2 \,|\, r_1) \ldots f(r_n \,|\, r_1, r_2, \ldots, r_{n-1}) \\ &= \prod_{j=1}^{n} \frac{\psi(\boldsymbol{z}_{r_j})}{\sum_{k \in R(t_{(j)})} \psi(\boldsymbol{z}_k)} \\ &= \prod_{j=1}^{n} \frac{e^{\boldsymbol{\beta}' \boldsymbol{z}_{r_j}}}{\sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k}}. \end{aligned}$$

The log likelihood is

$$\ln L(\boldsymbol{\beta}) = \sum_{j=1}^{n} \left[\boldsymbol{\beta}' \boldsymbol{z}_{r_j} - \ln \sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k} \right].$$

The score vector has sth component

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_s} = \sum_{j=1}^{n} \left[z_{sr_j} - \frac{\sum_{k \in R(t_{(j)})} z_{sk} e^{\boldsymbol{\beta}' \boldsymbol{z}_k}}{\sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k}} \right]$$

for $s = 1, 2, \ldots, q$. The vector of maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ is obtained when the elements of the score vector are equated to zero and solved via numerical methods. To determine an estimate for the variance of $\hat{\boldsymbol{\beta}}$, the score vector must be differentiated to calculate the observed information matrix. The diagonal elements of the inverse of the observed information matrix are asymptotically valid estimates of the variance of $\hat{\boldsymbol{\beta}}$.

There are two approaches to handle right censoring that do not significantly complicate the derivation presented thus far. The first approach is to assume that right censoring occurs immediately after a failure occurs when a failure time and right-censoring time coincide. This assumption is valid for a Type II censored data set, but will involve an approximation for more general right-censoring schemes. In this case the rank vector is shortened to only r elements, corresponding to the indexes of the observed failure times $t_{(1)}, t_{(2)}, \ldots, t_{(r)}$. The likelihood function is

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\psi(\boldsymbol{z}_{r_j})}{\sum_{k \in R(t_{(j)})} \psi(\boldsymbol{z}_k)} = \prod_{j=1}^{r} \frac{e^{\boldsymbol{\beta}' \boldsymbol{z}_{r_j}}}{\sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k}}.$$

The log likelihood function is

$$\ln L(\boldsymbol{\beta}) = \sum_{j=1}^{r} \left[\boldsymbol{\beta}' \boldsymbol{z}_{r_j} - \ln \sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k} \right].$$

The score vector has sth component

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_s} = \sum_{j=1}^{r} \left[z_{sr_j} - \frac{\sum_{k \in R(t_{(j)})} z_{sk} e^{\boldsymbol{\beta}' \boldsymbol{z}_k}}{\sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k}} \right]$$

for $s = 1, 2, \ldots, q$. Using the quotient rule, the derivative of the score vector is

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_s \partial \beta_t} = -\sum_{j=1}^{r} \frac{\left(\sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k} \right) \left(\sum_{k \in R(t_{(j)})} z_{sk} z_{tk} e^{\boldsymbol{\beta}' \boldsymbol{z}_k} \right) - \left(\sum_{k \in R(t_{(j)})} z_{sk} e^{\boldsymbol{\beta}' \boldsymbol{z}_k} \right) \left(\sum_{k \in R(t_{(j)})} z_{tk} e^{\boldsymbol{\beta}' \boldsymbol{z}_k} \right)}{\left(\sum_{k \in R(t_{(j)})} e^{\boldsymbol{\beta}' \boldsymbol{z}_k} \right)^2}$$

for $s = 1, 2, \ldots, q$ and $t = 1, 2, \ldots, q$. The elements of the observed information matrix are obtained by using the maximum likelihood estimates as arguments in the negative of this expression.

The second approach to right censoring is to write the likelihood function as the sum of all likelihoods for complete data sets that are consistent with the censoring pattern. Fortunately, this second approach yields the same likelihood function as the first approach, as illustrated by the following example.

Example 5.17 In the previous example, the data set consisted of three observed failure times: 80, 20, 50. Now, if the situation changes so that the lifetime of the third light bulb is right censored at time 50, the data set is 80, 20, 50*, as is illustrated in Figure 5.21. Using the first approach to right censoring, the observed rank vector is now $\boldsymbol{r} = (2, 1)'$, and the likelihood function is

$$\begin{aligned} L(\beta_1) &= \prod_{j=1}^{2} \frac{\psi(z_{r_j})}{\sum_{k \in R(t_{(j)})} \psi(z_k)} \\ &= \frac{\psi(z_{21})}{\psi(z_{11}) + \psi(z_{21}) + \psi(z_{31})} \cdot \frac{\psi(z_{11})}{\psi(z_{11})} \\ &= \frac{\psi(z_{21})}{\psi(z_{11}) + \psi(z_{21}) + \psi(z_{31})}. \end{aligned}$$

For the second approach to right censoring, there are two possibilities for the rank vector if there was no censoring: if the third bulb failed before time 80, the observed rank vector would be $\boldsymbol{r} = (2, 3, 1)'$; if the third bulb failed after time 80, the observed rank vector would be $\boldsymbol{r} = (2, 1, 3)'$. In the first case, the likelihood function would be that from the previous example:

$$\frac{\psi(z_{21})}{\psi(z_{11}) + \psi(z_{21}) + \psi(z_{31})} \cdot \frac{\psi(z_{31})}{\psi(z_{11}) + \psi(z_{31})} \cdot \frac{\psi(z_{11})}{\psi(z_{11})}.$$

In the second case, the likelihood function would be

$$\frac{\psi(z_{21})}{\psi(z_{11}) + \psi(z_{21}) + \psi(z_{31})} \cdot \frac{\psi(z_{11})}{\psi(z_{11}) + \psi(z_{31})} \cdot \frac{\psi(z_{31})}{\psi(z_{31})}.$$

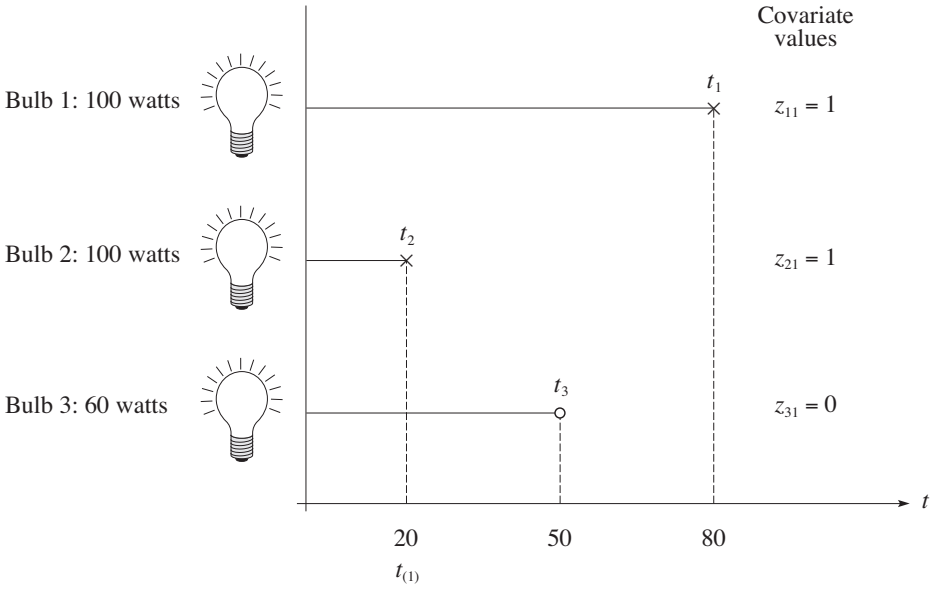

Figure 5.21: Proportional hazards model with censoring.

The sum of these two likelihood functions is

$$\frac{\psi(z_{21})}{\psi(z_{11}) + \psi(z_{21}) + \psi(z_{31})},$$

which is the same result as in the first approach to handling right censoring.

Tied lifetimes are typically handled by an approximation. When there are several failures at the same time value, each is assumed to contribute the same term to the likelihood function. Consequently, all the items with tied failure times are included in the risk set at the time of the tied observation. This approximation works well when there are not many tied observations in the data set and has been implemented in many software packages that estimate the vector of regression coefficients $\boldsymbol{\beta}$.

Example 5.18 Fit the Cox proportional hazards model via maximum likelihood to the remission times in the 6–MP clinical trial with a single binary covariate z_1 for the control ($z_1 = 0$) and treatment ($z_1 = 1$) groups. The data values are given in Example 5.6.

Using numerical methods, the maximum likelihood estimate for the single regression parameter is $\hat{\beta}_1 = -1.51$. The log likelihood function attains a value of -86.38 at the maximum likelihood value, the observed information matrix has a single value 5.962, and the inverse of the observed information matrix is $1/5.962 = 0.168$. This means that an asymptotic estimate of the standard deviation of the maximum likelihood estimate is

$$\sqrt{\hat{V}\left[\hat{\beta}_1\right]} = \sqrt{0.168} = 0.41,$$

which indicates that the maximum likelihood estimate is $1.51/0.41 = 3.7$ standard deviations units away from 0. It can be concluded, with a p-value less than 0.001, that

the 6–MP drug is effective in increasing the remission times for leukemia patients, assuming that the proportional hazards model is appropriate here. Regardless of the baseline hazard function $h_0(t)$ chosen, the hazard function in the treatment case is $e^{\hat{\beta}_1} = e^{-1.51} = 0.221$ times that of the baseline hazard function for all time values. Note that no work has been done here to assess model adequacy, and all these conclusions have been based on the fact that the proportional hazards model adequately describes the distribution of the remission time with the single binary covariate.

The R code below uses the `coxph` function in the `survival` package to compute $\hat{\beta}_1$ and a p-value for the appropriate hypothesis test.

```
library(survival)
x1 = c(1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12,
       12, 15, 17, 22, 23)
d1 = rep(1, length(x2))
z1 = rep(0, length(x2))
x2 = c(6, 6, 6, 6, 7, 9, 10, 10, 11, 13, 16, 17, 19, 20, 22, 23,
       25, 32, 32, 34, 35)
d2 = c(1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0)
z2 = rep(1, length(x1))
x  = c(x1, x2)
d  = c(d1, d2)
z  = c(z1, z2)
ph  = coxph(Surv(x, d) ~ z)
summary(ph)
```

The data set is contained in the `gehan` data frame in the `MASS` package, so the Cox proportional hazards model can also be fitted with the statements

```
library(survival)
library(MASS)
summary(coxph(Surv(time, cens) ~ treat, data = gehan))
```

which reverses the roles of the treatment and control groups, resulting in the reversal of the sign of $\hat{\beta}_1$.

The last example moves from the single binary covariate case to the case in which there are $q > 1$ covariates which can assume discrete and continuous values. The survival analysis application comes from sociology, and the analyst is attempting to determine which of the covariates significantly influences survival.

Example 5.19 The proportional hazards model has been used in diverse applications. Recidivism considers the probability that an inmate will return to prison in the future after release. Recidivism can be predicted using survival models. Several factors related to inmate background that could affect an inmate's adjustment to society are potential screening variables. North Carolina collected recidivism data on $n = 1540$ prisoners in 1978. The lifetime of interest here is the time of release until the time of return to prison. Obviously, not all inmates will return to prison, so a more complicated *split model*, for which some of the lifetimes are assumed to be infinite, may also be used. In addition,

there is significant right censoring in the data set. The purpose of the study is to assess the impact of the $q = 15$ covariates. The covariates $z_1, z_2, \ldots, z_{15}$ are time served, age, number of prior convictions, number of rule violations in prison, education, race, gender, alcohol problems, drug problems, marital status, probationary period, participation in a work release program, type of crime, crime against person, and crime against property. Many of these covariates are coded as indicator variables. Table 5.3 presents the estimates of the regression coefficients and their standard deviations in order of their significance. The column labeled *Covariate* gives a short description of the covariate considered. The next two columns give the regression coefficient estimator and an asymptotic estimate of its standard deviation. The column labeled $\hat{\beta}/\sqrt{\hat{V}\left[\hat{\beta}\right]}$ gives a test statistic for testing $H_0 : \beta_i = 0$ versus $H_1 : \beta_i \neq 0$, for $i = 1, 2, \ldots, 15$. The column labeled *p-value* indicates the attained significance of the covariates. A value less than $\alpha = 0.05$ indicates that a covariate is a statistically significant indicator of recidivism. Ten of the fifteen covariates are statistically significant. This example includes indicator variables (such as gender) and can easily be extended to include other regression modeling tools such as nonlinear and interaction terms in the regression model.

Name	Covariate	$\hat{\beta}$	$\sqrt{\hat{V}\left[\hat{\beta}\right]}$	$\frac{\hat{\beta}}{\sqrt{\hat{V}\left[\hat{\beta}\right]}}$	p-value	Significant
z_2	AGE	-3.3420	0.5195	-6.4328	0.0000	•
z_3	PRIORS	0.8355	0.1371	6.0957	0.0000	•
z_1	TSERVD	1.1666	0.1957	5.9616	0.0000	•
z_6	WHITE	-0.4444	0.0876	-5.0701	0.0000	•
z_8	ALCHY	0.4285	0.1043	4.1103	0.0000	•
z_{13}	FELON	-0.5782	0.1633	-3.5412	0.0002	•
z_9	JUNKY	0.2819	0.0970	2.9058	0.0018	•
z_7	MALE	0.6745	0.2423	2.7834	0.0027	•
z_{15}	PROPTY	0.3894	0.1578	2.4678	0.0068	•
z_4	RULE	3.0788	1.6890	1.8229	0.0342	•
z_{10}	MARRIED	-0.1532	0.1077	-1.4227	0.0774	
z_5	SCHOOL	-0.2507	0.1933	-1.2966	0.0974	
z_{12}	WORKREL	0.0865	0.0902	0.9587	0.1688	
z_{14}	PERSON	0.0737	0.2425	0.3039	0.3806	
z_{11}	SUPER	-0.0088	0.0966	-0.0914	0.4636	

Table 5.3: North Carolina recidivism model.

This chapter has contained a brief introduction to some of the statistical methods that are used in survival analysis. The key modeling features that indicate the use of survival analysis are (a) a population lifetime distribution with nonnegative support, (b) appreciable dispersion, (c) possibly right-censored data values, (d) possibly a vector of covariates which might influence the lifetime distribution. The exponential, Weibull, and Cox proportional hazards model were fitted to complete and right-censored data sets in this chapter.

5.7 Exercises

5.1 Consider a large batch of light bulbs whose lifetimes are known to have exponential(1) lifetimes. Gina knows that the population distribution is exponential, but she does not know the value of the population mean. She estimates the population mean lifetime of the light bulbs by averaging n observed lifetimes from bulbs chosen at random from the batch. Find the smallest value of n that assures, with probability of at least 0.95, that the sample mean is within 0.2 of the population mean

(a) exactly,

(b) approximately, using the central limit theorem.

5.2 Libby is a statistician for a light bulb company. She knows that the lifetimes of the 60-watt bulbs that her company manufactures are exponentially distributed with population mean 1500 hours. She conducts a life test in which 39 of their 60-watt bulbs are placed on life test until they fail and the average of the failure times is recorded. Find the probability that the sample mean exceeds 1600 hours using

(a) the central limit theorem,

(b) the exact distribution of the sample mean.

5.3 Let $t_1, t_2, \ldots, t_n$ be a random sample from an exponential(λ) population, where λ is a positive unknown failure rate parameter. Find an unbiasing constant c_n so that $c_n t_{(1)}$ is an unbiased estimator of $1/\lambda$, where $t_{(1)} = \min\{t_1, t_2, \ldots, t_n\}$ is the first order statistic. *Hint*: the unbiasing constant c_n is a function of the number of items on test n.

5.4 Debbie purchases a laptop computer with a random lifetime T whose probability distribution is a special case of the log logistic distribution with survivor function

$$S(t) = \frac{1}{1+\lambda t} \qquad t > 0,$$

where λ is a positive unknown scale parameter. From just a single observation of the lifetime of her laptop computer, find an exact two-sided 90% confidence interval for λ.

5.5 Let $t_1, t_2, \ldots, t_n$ be a random sample from an exponential population with mean θ, where θ is a positive unknown parameter. An exact two-sided 90% confidence interval for θ is

$$27 < \theta < 55.$$

Carol is not concerned about large values of θ. Only small values of θ are of concern. What is an exact one-sided 95% confidence interval of the form $\theta > k$, for some constant k?

5.6 If $t_1, t_2, \ldots, t_n$ are n mutually independent observations from a log normal distribution with probability density function

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2} \qquad t \geq 0$$

for $\sigma > 0$ and $-\infty < \mu < \infty$, find the maximum likelihood estimators of μ and σ and exact two-sided $100(1-\alpha)\%$ confidence intervals for μ and σ in terms of $t_1, t_2, \ldots, t_n$.

5.7 Let $t_1, t_2, \ldots, t_7$ be a random sample of the lifetimes of $n = 7$ items on test drawn from an exponential population with positive unknown mean θ.

(a) Find an exact two-sided 90% confidence interval for the median by finding a pivotal quantity based on the sample median $t_{(4)}$.

(b) Give an exact two-sided 90% confidence interval for the median for the $n = 7$ rat survival times in the treatment group from Efron and Tibshirani (1993, page 11):

$$16, 23, 38, 94, 99, 141, 197.$$

(c) Conduct a Monte Carlo simulation experiment to provide convincing numerical evidence that the exact two-sided 90% confidence interval for the median is indeed an exact two-sided 90% confidence interval for an exponential population when θ is arbitrarily set to 1.

5.8 This chapter has emphasized confidence intervals. Another type of statistical interval is known as a *prediction interval*, which contains a future value of an observation with a prescribed probability. Let $t_1, t_2, \ldots, t_n$ be a random sample from an exponential population with a positive unknown mean θ. Conduct a Monte Carlo simulation experiment that provides convincing numerical evidence that the $100(1-\alpha)\%$ prediction interval for t_{n+1}

$$\frac{\bar{t}}{F_{2n,2,\alpha/2}} < t_{n+1} < \frac{\bar{t}}{F_{2n,2,1-\alpha/2}}$$

is an *exact* prediction interval for the arbitrary parameter settings $n = 11$, $\alpha = 0.05$, and $\theta = 19$.

5.9 Let T_1, T_2, T_3 be mutually independent random variables such that T_i is exponentially distributed with mean $i\theta$, for $i = 1, 2, 3$, where θ is a positive unknown parameter.

(a) Find the maximum likelihood estimator $\hat{\theta}$.

(b) Find the probability density function of the maximum likelihood estimator $\hat{\theta}$.

(c) Is $\hat{\theta}$ an unbiased estimator of θ?

(d) Find an exact two-sided $100(1-\alpha)\%$ confidence interval for θ.

(e) Perform a Monte Carlo simulation experiment to evaluate the coverage of the confidence interval for $\theta = 10$ and $\alpha = 0.1$.

5.10 Let $t_1, t_2, \ldots, t_n$ be a random sample from a population with probability density function

$$f(t) = \frac{\theta}{t^{\theta+1}} \qquad t \geq 1,$$

where θ is a positive unknown parameter.

(a) Find the maximum likelihood estimator of θ.

(b) Use the invariance property to find the maximum likelihood estimator of the median of the distribution.

5.11 Let $t_1, t_2, \ldots, t_n$ be a random sample from a population with probability density function

$$f(t) = \sqrt{\frac{\lambda}{2\pi t^3}}\, e^{-\lambda(t-1)^2/(2t)} \qquad t > 0,$$

where λ is a positive unknown parameter. This distribution is known as the *standard Wald distribution* which is a special case of the inverse Gaussian distribution Find the maximum likelihood estimator of λ.

5.12 Let $T_1, T_2, \ldots, T_n$ be mutually independent and identically distributed random variables from a population having probability density function

$$f(t) = 7e^{-7(t-\theta)} \qquad t \geq \theta.$$

Find the limiting distribution of $n\left(T_{(1)} - \theta\right)$. Support this limiting distribution by conducting a Monte Carlo simulation experiment.

5.13 Let $t_1, t_2, \ldots, t_n$ be a random sample from a population with probability density function

$$f(t) = \frac{\theta}{(1+t)^{\theta+1}} \qquad t \geq 0,$$

where θ is a positive unknown parameter. Calculate an asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for θ based on the asymptotic normality of the maximum likelihood estimator.

5.14 Let $t_1, t_2, \ldots, t_n$ be a random sample from a population with probability density function

$$f(t) = \sqrt{\frac{1}{2\pi t^3}}\, e^{-(t-\theta)^2/(2t\theta^2)} \qquad t > 0,$$

where θ is a positive unknown parameter. This population distribution is a special case of the inverse Gaussian distribution. Calculate an asymptotically exact two-sided $100(1-\alpha)\%$ confidence interval for θ based on the asymptotic normality of the maximum likelihood estimator. *Hint*: the expected value of T is $E[T] = \theta$.

5.15 Let $t_1, t_2, \ldots, t_n$ be a random sample from a population with probability density function

$$f(t) = \frac{\theta}{(1+\theta t)^2} \qquad t \geq 0,$$

where θ is a positive unknown parameter. This is a special case of the log logistic distribution.

(a) Find the maximum likelihood estimator of θ. *Hint*: The maximum likelihood estimator cannot be expressed in closed form.

(b) Find the maximum likelihood estimate of θ for the $n = 7$ rat survival times (in days) of the treatment group from Efron and Tibshirani (1993, page 11):

$$16, 23, 38, 94, 99, 141, 197.$$

(c) Find an asymptotically exact two-sided 95% confidence interval for θ based on the likelihood ratio statistic for the rat survival times from part (b).

5.16 If n items from an exponential population with failure rate λ are placed on a life test that is terminated after r failures have occurred, show that

$$V\left[\sum_{i=1}^{n} x_i\right] = \frac{r}{\lambda^2},$$

where $x_i = \min\{t_i, c_i\}$, t_i is the time to failure of the ith item, and c_i is the right-censoring time for the ith item, $i = 1, 2, \ldots, n$.

5.17 If n items from an exponential population with failure rate λ are placed on a life test that is terminated after r failures have occurred, find the expected time to complete the test if

(a) failed items are not replaced,

(b) failed items are immediately replaced with new items.

5.18 Find the score, maximum likelihood estimator, and Fisher information matrix for a Type II censored random sample from a population with

$$f(t) = \frac{\theta}{t^{\theta+1}} \qquad t \geq 1,$$

where θ is a positive unknown parameter.

5.19 The lifetimes of studio light bulbs, measured in days, is exponentially distributed with an unknown failure rate λ. James places n studio light bulbs on test at noon on one day and subsequently checks for failed bulbs at noon on subsequent days until all bulbs have failed. Let $r_1, r_2, \ldots, r_k$ be the number of observed bulb failures, some of which may be zero, on the k days that the bulbs are inspected. Find the maximum likelihood estimator for λ. Also, give the maximum likelihood estimate for the data values $r_1 = 8$, $r_2 = 5$, $r_3 = 2$, $r_5 = 1$, and all other r_i values equal zero.

5.20 James's friend Alexandra decides to simplify matters from the previous question by assuming that all failures that occur during any interval occur at midnight. What is Alexandra's maximum likelihood estimator for λ as a function of n and $r_1, r_2, \ldots, r_k$?

5.21 Dre conducts a life test on n items from an exponential population with mean θ. He observes only the value of a single order statistic $t_{(k)}$, where k is known. So $k-1$ lifetimes are left censored at $t_{(k)}$, one lifetime is observed at $t_{(k)}$, and $n-k$ lifetimes are right censored at $t_{(k)}$.

(a) What is the score statistic for estimating θ?

(b) What is the maximum likelihood estimator for θ when $n = 30$, $k = 11$, and $t_{(11)} = 15.5$?

5.22 Consider a Type II right-censored life test with n items on test and $r = 1$ failure is observed at time $t_{(1)}$. Assume that the items placed on the life test have lifetimes that are well described by a Rayleigh(λ) population.

(a) What is the maximum likelihood estimator for λ?

(b) What is an exact confidence interval for λ?

(c) What is the expected width of the confidence interval from part (b)?

(d) Verify the coverage and expected width of the exact confidence interval for $\lambda = 2$, $n = 7$, and $\alpha = 0.05$ via Monte Carlo simulation.

5.23 A randomly right-censored data set is collected from a population with hazard function

$$h(t) = \theta(1+t) \qquad t \geq 0,$$

where θ is a positive parameter.

(a) Find the maximum likelihood estimator $\hat{\theta}$.

(b) Give an expression for the observed information matrix.

(c) Give an asymptotically exact confidence interval for θ based on the observed information matrix.

5.24 Candice conducts a life test in which n items are simultaneously placed on test at time 0. The test is concluded at time $c > 0$. Assuming that the lifetimes of the items are from an exponential population with mean θ, find the distribution of the *number* of failures that occur by time c.

5.25 Show that when a random sample is drawn an exponential(λ) population with Type II right censoring

$$\frac{2r\lambda}{\hat{\lambda}} \sim \chi^2(2r),$$

where $\chi^2(2r)$ is the chi-square distribution with $2r$ degrees of freedom.

5.26 Consider a Type II right censored sample of n items on test and r observed failures drawn from an exponential population with mean θ. Show that the maximum likelihood estimate $\hat{\theta}$ is unbiased.

5.27 Assume that a life test without replacement is conducted on n items from an exponential population with failure rate λ. The exact failure times are not known, but the test is terminated upon the rth ordered failure at time $t_{(r)}$. Find a point estimator for λ.

5.28 Consider a population of items with exponential(λ) lifetimes. A life test with replacement is terminated when r failures occur or at time c, whichever occurs first. This is a combination of Type I and Type II right censoring. Find the expected number of items that fail during the test as a function of λ.

5.29 For a life test of n items with exponential(λ) lifetimes (items are not replaced upon failure) which is continued until all items fail, show that

$$E\left[\hat{\lambda}\right] = \frac{n}{n-1}\lambda,$$

where λ is the population failure rate and $\hat{\lambda}$ is the maximum likelihood estimator for λ. Thus, an unbiasing constant for $\hat{\lambda}$ is $u_n = (n-1)/n$. Equivalently,

$$E\left[\frac{n-1}{n}\hat{\lambda}\right] = E\left[u_n\hat{\lambda}\right] = \lambda.$$

Find an unbiasing constant for the case of Type II right censoring.

5.30 Give a point and 95% interval estimator for the median lifetime of the 6–MP treatment group assuming that the data have been drawn from an exponential(λ) population.

5.31 Consider the following Type II right censored data set for the lifetime of a product ($n = 5$ and $r = 3$) drawn from an exponential population with failure rate λ:

$$3.6 \qquad 3.9 \qquad 8.5.$$

(a) Find the maximum likelihood estimator for the mean of the population.

(b) Find the maximum likelihood estimator for $S(5)$.

(c) Find an exact two-sided 80% confidence interval for $E\left[T^3\right]$.

(d) Find an exact one-sided 95% lower confidence interval for $S(5)$.

(e) Find the p-value for the test $H_0 : \lambda = 0.04$ versus $H_1 : \lambda > 0.04$.

(f) Find the value of the log likelihood function at the maximum likelihood estimate.

(g) Find the value of the observed information matrix.

(h) Assume the data values

$$3.8 \qquad 4.6 \qquad 6.0 \qquad 9.6$$

constitute a complete data set for a different product. Find an exact two-sided 90% confidence interval for the ratio of the failure rates of the two products if both are assumed to come from exponential populations.

5.32 Sara observes a single observed lifetime T from an exponential(λ) population, where λ is a positive unknown rate parameter. Find an exact two-sided 95% confidence interval for λ.

5.33 Justin places a single item is placed on test ($n = 1$). The only information that is available is that the item failed between times a and b, where $a < b$. In other words, the single item's lifetime is *interval censored*. Assuming that the population time to failure is exponential(λ), what is the maximum likelihood estimator of λ?

5.34 Natalie conducts a life test with $n = 19$ items on test and random right censoring. Let $t_1, t_2, \ldots, t_{19}$ be the independent exponential(2) times to failure. Let $c_1, c_2, \ldots, c_{19}$ be the independent exponential(1) censoring times, which are independent of the times to failure. Use Monte Carlo simulation to estimate the actual coverage of the following approximate confidence interval procedures for the population failure rate λ at for $\alpha = 0.05$.

(a) The confidence interval consisting of all λ satisfying

$$\frac{\hat{\lambda}\chi^2_{2r,\,1-\alpha/2}}{2r} < \lambda < \frac{\hat{\lambda}\chi^2_{2r,\,\alpha/2}}{2r}.$$

(b) The confidence interval consisting of all λ satisfying

$$\hat{\lambda} - z_{\alpha/2} O\big(\hat{\lambda}\big)^{-1/2} < \lambda < \hat{\lambda} + z_{\alpha/2} O\big(\hat{\lambda}\big)^{-1/2}.$$

(c) The confidence interval consisting of all λ satisfying

$$2[\ln L(\hat{\lambda}) - \ln L(\lambda)] < \chi^2_{1,\alpha}.$$

Replicate the experiment so as to estimate the actual coverages to three digits of accuracy.

5.35 Sixty-watt light bulb lifetimes are known to be exponentially distributed with unknown positive population mean θ from previous test results. The company that produces these light bulbs would like to estimate θ by testing n bulbs to failure at one facility and m bulbs to failure at a second facility. Let $X_1, X_2, \ldots, X_n$ be the independent lifetimes of the bulbs tested at the first facility; let $Y_1, Y_2, \ldots, Y_m$ be the independent lifetimes of the bulbs tested at the second facility. An unbiased estimate of θ is the convex combination

$$\hat{\theta} = p\hat{\theta}_X + (1-p)\hat{\theta}_Y,$$

where $0 < p < 1$, $\hat{\theta}_X = \bar{X}$ is the maximum likelihood estimator of θ for the data from the first facility, and $\hat{\theta}_Y = \bar{Y}$ is the maximum likelihood estimator of θ for the data from the second facility. Find the value of p that minimizes $V\left[\hat{\theta}\right]$.

5.36 Ash would like to test the hypothesis

$$H_0 : \lambda = 17$$

versus

$$H_1 : \lambda > 17$$

using a single value T from an exponential(λ) population, where λ is a positive unknown population failure rate. The null hypothesis is rejected if $T < 0.01$. Find the significance level α for the test.

5.37 Let T be an observation from an exponential population with positive unknown population mean θ. This observation is used to test

$$H_0 : \theta = 6$$

versus

$$H_1 : \theta = 2.$$

(a) Find the critical value for the test for a fixed significance level α.

(b) Find β for a fixed significance level α.

5.38 Paul collects a random sample $t_1, t_2, \ldots, t_n$ from an exponential population with positive unknown mean θ. Show that the sample mean, $\bar{t}$, and n times the first order statistic, $nt_{(1)}$, are both unbiased estimators of θ.

5.39 Jessica and Mary collect a random sample $t_1, t_2, \ldots, t_n$ of light bulb lifetimes drawn from an exponential(λ) population, where λ is a positive unknown failure rate. The bulbs are stamped with "1000 hour MTTF," indicating that the *mean time to failure* equals 1000 hours. They would like to determine whether there is statistical evidence in the sample that indicates the bulbs last *longer* than 1000 hours.

(a) State the appropriate H_0 and H_1.

(b) Jessica uses the test statistic $\bar{t}$ and Mary uses the test statistic $nt_{(1)}$ to test the hypothesis. Find the critical values for their test statistics when $\alpha = 0.05$ and $n = 10$.

(c) Draw the power curves associated with each of the test statistics from part (b) on the same set of axes using a computer. Again assume that $\alpha = 0.05$ and $n = 10$.

5.40 Camille observes a single lifetime T from an exponential population with a positive unknown population mean θ. She would like to test

$$H_0 : \theta = 1$$

versus

$$H_1 : \theta > 1$$

at $\alpha = 0.07$ using T as a test statistic.

(a) Find the critical value c for this test.

(b) Plot the power function for this test.

5.41 Ellen collects a random sample $t_1, t_2, \ldots, t_{10}$ of light bulb lifetimes from an exponential(λ) population, where λ is a positive unknown failure rate. Ellen is a reliability engineer. She is confident from previous test results that the time to failure for these light bulbs is exponentially distributed. She is interested in testing whether a manufacturer's claim that the population mean time to failure for the bulbs is 1000 hours. So she would like to test

$$H_0 : \lambda = 0.001$$

versus

$$H_1 : \lambda > 0.001.$$

She is in a hurry. She places ten bulbs on test and only observes the first bulb fail at $t_{(1)} = 14$ hours, and would like to draw a conclusion at 14 hours. Give the p-value for the test based on the value of this single order statistic.

5.42 Liz collects a random sample of lifetimes $t_1, t_2, \ldots, t_n$ from an exponential(λ) distribution, where λ is a positive unknown failure rate parameter. She conducts a significance test of

$$H_0 : \lambda = 1$$

versus

$$H_0 : \lambda \neq 1,$$

which achieves a p-value of $p = 0.07$ for a particular data set. If she then computes an exact two-sided 95% confidence interval for λ for this particular data set, will the confidence interval contain 1?

5.43 Karen fits the ball bearing data set to the Weibull distribution parameterized as

$$S(t) = e^{-(\lambda t)^\kappa} \qquad t \geq 0,$$

yielding maximum likelihood estimates $\hat{\lambda} = 0.0122$ and $\hat{\kappa} = 2.10$. Ute also wants to fit the same data set to the Weibull distribution, but she uses the parameterization

$$S(t) = e^{-\rho t^\beta} \qquad t \geq 0.$$

What will be the maximum likelihood estimates $\hat{\rho}$ and $\hat{\beta}$ that Ute obtains for the ball bearing data set?

5.44 Jay conducts a life test with $n = 5$ items on test which is terminated when $r = 3$ items have failed. Failed items are not replaced in this traditional Type II right-censored data set. Assuming that the time to failure of an item in the population has a Weibull(λ, κ) distribution with known, positive parameters λ and κ, what is the probability density function of the time to complete the life test?

5.45 Jennie collects a random sample $t_1, t_2, \ldots, t_7$ from a Rayleigh population with probability density function

$$f(t) = 2\theta^{-2} t e^{-(t/\theta)^2} \qquad t > 0,$$

where θ is a positive unknown parameter. She would like to test

$$H_0 : \theta = 10$$

versus

$$H_1 : \theta > 10$$

using the test statistic $t_{(1)} = \min\{t_1, t_2, \ldots, t_7\}$, which assumes the value $t_{(1)} = 6$. Find the p-value for her test.

5.46 Mildred collects a random sample $t_1, t_2, \ldots, t_n$ from a Rayleigh(λ) population with survivor function

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

where λ is a positive unknown parameter.

(a) Find the maximum likelihood estimator of λ.

(b) Show that the log likelihood function is maximized at the maximum likelihood estimator $\hat{\lambda}$.

(c) Given that the expected value of T is $E[T] = \sqrt{\pi}/(2\lambda)$, find the method of moments estimator of λ.

5.47 Find the elements of the score vector for the log logistic distribution for a randomly right-censored data set.

5.48 Bryan places n items on test and observes r failures. Assuming that the failure times of the items follow the log logistic distribution and censoring is random, set up an expression for the boundary of a 95% confidence region for the shape parameter κ and scale parameter λ of the log logistic distribution based on the likelihood ratio statistic. Assume that the survivor function for the log logistic distribution is

$$S(t) = \frac{1}{1 + (\lambda t)^\kappa} \qquad t \geq 0,$$

for $\lambda > 0$ and $\kappa > 0$. It is not necessary to solve for the maximum likelihood estimators.

5.49 Consider a proportional hazards model with $n = 3$ items on test and distinct failure times t_1, t_2, t_3. Compute the joint probability mass function values for the $3! = 6$ possible rank vectors, and show that they sum to 1.

5.50 Give the equations that must be solved in order to find the maximum likelihood estimators $\hat{\lambda}$, $\hat{\kappa}$, and $\hat{\boldsymbol{\beta}}$ for a proportional hazards model with log logistic baseline distribution and log linear link function. A random right-censoring scheme is used.

5.51 Joyce fits the Cox proportional hazards model with unknown baseline distribution given in Examples 5.14, 5.15, and 5.16 to the $n = 3$ light bulb failure times. The purpose of the study was to determine the effect of wattage on survival for 60-watt and 100-watt light bulbs.

(a) What is the value of the regression coefficient for wattage if it were coded as $z = 60$ and $z = 100$ rather than as a binary covariate?

(b) Write a short paragraph indicating whether or not these two approaches are fundamentally equivalent ways of coding the covariate. If they differ, is one method of coding the covariate superior to the other for the purpose of the study?

5.52 Survival times (in weeks) for two groups of leukemia patients (AG positive and AG negative blood types), along with an additional covariate, white blood cell count are given in Feigl, P. and Zelen, M., "Estimation of Exponential Survival Probabilities with Concomitant Information," *Biometrics*, Vol. 21, No. 4, pp. 826–838, 1965, and are displayed below.

AG positive group		AG negative group	
Survival time	White blood count	Survival time	White blood count
65	2300	56	4400
156	750	65	3000
100	4300	17	4000
134	2600	7	1500
16	6000	16	9000
108	10500	22	5300
121	10000	3	10000
4	17000	4	19000
39	5400	2	27000
143	7000	3	28000
56	9400	8	31000
26	32000	4	26000
22	35000	3	21000
1	100000	30	79000
1	100000	4	100000
5	52000	43	100000
65	100000		

(a) Fit the Cox proportional hazards model to the survival times. Code the blood type as the indicator variable z_1, using 1 for AG positive and 0 for AG negative. The second covariate z_2 is the natural logarithm of the white blood cell counts minus the sample mean of the natural logarithms of the white blood cell counts. Include the interaction term $(z_1 - \bar{z}_1)z_2$ in the model. Use the Breslow method for handling tied survival times.

(b) Write a sentence interpreting the *sign* of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ in terms of risk to the patient.

(c) Give a 95% confidence interval for β_1.

(d) If covariates associated with p-values that are less than 0.10 are considered statistically significant, what is the fitted hazard function for a leukemia patient with baseline hazard function $h_0(t)$, white blood cell count 9000 who has AG positive blood type? *Hint*: The sample mean of the natural logarithms of the white blood cell types is 9.52 and the mean of the blood types coded as an indicator variable is $17/33 = 0.515$.

5.53 Consider the Cox proportional hazards model with a single ($q = 1$) binary covariate z_1, an exponential(λ) baseline distribution, and a log linear link function. The baseline distribution can be absorbed into the link function by creating an artificial covariate $z_0 = 1$ and setting $\lambda = e^{\beta_0 z_0}$.

(a) For a randomly right-censored data set, find the score vector.

(b) For a randomly right-censored data set, find closed-form expressions for the maximum likelihood estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

(c) For the $n = 3$ observations given in vector form below, calculate the maximum likelihood estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\boldsymbol{x} = \begin{bmatrix} 80 \\ 20 \\ 50 \end{bmatrix} \qquad \boldsymbol{\delta} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{Z} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

(d) What is the hazard function of the fitted model for the data from part (c)?

(e) Use the observed information matrix to give approximate two-sided 95% confidence intervals for β_0 and β_1 for the data from part (c).

(f) Give the p-values for testing the hypotheses

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

for $i = 0, 1$, for the data from part (c).

5.54 The wattage of the $n = 3$ light bulbs in Example 5.16 was coded as the covariate $z_1 = 0$ for a 60-watt bulb and $z_1 = 1$ for a 100-watt bulb. When the Cox proportional hazards model with an unspecified baseline hazard function was fit to the data set, the point estimate for the regression parameter was $\hat{\beta}_1 = -0.347$. Without doing the derivation from scratch, what is the point estimate for the regression parameter if the wattage (that is, 60 watts or 100 watts) of the bulb were used as the covariate.

5.55 Mark fits a Cox proportional hazards model with unknown baseline distribution to a data set of drill bit failure times (measured in number of items drilled) with $q = 2$, for which the covariates denote the turning speed (revolutions per minute, rpm) and the hardness of the material (Brinell hardness number, BHN) being drilled. The turning speeds range from 2400 to 4800 rpm and the hardness of the materials ranges from 250 to 440 BHN. Interactions are not considered and the variables are not centered. The fitted model has estimated regression vector $\hat{\boldsymbol{\beta}} = (0.014, 0.45)'$, and the inverse of the observed information matrix is

$$O^{-1}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 0.000081 & 0.000016 \\ 0.000016 & 0.010000 \end{bmatrix}.$$

Write a paragraph interpreting these results.