7.5 Censoring

Censoring occurs frequently in lifetime data 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 7.11 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 manufactured or the time a product is purchased). 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*.

As mentioned at the beginning of this chapter, it is always best to obtain exact failure times, rather than grouping the failure times into intervals. So with the exception of life tables (used by actuaries, in which a time interval is typically one year), exact failure times are assumed to be known for uncensored observations.

The most frequent 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, patients can either (a) still be alive at the end of a study, (b) die of a cause other than the particular type of cancer, or (c) lose contact with the study (for example, if they leave town), constituting right-censored observations.

Three special cases of right censoring are common in reliability and life testing situations. The first is Type II or *order statistic* censoring. As shown in Figure 7.12, 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 7.12 and the failure of the first item (corresponding to the third

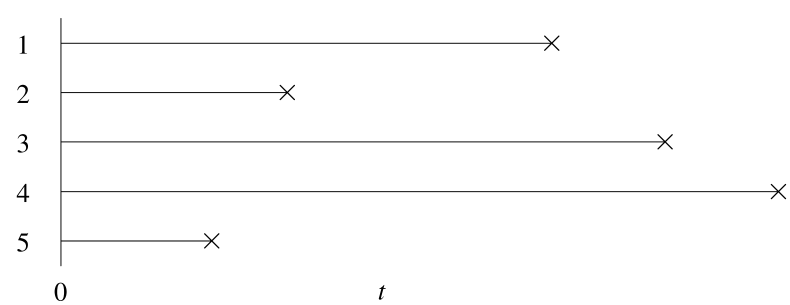

Figure 7.11: Complete data set with $n = 5$.

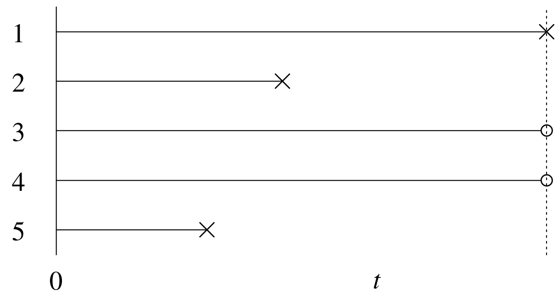

Figure 7.12: Type II right-censored data set with $n = 5$ and $r = 3$.

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.

The second special case of right censoring is Type I or *time* censoring. As shown in Figure 7.13, 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 7.13, 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 7.14 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 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.

Left censoring occurs less frequently than right censoring. An observation is left censored if only an upper bound is known on the failure time. Scientific applications for which the precision of the measuring equipment is finite (for example, in the earth sciences, when a gas cannot be measured below a threshold of six parts per million with a particular measuring device) may yield a data set

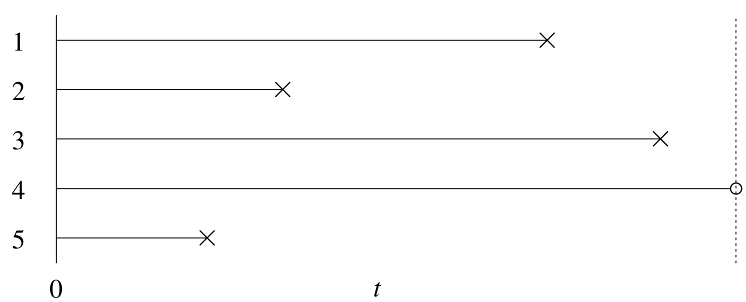

Figure 7.13: Type I right-censored data set with $n = 5$ and $r = 4$.

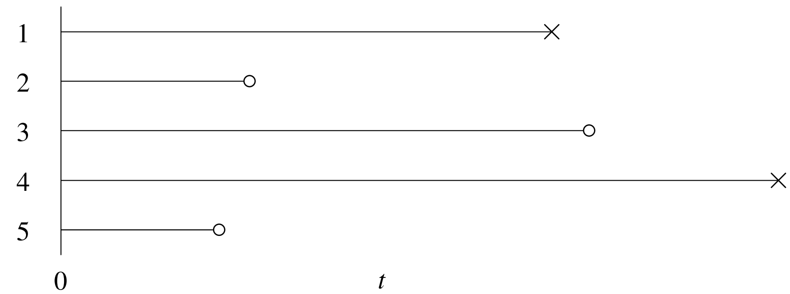

Figure 7.14: Randomly right-censored data set with $n = 5$ and $r = 2$.

containing left-censored observations.

The following example of a situation for which a data set contains both left-censored, right-censored, and observed failures is referenced at the end of the chapter. A psychiatrist traveled to an African village to determine the probability distribution of the age at which children learned to perform a particular task. In this case, the lifetime T of interest was the time between the date of birth and the date the child is able to perform the task. Those children who already knew how to perform the task when the psychiatrist arrived to the village were associated with left-censored observations; those that learned the task while he was there were associated with uncensored observations; those that did not learn the task by the time he departed were associated with right-censored observations.

Interval censoring is still another type of censoring for which the lifetime falls into an interval. This is the case when data are grouped into intervals. Interval censoring also occurs when an item is checked periodically (for example, once a week) for failure. In this case, the information known about the lifetime is that its failure time occurred during the interval prior to when failure was detected.

Current status data, sometimes known as *quantal response data*, arises frequently in reliability settings. It involves a unique form of censoring in which each item with age C_i (for $i = 1, 2, \ldots, n$) is inspected simultaneously, and is deemed to be functioning or failed. If failure has occurred upon inspection, then $T_i \le C_i$ which corresponds to a left-censored observation. If the item is functioning upon inspection, then $T_i > C_i$ which corresponds to a right-censored observation. In some applications the inspection does not require the destruction of the item and in other applications the item must be destroyed. When an entire metro system in a large city is shut down to inspect the width between railroad tracks at several random points in the network, no destruction is necessary. On the other hand, when a food scientist inspects several packaged food items to test their shelf life, the food items are consumed in the test.

A taxonomy for some of the common censoring schemes that occur in practice is shown in Figure 7.15. The vast majority of censored data sets will fall into one of the categories listed in the taxonomy.

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. Many data sets in reliability are heavily censored because the items on test have high reliabilities and the testing time is relatively short. Since the observed failures in a heavily-censored data set will consist mainly of early failures, care must be taken not to extrapolate the estimated failure time distribution beyond the largest observed failure.

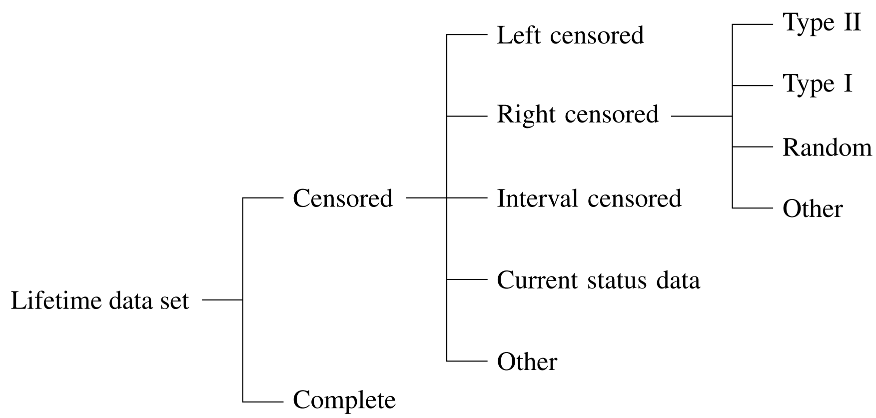

Figure 7.15: Lifetime data set taxonomy.

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 this approach is used on a right-censored data set, for example, 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 independent observations denoting lifetimes sampled randomly from a population. The corresponding right-censoring times are denoted by $c_1, c_2, \ldots, c_n$. 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 (the uncensored observations):

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

The set C contains the indexes of the items whose failure time exceeds the corresponding censoring time (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 7.13 Consider the case of $n = 5$ items placed on test as indicated in Figure 7.16. Find the sets U and C.

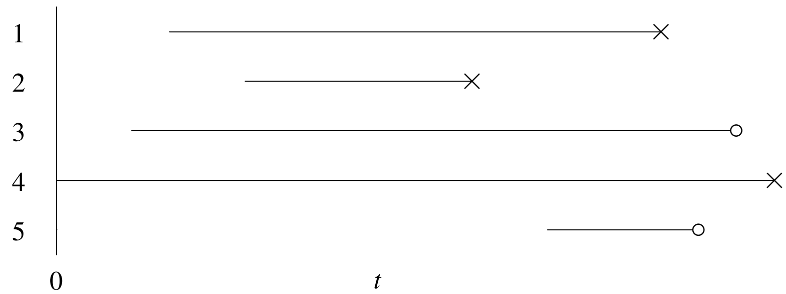

Figure 7.16: Randomly right-censored data set.

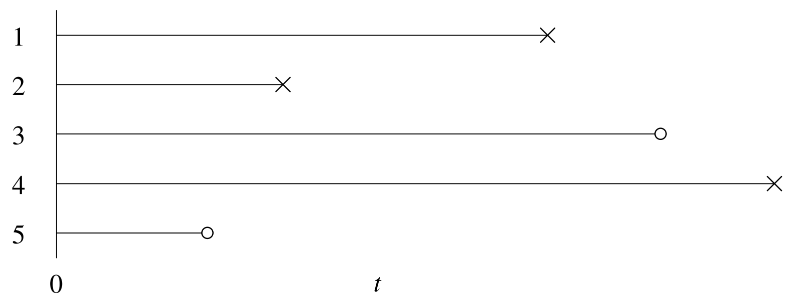

Figure 7.17: Aligned randomly right-censored data set.

Observe that the right-censored data set depicted in Figure 7.16, 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, 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 design changes, reliability growth, reliability degradation, etc. over the time window of observation, aligning the data values in this fashion is appropriate. Figure 7.17 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

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

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:

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

for $i = 1, 2, \ldots, n$. The (x_i, δ_i) pairs can be constructed from the (t_i, c_i) pairs. 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 7.14 Consider a life test with n items on test with random right censoring. 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 (a Weibull distribution with shape parameter $\kappa = 2$) 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

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

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 mean and dividing by the 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:

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

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 7.17 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

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

or

$$0.259 < \lambda < 0.937.$$

To summarize 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.