CHAPTER 1

Introduction

1.1 FAILURE TIME DATA

We consider methods for the analysis of data when the response of interest is the time until some event occurs. Such events are generically referred to as *failures*, although the event may, for instance, be the performance of a certain task in a learning experiment in psychology or a change of residence in a demographic study. Major areas of application, however, are biomedical studies and industrial life testing.

We assume that observations are available on the failure time of n individuals usually taken to be independent. A principal problem examined is that of developing methods for assessing the dependence of failure time on explanatory variables. Typically, such explanatory variables will describe prestudy heterogeneity in the experimental material or differential allocations of treatments resulting from the study design. A secondary problem involves the estimation and specification of models for the underlying failure time distribution.

Additional problems arise in the analysis of multivariate failure times and failure types. These problems entail assessing the frequency of recurrent failures and estimating the correlation among failure times and types. There are a number of reasons why special methods and special treatment is required for failure time data, and it is convenient to illustrate some of the distinguishing features through the following examples.

1.1.1 Carcinogenesis

Table 1.1 gives the times from insult with the carcinogen DMBA to mortality from vaginal cancer in rats. Two groups were distinguished by a pretreatment regimen. We might consider comparing the two regimes using the *t*-test (presumably to transformed data) or one of several nonparametric tests. Such procedures cannot be applied immediately, however, because of a feature very prevalent in failure time studies. Specifically, four failure times in Table 1.1 are *censored*. For these four rats, we can see that the failure times exceed 216, 244, 204, and 344 days,

Table 1.1   Days to Vaginal Cancer Mortality in Rats

Group 1	143,	164,	188,	188,	190,	192,	206,	209,	213,	216,	220
	227,	230,	234,	246,	265,	304,	216*,	244*			
Group 2	142,	156,	163,	198,	205,	232,	232,	233,	233,	233,	233
	239,	240,	261,	280,	280,	296,	296,	323,	204*,	344*	

Source: Pike (1966).

* These four items are right censored.

respectively, but we do not know the failure times exactly. In this example, the (right) censoring may have arisen because these four rats died of causes unrelated to application of the carcinogen and were free of tumor at death, or they may simply not have died by the time of data analysis. The necessity of obtaining methods of analysis that accommodate censoring has been a principal motivating factor for the development of specialized models and procedures for failure time data.

A larger set of animal carcinogenesis data is given in Appendix A (data set V). Two groups of male mice were given 300 rads of radiation and followed for cancer incidence. One group was maintained in a germ-free environment. The new feature of these data is that more than one failure mode occurs. It is of interest, for example, to evaluate the effect of a germ-free environment on the incidence rate of reticulum cell sarcoma while accommodating the competing risks of developing thymic lymphoma or other causes of failure.

1.1.2   Randomized Clinical Trial

Table 1.2 gives some data from a randomized clinical trial on 64 patients with severe aplastic anemia. Prior to the trial, all the patients were treated with high-dose cyclophosphamide followed by an infusion of bone marrow from an HLA-identical family member. Patients were then assigned to each of two treatment groups: cyclosporine and methotrexate (CSP + MTX) or methotrexate alone (MTX). One endpoint of interest was the time from assignment until the diagnosis of a life-threatening stage (≥ 2) of acute graft versus host disease (AGVHD). The times are given in days. Also included are two covariates measured at the outset: the patient's age in years at the time of transplant and an indicator of whether or not the patient was assigned to a laminar airflow (LAF) isolation room. Storb et al. (1986) report on the subset of 46 patients who were randomly assigned to treatment, with stratification by age group and LAF. For purposes of illustration, we shall treat the data as though all 64 patients had been randomly assigned. In this trial, only 20 of the 64 patients actually reached the endpoint; the remaining 44 patients were right censored.

Appendix A (data set II) gives a part of the data from a much larger clinical trial carried out by the Radiation Therapy Oncology Group. The full study included patients with squamous cell carcinoma of 15 sites in the mouth and throat, with 16 participating institutions, although only the data on three sites in the oropharynx

Table 1.2   Time in Days to Severe (Stage $\geq$ 2) Acute Graft Versus Host Disease (AGVHD), Death, or Last Contact for Bone Marrow Transplant Patients Treated with Cyclosporine and Methotrexate (CSP + MTX) or with MTX Only [a]

CSP + MTX						MTX					
Time	LAF	Age	Time	LAF	Age	Time	LAF	Age	Time	LAF	Age
3*	0	40	324*	0	23	9	1	35	104*	1	27
8	1	21	356*	1	13	11	1	27	106*	1	19
10	1	18	378*	1	34	12	0	22	156*	1	15
12*	0	42	408*	1	27	20	1	21	218*	1	26
16	0	23	411*	1	5	20	1	30	230*	0	11
17	0	21	420*	1	23	22	0	7	231*	1	14
22	1	13	449*	1	37	25	1	36	316*	1	15
64*	0	20	490*	1	37	25	1	38	393*	7	27
65*	1	15	528*	1	32	25*	0	20	395*	0	2
77*	1	34	547*	1	32	28	0	25	428*	0	3
82*	1	14	691*	1	38	28	0	28	469*	1	14
98*	1	10	769*	0	18	31	1	17	602*	1	18
155*	0	27	1111*	0	20	35	1	21	681*	0	23
189*	1	9	1173*	0	12	35	1	25	690*	1	9
199*	1	19	1213*	0	12	46	1	35	1112*	1	11
247*	1	14	1357*	0	29	49	0	19	1180*	0	11

[a] Asterisks indicate that time to severe AGVHD is right censored; that is, the patient died without severe AGVHD or was without severe AGVHD at last contact.

reported by the six largest institutions are given. Patients entering the study were randomly assigned to one of two treatment groups: radiation therapy alone or radiation therapy together with a chemotherapeutic agent. One objective of the study was to compare the two treatment policies with respect to patient survival.

Approximately 30% of the survival times are censored, owing primarily to patients surviving to the time of analysis. Some patients were lost to follow up because the patient moved and was unable to continue, but these cases were relatively rare. From a statistical point of view, a key feature of these data is the considerable lack of homogeneity between individuals being studied. Of course, as a part of the study design, certain criteria for patient eligibility had to be met which eliminated extremes in the extent of disease, but still many factors are not controlled. This study included measurements of many covariates that would be expected to relate to survival experience. Six such variables are given in the data of Appendix A (sex, T staging, N staging, age, general condition, and grade). The site of the primary tumor and possible differences between participating institutions require consideration as well.

The TN staging classification gives a measure of the extent of the tumor at the primary site and at regional lymph nodes. T_1 refers to a small primary tumor, 2 cm or less in largest diameter, whereas T_4 is a massive tumor with extension to adjoining tissue. T_2 and T_3 refer to intermediate cases. N_0 refers to the absence of clinical

evidence of a lymph node metastasis and N_1, N_2, and N_3 indicate, in increasing magnitude, the extent of existing lymph/node involvement. Patients with classifications T_1N_0, T_1N_1, T_2N_0, or T_2N_1 or with distant metastasis were excluded from study.

The variable "general condition" gives a measure of the functional capacity of the patient at the time of diagnosis (1 refers to no disability, whereas 4 denotes bed confinement; 2 and 3 refer to intermediate levels). The variable grade is a measure of the degree of differentiation of the tumor (the degree to which the tumor cell resembles the host cell) from 1 (well differentiated) to 3 (poorly differentiated).

In addition to the primary question of whether the combined treatment mode is preferable to the conventional radiation therapy, it is of considerable interest to determine the extent to which the several covariates are related to subsequent survival. In answering the primary question, it may also be important to adjust the survival times for possible imbalance that may be present in the study with regard to the other covariates. Such problems are similar to those encountered in the classical theory of regression and the analysis of covariance. Again, the need to accommodate censoring is an important distinguishing point. In many situations, nonparametric and robust procedures are desirable since there is frequently little empirical or theoretical work to support a particular family of failure time distributions.

1.1.3 Heart Transplant Data

Crowley and Hu (1977) give survival times of potential heart transplant recipients from their date of acceptance into the Stanford heart transplant program. These data are reproduced in Appendix A, data set IV. One problem of considerable interest is to evaluate the effect of heart transplantation on subsequent survival.

For each study subject the explanatory variables "age" and "prior surgery" were recorded. There were also donor–recipient variables that may be predictive of post-transplant survival time. The main new feature here is that patients change treatment status during the course of the study. Specifically, a patient is part of the control group until a suitable donor is located and transplantation takes place, at which time he or she joins the treatment group. Correspondingly, some explanatory variables, such as waiting time for transplant, are observed during the course of the study and depend on the time elapsed to transplant. This study is examined in some detail in Chapter 6 using the ideas of time-dependent covariates and time-dependent stratification.

The existence of covariates that change over time is yet another unusual feature of failure time data that requires special methods and attention to model characteristics and implications. Transplant studies, such as the heart transplant study, provide a class of examples where such covariates arise because of the very nature of the treatment. Alternatively, we can imagine a system operating under stress where the stress factor is varied as time elapses. In such a situation, it would be common to examine the relationship between the stress applied now and the current risk of failure. Other examples arise in clinical studies, such as, for example, measures

of immune function taken at regular intervals for leukemia patients in remission. One may wish, in this instance, to study the relationship between changes in immune function and corresponding propensity to relapse. Such examples are also discussed in Chapter 6. In comparative trials, time-dependent covariates such as measures of immune function can be *responsive*; that is, they can be affected by the treatments under investigation. Responsive covariates have the potential to be useful in examining the mechanism of a treatment effect (does the treatment work by improving immune function?) or even in serving as a surrogate for the primary failure time outcome. If, however, they are treated as ordinary covariates in a regression model to investigate the effect of treatments, they can mask a treatment effect.

1.1.4 Accelerated Life Test

Nelson and Hahn (1972) present data on the number of hours to failure of motorettes operating under various temperatures. The name *accelerated life test* for this type of study derives from the use of a stress factor, in this case temperature, to increase the rate of failure over that which would be observed under normal operating conditions. The data are presented in Table 1.3 and exhibit severe censoring, with only 17 of 40 motorettes failing. Note that the stress (temperature) is constant for any particular motorette over time. The principal interest in such a study involves determination of the relationship between failure time and temperature for the purpose of extrapolating to usual running temperatures. Of course, the validity of such an extrapolation depends on the constancy of certain relationships over a very wide range of temperatures. For this study, the failure time distribution at the regular operating temperature of 130°C was of interest.

As in earlier examples, the censoring here is *type I* or *time censoring*. That is, censored survival times were observed only if failure had not occurred prior to a predetermined time at which the study was to be terminated. Experiments of this type, where considerable control is available to the experimenter, offer the possibility of other censoring schemes. For instance, in the study above it might have been decided in advance to continue the study until specified numbers of motorettes had failed at each of the temperatures (e.g., until one, three, five, and seven motorettes had failed at 150°C, 170°C, 190°C, and 220°C, respectively). Such censoring is usually referred to as *type II* or *order statistic censoring*, in that the study terminates as soon as certain order statistics are observed. With certain models, some

Table 1.3 Hours to Failure of Motorettes

150°C	All 10 motorettes without failure at 8064 hours
170°C	1764, 2772, 3444, 3542, 3780, 4860, 5196
	3 motorettes without failure at 5448 hours
190°C	408, 408, 1344, 1344, 1440
	5 motorettes without failure at 1680 hours
220°C	408, 408, 504, 504, 504
	5 motorettes without failure at 528 hours

inferential procedures (e.g., exact significance tests) are simpler for type II than for type I censoring. It should be noted, however, that type II censoring usually does not allow an upper bound to be placed on the total duration of the study and is generally not a feasible study design if there is staggered entry to the study.

Some of the examples above are considered further throughout the book. We turn now, however, to mathematical representations of failure times and consider the very simplest case of an independent sample from a homogeneous population (no explanatory variables) with a single failure mode.

1.2   FAILURE TIME DISTRIBUTIONS

Let T be a nonnegative random variable representing the failure time of an individual from a homogeneous population. The probability distribution of T can be specified in many ways, three of which are particularly useful in survival applications: the survivor function, the probability density function, and the hazard function. Interrelations among these three representations are given below for discrete and continuous distributions.

The *survivor function* is defined for discrete and continuous distributions by the probability that T exceeds a value t in its range; that is,

$$F(t) = P(T > t), \qquad 0 < t < \infty.$$

Note that F in some settings refers to the cumulative distribution function, $P(T \leq t)$, and therefore gives the probabilities in the left tail rather than in the right tail of the distribution. The right tail, however, is the important component for the incorporation of right censoring, so it is more convenient to concentrate on the survivor function in dealing with failure time distributions. Clearly, $F(t)$ is a nonincreasing right-continuous function of t with $F(0) = 1$ and $\lim_{t \to \infty} F(t) = 0$.

1.2.1 T (Absolutely) Continuous

The *probability density function* (PDF) of T is

$$f(t) = -dF(t)/dt.$$

The range of T is $[0, \infty)$, and this should be understood as the domain of definition for functions of t. It is convenient to remember that $f(t)$ gives the density of probability at t and for h small has the interpretation

$$f(t)h \simeq P(t \leq T < t + h) = F(t) - F(t + h),$$

provided that $f(t)$ is continuous at t. We note also that $f(t) \geq 0$, $\int_0^\infty f(t)\, dt = 1$, and

$$F(t) = \int_t^\infty f(s)\, ds.$$

The *hazard function* is defined as

$$\lambda(t) = \lim_{h \to 0^+} P(t \le T < t + h \mid T \ge t)/h \tag{1.1}$$

and specifies the instantaneous rate at which failures occur for items that are surviving at time t. The hazard function fully specifies the distribution of t and so determines both the density and the survivor functions. From (1.1) and using the definition of the density function, it follows that

$$\lambda(t) = -f(t)/F(t)$$
$$= -d \log F(t)/dt.$$

Now integrating with respect to t and using $F(0) = 1$, we obtain

$$F(t) = \exp\left[-\int_0^t \lambda(s)\, ds\right]$$
$$= \exp[-\Lambda(t)], \tag{1.2}$$

where $\Lambda(t) = \int_0^t \lambda(s)\, ds$ is called the *cumulative hazard function*. The PDF of T can be obtained by differentiating (1.2) to find that

$$f(t) = \lambda(t)\exp[-\Lambda(t)]. \tag{1.3}$$

Examination of (1.2) indicates that any nonnegative function $\lambda(t)$ that satisfies

$$\int_0^t \lambda(s)\, ds < \infty$$

for some $t > 0$ and

$$\int_0^\infty \lambda(s)\, ds = \infty$$

can be the hazard function of a continuous random variable.

Other representations of the failure time distribution are occasionally useful. An example is the *expected residual life* at time t,

$$r(t) = E(T - t \mid T \ge t),$$

which uniquely determines a continuous survival distribution with finite mean. To see this, note that

$$r(t) = \frac{\int_t^\infty (s - t) f(s)\, ds}{F(t)}$$

and integrate by parts to obtain

$$r(t) = \frac{\int_t^\infty F(s)\,ds}{F(t)}, \tag{1.4}$$

where we have used the fact that $E(T) < \infty$ implies that $\lim_{t \to \infty} tF(t) = 0$. Substituting $t = 0$ in (1.4) gives the useful result

$$E(T) = r(0) = \int_0^\infty F(s)\,ds. \tag{1.5}$$

Taking the reciprocal of both sides of (1.4), we obtain

$$\frac{1}{r(t)} = -\frac{d}{dt} \log \int_t^\infty F(s)\,ds,$$

so that

$$\int_0^t \frac{ds}{r(s)} = -\log \int_t^\infty F(s)\,ds + \log r(0).$$

This leads finally to the expression

$$F(t) = \frac{r(0)}{r(t)} \exp\left[-\int_0^t \frac{du}{r(u)}\right]$$

for the survivor function.

1.2.2   *T* Discrete

If T is a discrete random variable taking values $a_1 < a_2 < \cdots$ with associated probability function

$$f(a_i) = P(T = a_i), \qquad i = 1, 2, \ldots,$$

the survivor function is

$$F(t) = \sum_{j|a_j > t} f(x_j).$$

The hazard at a_i is defined as the conditional probability of failure at a_i given that the individual has survived to a_i,

$$\lambda_i = P(T = a_i \mid T \geq a_i) = \frac{f(a_i)}{F(a_i^-)}, \qquad i = 1, 2, \ldots,$$

where $F(a^-) = \lim_{t \to a^-} F(t)$. Corresponding to (1.2) and (1.3), the survivor function and the probability function are given by

$$F(t) = \prod_{j|a_j \leq t} (1 - \lambda_j) \tag{1.6}$$

and

$$f(a_i) = \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j). \tag{1.7}$$

As in the continuous case, the discrete hazard function $(\lambda_i, i = 1, 2, \ldots)$ uniquely determines the distribution of the failure time variable T.

The results in (1.6) and (1.7) are quite easily deduced by considering the failure time process unfolding over time and a sequence of trials, each of which may or may not result in a failure. For example, the result in (1.7) follows from noting that an individual fails at time a_i if and only if:

- The individual survives in sequence each of the preceding discrete failure times $a_1, \ldots, a_{i-1}$ with corresponding (conditional) probabilities $(1 - \lambda_1), \ldots, (1 - \lambda_{i-1})$.
- Having survived to a_i, the individual fails at a_i with (conditional) probability λ_i.

1.2.3 *T* has Discrete and Continuous Components

More generally, the distribution of T may have both discrete and continuous components. In this case, the hazard function can be defined to have the continuous component $\lambda_c(t)$ and discrete components $\lambda_1, \lambda_2, \ldots$ at the discrete times $a_1 < a_2 < \cdots$. The overall survivor function can then be written

$$F(t) = \exp\left[-\int_0^t \lambda_c(u)\, du\right] \prod_{j|a_j \leq t} (1 - \lambda_j).$$

The discrete, mixed, and continuous cases can be combined. The cumulative hazard function,

$$\Lambda(t) = \int_o^t \lambda_c(u)\, du + \sum_{j|a_j \leq t} \lambda_j,$$

is a right-continuous nondecreasing function. From $\Lambda(t)$ we define the differential increment

$$
\begin{aligned}
d\Lambda(t) &= \Lambda(t^- + dt) - \Lambda(t^-) \\
&= P[T \in [t, t+dt)|T \geq t] \\
&= \begin{cases} \lambda_i, & t = a_i, \quad i = 1, 2, \ldots \\ \lambda_c(t)\, dt, & \text{otherwise.} \end{cases}
\end{aligned}
$$

which specifies the hazard of failure over the infinitesimal interval $[t, t + dt)$.

The survivor function in the discrete, continuous, or mixed cases can then be written as

$$F(t) = \mathscr{P}_0^t[1 - d\Lambda(u)], \tag{1.8}$$

where the *product integral* $\mathscr{P}$ is defined by

$$\mathscr{P}_0^t[1 - d\Lambda(u)] = \lim \prod_{k=1}^{r} \{1 - [\Lambda(u_k) - \Lambda(u_{k-1})]\}.$$

Here $0 = u_0 < u_1 < \cdots < u_r = t$ and the limit is taken as $r \to \infty$ and $\max(u_i - u_{i-1}) \to 0$. In the continuous case ($\lambda_i = 0$ for all i), it can be shown that this reduces to

$$F(t) = \mathscr{P}_0^t[1 - d\Lambda(u)] = \mathscr{P}_0^t[1 - \lambda_c(u)\,du] = \exp\left[-\int_0^t \lambda_c(u)\,du\right].$$

In the discrete case $[\lambda_c(t) = 0$ for all $t]$, it is easily seen that

$$\mathscr{P}_0^t[1 - d\Lambda(u)] = \prod_{j|a_j \leq t} (1 - \lambda_i).$$

This unification shows that failure time data can be considered to arise in essentially the same way in both the discrete and continuous cases. The product representation in (1.8) can be thought of as describing a coin-tossing experiment in which the probability of heads varies over time. The coin is tossed repeatedly and failure corresponds to the first occurrence of a tail. Thus, in general, the survival probability at time t is obtained by taking the product of the conditional survival probabilities $1 - d\Lambda(u)$ over infinitesimal intervals up to time t. This way of viewing a failure mechanism has led to many developments in the area and is crucial in understanding many of the ideas and techniques. In effect, it is possible to examine survival experience by looking at the survival experience over each interval conditional upon the experience to that point. Simple arguments for estimating the survivor function (Section 1.4) or for constructing censored data tests (Section 1.5) depend on this idea. It also underlies failure time analysis by counting processes and martingales (Chapter 5), the construction of the likelihood under independent censoring (Section 6.2), the construction of partial likelihood in the Cox model (Section 4.3), and the analysis of multivariate failure times and life-history processes (Chapter 9).

Note that $f(t)$ and $F(t)$ [or more usually, the cumulative distribution function $\bar{F}(t) = 1 - F(t)$] are common representations of the distribution of a random variable. The hazard function $\lambda(t)$ is a more specialized characterization but is particularly useful in modeling survival time data. In many instances, information is
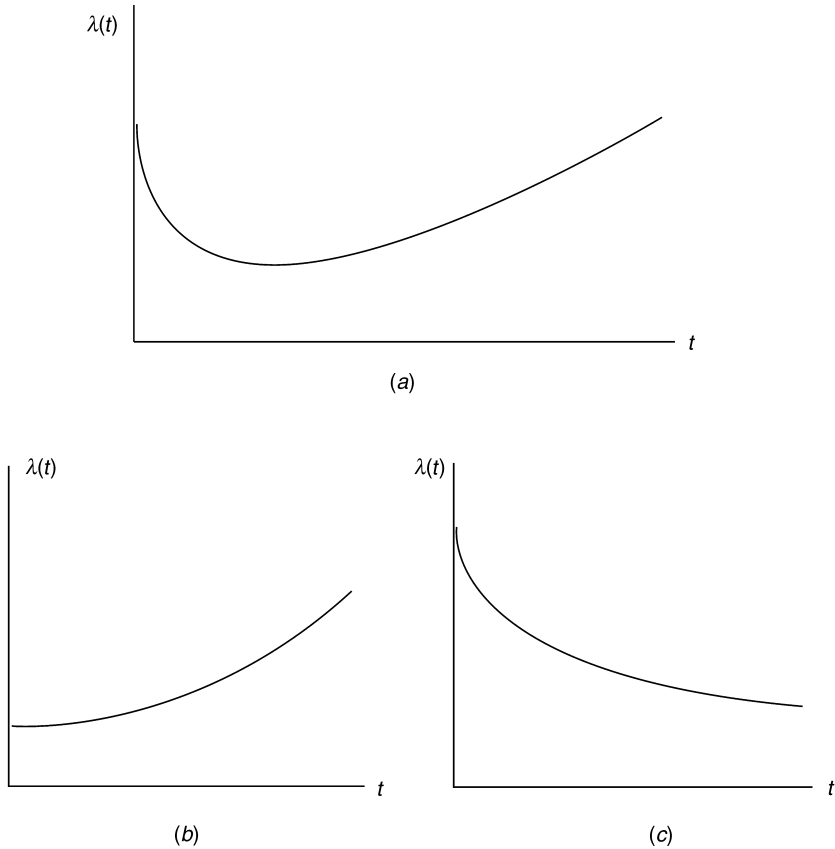
Figure 1.1 Examples of hazard functions: (*a*) hazard for human mortality; (*b*) positive aging; (*c*) negative aging.

available as to how failure rates change with the amount of time on test. This information can be used to model $\lambda(t)$ and easily translated into implications for $F(t)$ and $f(t)$ using the formulas above. For example, in modeling age at death of human populations, it is clear that initially, $\lambda(t)$ is elevated, owing to infant mortality and childhood diseases. This is followed by a period of relatively low mortality, after which the mortality rate increases very rapidly (see Figure 1.1*a*). In other applications, monotone increasing hazards (positive aging) or decreasing hazards (negative aging) may be suggested (Figure 1.1*b* and *c*). Such qualitative information on $\lambda(t)$ can be useful in selecting a family of probability models for *T*. In Chapter 2 we discuss and examine some commonly used models for failure time and their associated hazard functions.

In the discussion above, we have specified models for a homogeneous population in which all individuals independently experience the same probability laws governing their failure. As noted earlier, there are many applications where we

wish to incorporate measured covariates into the model. With covariates x measured at the time origin of the study, we can then think of models for the corresponding hazard function

$$\lambda(t; x) = \lim_{h \to 0} P\{T \in [t, t + h] | T \geq t, x\}/h,$$

which applies to those individuals with covariate value x. Corresponding to this, there are density and survivor functions, written $f(t; x)$ and $F(t; x)$, respectively.

1.3 TIME ORIGINS, CENSORING, AND TRUNCATION

In considering failure time data, it is important to have a clear and unambiguous definition of the time origin from which survival is measured. In some instances, time may represent age, with the time origin the birth of the individual. In other instances, the natural time origin may be the occurrence of some event, such as randomization or entry into a study or diagnosis of a particular disease. In like manner, one must have a clear definition of what constitutes failure. For example, in a trial to compare treatments of heart disease, one might take previous documented occurrence of a heart attack as providing eligibility for study. The time origin might be admission and randomization to the study, and failure may correspond to the recurrence of a heart attack. One would need to define carefully the clinical medical conditions that correspond to failure (and eligibility for the study). We will not talk about this further, but the clear identification of an origin and an endpoint are crucial applied aspects of failure time studies.

As noted earlier, failure time data often include some individuals who do not fail during their observation period; the data on these individuals are said to be *right censored*. In some situations, right censoring arises simply because some individuals are still surviving at the time that the study is terminated and the analysis is done. In other instances, individuals may move away from the study area for reasons unconnected with the failure time endpoint, so contact is lost. In yet other instances, individuals may be withdrawn or decide to withdraw from the study because of a worsening or improving prognosis. As is intuitively apparent, some censoring mechanisms have the potential to introduce bias into the estimation of survival probabilities or into treatment comparisons.

A right-censoring mechanism is said to be *independent* if the failure rates that apply to individuals on trial at each time $t > 0$ are the same as those that would have applied had there been no censoring. We discuss this idea more thoroughly in Chapter 6, but a brief discussion here is useful to set the stage. Suppose that the failure rate at time t that applies in the absence of censoring for an individual selected at random from a group with covariate value x is $\lambda(t; x)$. Here, as before, x consists of measurements taken on the individual at the time that he or she enters the study, such as age, sex, measures of physical condition, and so on. Suppose that within this group, individuals are to be censored according to a specific mechanism.

Consider the subset of individuals who are at risk of failure (neither failed nor censored) at some time $t > 0$. The censoring mechanism or scheme is independent if for an individual selected at random from this subset, the failure rate is $\lambda(t; x)$. Thus we require that at each time t,

$$\lim_{h \to 0} \frac{P\{T \in [t, t+h)|x, T \geq t\}}{h} = \lim_{h \to 0} \frac{P\{T \in [t, t+h)|x, T \geq t, Y(t) = 1\}}{h}, \quad (1.9)$$

where $Y(t) = 1$ indicates that the individual has neither failed nor been censored prior to time t (is at risk of failure at time t). If the censoring scheme is independent, it can be shown that an individual who is censored at time t contributes the term $P(T > t; x) = F(t; x)$ to the likelihood. Thus the information that the individual is censored at time t tells us only that the time to failure exceeds t.

As mentioned, independent censoring is examined more fully in Chapter 6. It is interesting to note, however, that some standard censoring schemes are independent. Consider, for example, a random censorship model where the ith individual has a time T_i to failure and a time C_i to censoring. Given the covariate value x_i, we suppose that C_i and T_i are independent random variables. Further, conditional on the x_i's, (T_i, C_i) are independent, $i = 1, \ldots, n$, where n is the number of subjects in the study. The time T_i to failure is observed if $T_i \leq C_i$. Otherwise, the individual is censored at C_i. For this case, it is easy to see that

$$\lim_{h \to 0} \frac{P\{T_i \in [t, t+h)|x_i, T_i \geq t\}}{h} = \lim_{h \to 0} \frac{P\{T \in [t, t+h)|x_i, T_i \geq t, C_i \geq t\}}{h},$$

which is equivalent to the condition (1.9). Type II censoring, in which individuals are put on trial until the kth item fails, for some fixed k, was discussed briefly Section 1.1.4. This censoring scheme is also independent.

In general, a censoring scheme is independent if the probability of censoring at each time t depends only on the covariate x, the observed pattern of failures and censoring up to time t in the trial, or on random processes that are independent of the failure times in the trial. Mechanisms in which the failure times of individuals are censored because the individuals appear to be at unusually high (or low) risk of failure are not independent. For these mechanisms, the condition (1.9) is violated, and the basic methods of survival analysis are not valid. Because of this, it is very important to follow the individuals entered into a study as completely as possible, so that the possibility of dependent censoring is minimized.

In some studies, individuals are not identified for observation at their respective time origin, but rather, at the occurrence of a subsequent event. Thus, there is a larger group of individuals who could have been observed, but the study is comprised of a subset of those in the cohort who experience some intermediate event. For these individuals, we observe the time origin and the follow-up time until they fail or are censored. For example, suppose that is the chosen time variable, so that time of birth is the time origin. Interest centers on the group of individuals who

were exposed to some environmental risk, and individuals are identified for study at the time they respond to an advertisement. Any individuals who died prior to the advertisement are not observed, and in fact may not even be known to exist. Those who are observed are subject to *delayed entry* or *left truncation*. There is a condition similar to (1.9) for independent left truncation which requires that the failure rates of individuals under observation at time t are representative of those in the study population. Many of the methods and analyses that we discuss extend easily to allow for independent left truncation as well as independent right censoring.

Individuals can also be subject to *left censoring*, which occurs if the individual is observed to fail prior to some time t, but the actual time of failure is otherwise unknown. In this case, we observe that $T \in [0, t]$, which is analogous to right censoring, where we observe that $T \in (t, \infty)$. Left censoring should not be confused with left truncation, as discussed in the preceding paragraph. With left censoring, we know the individual exists and failed prior to the time t. With left truncation, the existence of an individual who fails before the beginning of observation is hidden from us.

Other types of censoring also arise. For example, in some situations individuals are interval censored, so we observe only that the failure time falls within some interval $T \in (a, b)$. One might also have situations in which individuals are subject to right truncation. That is, an individual is observed if and only if its failure time is less than some given time t. Exercise 1.13 gives an example. We discuss these more general censoring schemes in Chapter 3 in the context of parametric analyses. Most of our attention, however, is focused on independent right censoring and extensions to allow independent delayed entry or left truncation.

1.4   ESTIMATION OF THE SURVIVOR FUNCTION

1.4.1   Kaplan–Meier or Product Limit Estimator

The *empirical distribution function*,

$$\bar{F}_n(x) = \frac{\text{no. sample values} \le x}{n}$$

is a simple estimate of the distribution function $\bar{F}(x) = P(X \le x)$ and is a familiar and convenient way to summarize and display data. A plot of $\bar{F}_n(x)$ versus x visually represents the sample and provides full information on the percentile points, the dispersion, and the general features of the sample distribution. Besides these obvious descriptive uses, it is an indispensable aid in studying the distributional shape of the population from which the sample arose; in fact, the empirical distribution function can serve as a basic tool in constructing formal tests of goodness of fit of the data to hypothesized probability models (see, e.g., Cox and Hinkley, 1974, pp. 69ff.).

In the analysis of survival data, it is very often useful to summarize the survival experience of particular groups of patients in terms of the empirical survivor

function. If an uncensored sample of n distinct failure times is observed from a continuous homogeneous population, the sample survivor function $F_n(t) = 1 - \bar{F}_n(t)$ is a step function that decreases by n^{-1} at each failure time observed. As noted earlier, survival data very often involve right censoring, and in this case a convenient method for estimating $F(t)$ is required.

Let $t_1 < t_2 < \cdots < t_k$ represent the observed failure times in a sample of size $n = n_0$ from a homogeneous population with (unknown) survivor function F. Suppose that d_j items fail at t_j and m_j items are censored in the interval $[t_j, t_{j+1})$ at times $t_{j1}, \ldots, t_{jm_j}, j = 0, \ldots, k$, where $t_0 = 0$ and $t_{k+1} = \infty$. Let $n_j = (m_j + d_j) + \cdots + (m_k + d_k)$ denote the number of items at risk at a time just prior to t_j. The probability of failure at t_j is

$$P(T = t_j) = F(t_j^-) - F(t_j).$$

We assume that the contribution to the likelihood of a censored survival time at t_{jl} is

$$P(T > t_{jl}) = F(t_{jl}).$$

Here we are assuming that the observed censoring time t_{jl} tells us only that the unobserved failure time is greater than t_{jl}. This is appropriate provided that the censoring is independent, as discussed in Section 1.3.

The probability of the data is then of the form

$$L = \prod_{j=0}^{k} \left\{ \left[F(t_j^-) - F(t_j) \right]^{d_j} \prod_{l=1}^{m_j} F(t_{jl}) \right\},$$

which, given the data, can be viewed as a likelihood function on the space of all survivor functions F. The (nonparametric) maximum likelihood estimate (MLE) is the survivor function $\hat{F}$ that maximizes L.

Clearly, $\hat{F}(t)$ is discontinuous at the failure times observed (i.e., places some positive probability mass at each t_j) since otherwise, $L = 0$. Further, since $t_{jl} \geq t_j$, $F(t_{jl})$ is maximized by taking $F(t_{jl}) = F(t_j)$ $(j = 1, \ldots, k; l = 1, \ldots, m_j)$. The required MLE, $\hat{F}(t)$, is therefore a discrete survivor function with hazard components $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ at $t_1, \ldots, t_k$, respectively. Thus

$$\hat{F}(t_j) = \prod_{l=1}^{j} (1 - \hat{\lambda}_l) \tag{1.10}$$

and

$$\hat{F}(t_j^-) = \prod_{l=1}^{j-1} (1 - \hat{\lambda}_l), \tag{1.11}$$

Table 1.4   Kaplan–Meier Survivor Function Estimates for Carcinogenesis Data

	Group 1					Group 2			
t_i	d_i	n_i	$\hat{F}(t_i)$	$\widehat{\text{var}}\,(\hat{F})$	t_i	d_i	n_i	$\hat{F}(t_i)$	$\widehat{\text{var}}\,(\hat{F})$
143	1	19	0.947	0.00262	142	1	21	0.952	0.00216
164	1	18	0.895	0.00496	156	1	20	0.905	0.00410
188	2	17	0.789	0.00875	163	1	19	0.857	0.00583
190	1	15	0.737	0.01021	198	1	18	0.810	0.00734
192	1	14	0.684	0.01137	205	1	16	0.759	0.00885
206	1	13	0.632	0.01225	232	2	15	0.658	0.01109
209	1	12	0.579	0.01283	233	4	13	0.455	0.01240
213	1	11	0.526	0.01312	239	1	9	0.405	0.01208
216	1	10	0.474	0.01312	240	1	8	0.345	0.01148
220	1	8	0.414	0.01311	261	1	7	0.304	0.01067
227	1	7	0.355	0.01264	280	2	6	0.202	0.00814
230	1	6	0.296	0.01170	296	2	4	0.101	0.00459
234	1	5	0.237	0.01029	323	1	2	0.051	0.00243
246	1	3	0.158	0.00873					
265	1	2	0.079	0.00530					
304	1	1	0.000						

where the $\hat{\lambda}_l$'s are chosen to maximize the function

$$\prod_{j=1}^{k}\left[\lambda_j^{d_j}\prod_{l=1}^{j-1}(1-\lambda_l)^{d_j}\prod_{l=1}^{j}(1-\lambda_l)^{m_j}\right]=\prod_{j=1}^{k}\lambda_j^{d_j}(1-\lambda_j)^{n_j-d_j}, \qquad (1.12)$$

obtained by substituting (1.10) and (1.11) in L. Clearly, $\hat{\lambda}_j = d_j/n_j (j = 1, \ldots, k)$ and the Kaplan–Meier or *product limit estimate* of the survivor function is

$$\hat{F}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j}. \qquad (1.13)$$

In the product limit estimate, we are in effect making the estimated hazard or conditional probability of failure at each t_j agree exactly with the observed proportion (d_j/n_j) of the n_j individuals at risk who fail at t_j. Again we are viewing the survival experience sequentially and at each failure time estimating the hazard of failure to be the observed proportion of failures. It should be noted that $\hat{F}(t)$ never reduces to zero if $m_k > 0$. In this instance, the largest time recorded is censored and it is usual to take $\hat{F}(t)$ as undefined for $t > t_{km_k}$.

The estimate $\hat{F}(t)$ is the direct generalization of the sample survivor function for censored data. It was first derived by Kaplan and Meier (1958), and as a consequence, is often referred to as the *Kaplan–Meier estimate*. Table 1.4 and Figure 1.2 exemplify the Kaplan–Meier estimate (1.13) for the carcinogenesis data of Section 1.1.1.
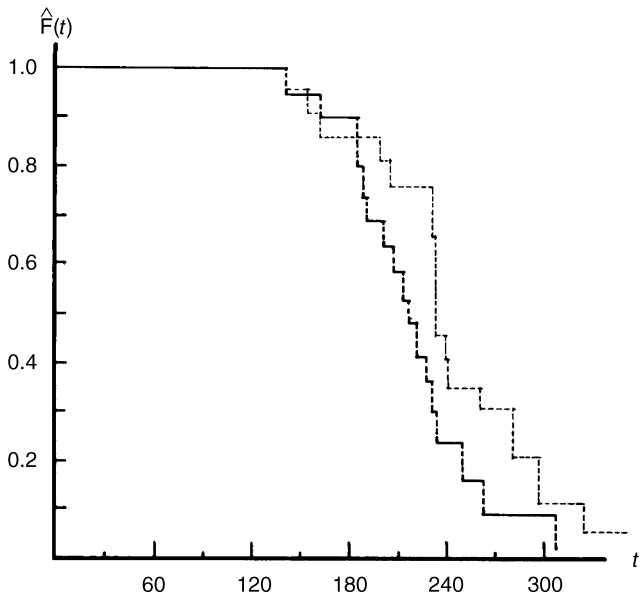
Figure 1.2 Kaplan–Meier survivor funtions estimates for carcinogenesis data: solid line, group; dashed line, group 2.

We consider now the asymptotic distribution of $\hat{F}(t)$ at a prespecified value of t. A heuristic derivation of an asymptotic variance can be obtained by regarding (1.12) as a parametric likelihood in the parameters $\lambda_1, \ldots, \lambda_k$. Standard likelihood methods, reviewed in Section 3.4, would yield an estimate $d_j(n_j - d_j)/n_j^3$ for the asymptotic variance of $\hat{\lambda}_j$ and hence for

$$\log \hat{F}(t) = \sum_{j|t_j \leq t} \log (1 - \hat{\lambda}_j),$$

an asymptotic variance estimate of

$$\widehat{\mathrm{var}}\left[\log \hat{F}(t)\right] = \sum_{j|t_j \leq t} (1 - \hat{\lambda}_j)^{-2} \, \widehat{\mathrm{var}}\,(1 - \hat{\lambda}_j)$$

$$= \sum_{j|t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}.$$

The induced expression for the asymptotic variance of $\hat{F}(t)$ is then

$$\hat{V}_F(t) = \widehat{\mathrm{var}}\left[\hat{F}(t)\right] = \hat{F}^2(t) \sum_{j|t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}. \tag{1.14}$$

Expression (1.14), known as *Greenwood's formula* (Greenwood, 1926), was first derived as the asymptotic variance of the classical life-table estimator, which is discussed below. The derivation above would be valid if the distribution of t were discrete with finitely many mass points. Proper treatment of the asymptotic properties of the Kaplan–Meier estimator in the continuous case can be based on counting process formulations and related martingale theory. We discuss these topics in Chapter 5, and asymptotics for the Kaplan–Meier and related estimates are discussed further in Section 1.7. Essentially, under reasonably mild conditions on the censoring and large n, the results justify the use of a normal approximation of the distribution of $\hat{F}(t)$ with mean $F(t)$ and variance estimate (1.14). These results hold whether T is discrete or continuous or mixed with discrete and continuous components.

An approximate 95% confidence interval for $F(t)$ is $\hat{F}(t) \pm 1.96[\widehat{\text{var}}\,\hat{F}(t)]^{1/2}$. At extreme values of t (e.g., $t \leq 188$ or $t > 246$ for the group 1 data of Table 1.3, such an approximate confidence interval may include impossible values outside the range $[0, 1]$. This problem can be avoided by applying the asymptotic normal distribution to a transformation of $F(t)$ for which the range is unrestricted. For example, the asymptotic variance of

$$\hat{v}(t) = \log\left[-\log \hat{F}(t)\right]$$

is, from Greenwood's formula and asymptotic theory (Section 3.4), estimated by

$$\hat{s}^2(t) = \widehat{\text{var}}\left[\log \hat{F}(t)\right] / \left[\log \hat{F}(t)\right]^2.$$

An asymptotic 95% confidence interval of $\hat{v}(t) \pm 1.96\hat{s}(t)$ for $v(t) = \log[-\log F(t)]$ gives a corresponding asymptotic 95% confidence interval for $F(t)$ of

$$\left[\hat{F}(t)\right]^{\exp[\pm 1.96\hat{s}(t)]},$$

which takes values in $[0, 1]$. Application of this method to the group 1 data of Table 1.1 gives an approximate 95% confidence interval for $F(t)$ at $t = 150$ of $(0.679, 0.992)$. A normal approximation to the distribution of $\hat{F}(150)$, in contrast, gives $(0.846, 1.047)$, a clearly unsatisfactory result.

It should be noted that many authors consider first the cumulative hazard function $\Lambda(t)$, which is most naturally estimated using the *Nelson–Aalen estimator*,

$$\hat{\Lambda}(t) = \sum_{t_i \leq t} d_i/n_i = \sum_{t_i \leq t} \hat{\lambda}_i, \tag{1.15}$$

which is a right-continuous step function whose increments are the empirical hazard estimates. Note that since the estimated distribution is discrete, the

Nelson–Aalen and Kaplan–Meier estimators are related in the way that one should expect [see (1.6) and (1.8)]

$$\hat{F}(t) = \mathscr{P}_0^t[1 - d\hat{\Lambda}(u)] = \prod_{t_i \leq t}(1 - \hat{\lambda}_i).$$

1.4.2   Life-Table and Related Estimates

Many other estimators of the survivor function have been considered. The oldest is that formed from the life table (see, e.g., Chiang, 1968). A *life table* is a summary of the survival data grouped into convenient intervals. In some applications (e.g., actuarial), the data are often collected in such a grouped form. In other cases, the data might be grouped to get a simpler and more easily understood presentation. Suppose, for example, that the data are grouped into intervals $I_1, \ldots, I_k$ such that

$$I_j = (b_0 + \cdots + b_{j-1}, b_0 + \cdots + b_j)$$

is of width b_j with $b_0 = 0$. The life table then presents the number of failures and censored survival times falling in each interval.

Suppose that m_j censored times and d_j failure times fall in the interval I_j, and let $n_j = \sum_{l \geq j}(d_l + m_l)$ be the number of individuals at risk at the start of the jth interval. The standard life-table estimator of the conditional probability of failure in I_j given survival to enter I_j, is $\hat{q}_j = 1$ if $n_j = 0$ and

$$\hat{q}_j = \frac{d_j}{n_j - m_j/2}$$

otherwise. The $m_j/2$ term in the denominator is used in an attempt to adjust for the fact that not all the n_j individuals are at risk for the whole of I_j. The corresponding life-table estimator of the survivor function at the end I_j is

$$\tilde{F}(b_1 = \cdots + b_j) = \prod_{l=1}^{j}(1 - \hat{q}_l). \tag{1.16}$$

Greenwood's formula (1.14), with n_j replaced by $n_j - m_j/2$, provides an estimator of the variance of $\tilde{F}$.

The life-table method is designed primarily for situations in which actual failure and censoring times are unavailable and only the d_j's and m_j's are given for the jth interval. A simple modification of the life-table method utilizes the additional information when the (continuous) failure times are known. Suppose, for example, that $t_{j1}, \ldots, t_{jr_j}$ are the observed times in I_j of which m_j are censored and d_j are failures, $r_j = d_j + m_j (j = 1, \ldots, k)$. Suppose that the hazard function $\lambda(t)$ is taken to be a

step function having constant value λ_j in the interval I_j. In this case, it can be shown that the maximum likelihood estimate of λ_j is

$$\hat{\lambda}_j = d_j/S_j,$$

where

$$S_j = \sum_{l=1}^{r_j} \left(t_{jl} - \sum_{0}^{j-1} b_i \right) + n_{j+1} b_j$$

is the total observed survival time in the interval I_j. The corresponding estimator of the survivor function is for $t \in I_j$,

$$\hat{F}(t) = \exp\left[-\hat{\lambda}_j \left(t - \sum_{l=0}^{j-1} b_l \right) - \sum_{i=1}^{j-1} \hat{\lambda}_i b_i \right]. \tag{1.17}$$

Unlike the preceding estimators, this is a continuous function of t and so relatively easier to view and to interpret shape. There is, however, an arbitrariness in the choice of intervals and in the piecewise constant model. Nonetheless, for explora- tory purposes, the estimator (1.17) can be very useful.

1.5 COMPARISON OF SURVIVAL CURVES

Often, it is of interest to determine whether two or more samples could have arisen from identical survivor functions. One approach would involve the use of the asymptotic results for $\hat{F}(t)$ mentioned above to devise a test for equality of the sur- vivor functions at some prespecified time t. Such a procedure, however, would not usually make efficient use of the data available, and attention has turned instead to test statistics that attempt to evaluate differences between survivor function estima- tors over the entire study period. The most commonly used statistics of this type can be viewed as censored data generalizations of such familiar nonparametric rank tests as the Wilcoxon test and the Savage (1956) or exponential scores test.

   In this section, a heuristic derivation of the *log-rank test* is given. This test is a censored data generalization of the Savage test and is particularly good when the ratio of hazard functions in the populations being compared is approximately con- stant. It can also be advocated on the basis of ease of presentation to nonstatistical personnel since the test statistic is particularly simple in form. It amounts to the difference between the number of failures observed in each group and a quantity that, for most purposes, can be thought of as the corresponding expected number of failures under the null hypothesis.

   Suppose that one wishes to test the hypothesis that the survivor functions $F_0(t), \ldots, F_p(t)$ are equal on the basis of samples from each of $p + 1$ populations.

Table 1.5 Frequency of Failures and Survivals at the Observed Failure Time t_j

	Sample 0	$\cdots$	Sample i	$\cdots$	Sample p	Total
Failures	d_{0j}	$\cdots$	d_{ij}	$\cdots$	d_{pj}	d_j
Survivors	$n_{0j} - d_{0j}$	$\cdots$	$n_{ij} - d_{ij}$	$\cdots$	$n_{pj} - d_{pj}$	$n_j - d_j$
At risk	n_{0j}	$\cdots$	n_{ij}	$\cdots$	n_{pj}	n_j

Let $t_1 < \cdots < t_k$ denote the failure times for the sample formed by pooling the $p + 1$ samples. Suppose that d_j failures occur at t_j and that n_j study subjects are at risk just prior to t_j $(j = 1, \ldots, k)$. Let d_{ij} and n_{ij} be the corresponding numbers in sample i $(i = 0, \ldots, p)$. The data at t_j can be summarized in the form of a $2 \times (p + 1)$ contingency table, as illustrated in Table 1.5. Conditional on the failure and censoring experience up to time t_j, the joint probability function of $d_{0j}, \ldots, d_{pj}$ is simply the product of independent binomial terms,

$$\prod_{i=0}^{p} \binom{n_{ij}}{d_{ij}} \lambda_j^{d_{ij}} (1 - \lambda_j)^{n_{ij} - d_{ij}},$$

where λ_j is the conditional failure probability (or hazard) at t_j, which under the null hypothesis is common for each of the $p + 1$ samples. The conditional distribution for $d_{0j}, \ldots, d_{pj}$ given d_j is then the multivariate hypergeometric distribution with probability function

$$\prod_{i=0}^{p} \binom{n_{ij}}{d_{ij}} \binom{n_j}{d_j}^{-1}. \tag{1.18}$$

The conditional mean and variance of d_{ij} from (1.18) are, respectively,

$$e_{ij} = n_{ij} d_j n_j^{-1}$$

and

$$(W_j)_{ii} = n_{ij}(n_j - n_{ij}) d_j (n_j - d_j) n_j^{-2} (n_j - 1)^{-1}. \tag{1.19}$$

The conditional covariance of d_{ij} and d_{lj} is

$$(W_j)_{il} = -n_{ij} n_{lj} d_j (n_j - d_j) n_j^{-2} (n_j - 1)^{-1}. \tag{1.20}$$

Thus, the statistic $w_j' = (d_{1j} - e_{1j}, \ldots, d_{pj} - e_{pj})$ has conditional mean 0 and $p \times p$ variance matrix W_j. Summing over the k failure times yields the log-rank statistic

$$w = \sum_{j=1}^{k} w_j = O - E, \tag{1.21}$$

where $O = (O_1, \ldots, O_p)'$, $E = (E_1, \ldots, E_p)'$, $O_i = \sum_{j=1}^{k} d_{ij}$, and $E_i = \sum_{j=1}^{k} e_{ij}$, $i = 1, \ldots, p$. Note that O is the vector of observed numbers of failures and E can informally be thought of as a vector of "expected" failures. This is informal only in that E is the sum of conditional expectations and its elements are random variables.

If the k contingency tables were independent, the variance of the log-rank statistic w would be $W = W_1 + \cdots + W_k$, and an approximate test of equality of the $p + 1$ survival distributions could be based on an asymptotic χ_p^2 distribution for

$$w'W^{-1}w. \tag{1.22}$$

Note that any of the $p + 1$ samples might be chosen as sample 0 and the log-rank statistic computed on the remaining p samples relabeled $1, \ldots, p$. It can be shown that the value of the statistic (1.22) is unchanged under any such relabeling.

Application of the log-rank method to a comparison of the two groups ($p = 1$) of survival data in Section 1.1.1 gives a log-rank statistic (1.21), $w = 19 - 23.763 = -4.763$, with corresponding variance estimate $W = 7.263$. The approximate χ_1^2 statistic has value $(4.763)^2 (7.263)^{-1} = 3.12$, which is just significant at the 10% level. The slight evidence of a difference that this test shows suggests improved survival for the group 2 rats. This is exhibited in the log-rank statistic, in which we see that the observed number (19) of failures in this group is less than the expected number (23.763).

The derivation of the log-rank test above is similar to that given by Mantel (1966). It is difficult, however, to formalize the distribution theory from this development since the contingency tables over failure times are clearly not independent. It can, however, be shown that the w_j's are uncorrelated and that W provides an estimate of the covariance matrix of w. The chi-squared limiting distribution of (1.22) can be shown to hold under fairly general conditions. The asymptotic results are most easily established using counting processes and martingale limit theorems, as outlined in Chapter 5.

There are two important extensions of the log-rank procedure which can be mentioned at this stage. The first is stratification, and the second concerns the inclusion of weights.

1.5.1 Stratified Log-Rank Test

A simple means of testing equality of several survival curves while allowing for heterogeneity in the populations to be compared involves stratification on auxiliary variables. An overall test statistic is obtained by summing the log-rank statistics (1.21) and corresponding variances obtained within each of the independent strata. Specifically, if the strata are indexed by h, and $w^{(h)}$ and $W^{(h)}$ are the corresponding log-rank and variance statistics based on the data in stratum h, the stratified log-rank test is based on the statistic

$$\left(\sum_{h=1}^{s} w^{(h)} \right)^T \left(\sum_{h=1}^{s} W^{(h)} \right)^{-1} \left(\sum_{h=1}^{s} w^{(h)} \right). \tag{1.23}$$

Under the null hypothesis, (1.23) typically has an asymptotic χ_p^2 distribution. It should be noted that this test will be most sensitive to differences among the $p + 1$ treatment groups that are similar across the strata. Examination of the individual log-rank tests in each of the strata can also provide some insights into possible treatment by strata interactions. This method can provide a valuable means of initial analysis and presentation for many data sets. As well, it is often a useful tool for communicating the results of a more complex analysis to nonstatistical personnel.

1.5.2 Weighted Log-Rank Test

The log-rank statistic as formulated above is most sensitive to departures from the null hypothesis in which the hazard ratios among the samples are roughly constant over time. In some instances, there may be reason to expect that any differences in the failure rates would occur early and that after the treatment has been in place for some time, treated and untreated individuals would show little difference. Conversely, there may be situations where any differences in failure rates between treatment groups might be expected to be small to begin and then larger later. Consider the weighted log-rank statistic

$$w(g) = \sum_{j=1}^{k} g_j w_j, \qquad (1.24)$$

where $g_1, \ldots, g_k$ are weights chosen in specific applications to emphasize or deemphasize in an appropriate way the differences measured by the w_j's. The g_j's may be functions of time or of j, or they may depend on the past failure and censoring experience in the study. For example, one might consider the weights $g_j^{(G)} = n_j$, which yields the Gehan–Breslow generalization of the Wilcoxon or Kruskal–Wallis statistic. Alternatively, the weights $g_j^{(P)} = \prod_{i \leq j}[1 - d_i/(n_i + 1)]$ yield the Peto and Prentice generalization of the Wilcoxon. Note that $g_j^{(P)}$ is a survivor function estimate, close to the Kaplan–Meier estimator at t_j. Both of these weighting schemes emphasize early differences in the failure rates.

Under the null hypothesis, arguments similar to those outlined above show that the weighted log-rank statistic (1.24) has mean 0 and variance estimated by $W(g) = \sum g_j^2 W_j$. This again yields a simple asymptotic χ_p^2 statistic,

$$w(g)' W(g)^{-1} w(g).$$

These statistics are considered much more comprehensively in Chapter 7.

1.6 GENERALIZATIONS TO ACCOMMODATE DELAYED ENTRY

The methods of survivor function estimation and log-rank and related tests are easily generalized to accommodate independent left truncation or delayed entry into the study sample. In fact, there are essentially no changes involved in the formula and results given. As individuals enter the study, they become at risk of failure

and so are included in the n_j or n_{ij}'s. With right censoring only, the number at risk in each sample will decrease over time as individuals fail or are censored. With left truncation, however, each new entry increases the number at risk in the appropriate group.

As a brief example, the Atomic Bomb Casualty Commision/Radiation Effects Research Foundation in Japan has, since 1950, followed a lifespan study cohort of over 100,000 persons who resided in Hiroshima or Nagasaki as of October 1, 1950. Data on this cohort are used to assess the effects of ionizing radiation exposure on mortality. The cohort includes a subsample who were residents of, but not in, either city at the time of the 1945 bombings. Key analyses from this cohort use date of bombing in the respective cities as the time origin, since mortality risk as a function of radiation exposure and time since exposure is of interest from both the public health and radiation biology perspectives. Data on time from exposure to death in this cohort are subject to left truncation since the cohort was not assembled until 1950. One can, however, estimate failure rates just as before as d_j/n_j, where d_j is the number of deaths at the jth chronological death time t_j and n_j is the number of cohort members alive and without censoring just prior to t_j. Similar changes generalize the log-rank procedures to this case. In this example it is not possible to estimate the failure rates or survival distribution for early times because no individuals who die early are included in the data. Typically, however, one can estimate the survival experience after some threshold time. Thus we can estimate that

$$F(t \mid T > a) = P(T > t \mid t > a) = F(t)/F(a)$$

for some suitably chosen a, where a might be October 1, 1950 or later in the illustration above.

In other instances, the data are subject to right truncation. In this case, the condition for study membership is that the event of interest occurs before some time of recruitment. Appendix A (data set III) gives data on transfusion-related AIDS cases in the United States. This study contains those individuals who were diagnosed with AIDS prior to 1988 and for whom the mode of infection was determined to be by blood transfusion. The distribution of the time from infection to diagnosis of AIDS (the incubation period) is of interest. In this study, individuals whose diagnosis occurs after the end of the study period are not included in the study, and the times included in the study are subject to very strong selection favoring the shorter incubation times. Right truncation is more difficult than left truncation to incorporate. Right truncation and this example are discussed further in Exercise 1.13.

1.7   COUNTING PROCESS NOTATION

Counting processes provide an alternative very compact notation for describing many of the results discussed above, and the related martingale theory provides a framework for deriving asymptotic properties. The theoretical framework and some of the asymptotic results are discussed in Chapter 5. In this section, some of the

counting process notation is introduced and the estimators, tests, and variance formulas are reexpressed in these terms. The counting process notation is widely used in the literature on failure time analysis, and a general acquaintance with it is important.

1.7.1 Kaplan–Meier and Related Estimators

As in Section 1.4, suppose that n individuals from a homogeneous population are put on study at time 0. Let F be the survivor function and Λ be the cumulative hazard function; these may be discrete, continuous, or mixed. For the ith individual, let $N_i(t)$ count the number of failures observed in the interval $(0, t]$ and let $N_i(0) = 0$. Note that N_i is right continuous and takes value 0 until a failure is observed to occur, at which time it jumps to 1. Let Y_i be the at-risk process defined such that $Y_i(t) = 1$ if the individual is without failure and uncensored just prior to time t, and $Y_i(t) = 0$ otherwise. By convention, Y_i is taken to be left continuous. Let $N.(t) = \sum_{i=1}^{n} N_i(t)$ and $Y.(t) = \sum_{i=1}^{n} Y_i(t)$, $0 < t < \infty$. Clearly, $Y.(t)$ is the number of individuals in the entire study group that are at risk at time t, and $N.(t)$ is the total number of observed failures in the interval $(0, t]$. In the notation of Section 1.4, $N.(t) = \sum_{t_i \le t} d_i$ is a right-continuous step function with a jump of d_i at t_i, $i = 1, \ldots, k$ and $Y.(t), 0 < t < \infty$ is a left-continuous step function that specifies the number of individuals who are uncensored and surviving at time t. Note that $Y.(t_i) = n_i, i = 1, \ldots, k$.

The Nelson–Aalen estimator of the cumulative hazard (1.15) can be written as the stochastic integral

$$\hat{\Lambda}(t) = \int_0^t \frac{J(u)}{Y.(u)} \, dN.(u), \tag{1.25}$$

where $J(u) = I[Y.(u) > 0]$ with the convention that $0/0$ is interpreted as 0. Note that $J(u)$ is used as a device to account for the possibility that at time u^-, there may be no items at risk. The Kaplan–Meier estimator of the survivor function is

$$\hat{F}(t) = \prod_{u \le t} [1 - d\hat{\Lambda}(u)] = \mathcal{P}_0^t [1 - \frac{J(u)}{Y.(u)} \, dN.(u)]. \tag{1.26}$$

We had previously considered the Kaplan–Meier and Nelson–Aalen estimators to be undefined for t values greater than the maximum observed time if that time corresponded to a censoring. The convention being used in (1.25) and (1.26), however, takes the estimates as defined at all t, but constant following the maximum observed time. The former convention is more appropriate in most contexts, but the latter is convenient for some theoretical arguments.

A variance estimator for the Nelson–Aalen estimator (1.15) or (1.25) is

$$\hat{V}(t) = \int_0^t \frac{J(u)}{[Y.(u)]^2} \left[1 - \frac{\Delta N.(u)}{Y.(u)} \right] dN.(u)$$

$$= \sum_{t_j \le t} \frac{d_j(n_j - d_j)}{n_j^3}, \tag{1.27}$$

where $\Delta N.(u) = N(u) - N(u^-)$. Large-sample properties of the Nelson–Aalen estimator can be shown to hold under relatively mild conditions, as outlined in Section 5.5. If for given t, $Y.(u) \to \infty$ for all $u \in (0, t]$ as $n \to \infty$, it is shown that $\hat{\Lambda}(t) \xrightarrow{\mathscr{P}} \Lambda(t)$ and

$$[\hat{\Lambda}(t) - \Lambda(t)]/\hat{V}(t)^{0.5} \xrightarrow{\mathscr{D}} N(0, 1),$$

where $\xrightarrow{\mathscr{P}}$ and $\xrightarrow{\mathscr{D}}$ indicate convergence in probability and convergence in distribution, respectively.

Greenwood's variance formula (1.14) can be written

$$\widehat{\text{var}}\,[\hat{F}(t)] = [\hat{F}(t)]^2 \int_0^t \frac{1}{Y.(s)[Y.(s) - \Delta N.(s)]}\, dN.(s). \tag{1.28}$$

1.7.2   Log-Rank and Related Tests

Consider the experimental situation described in Section 1.5, where n_{i0} items are placed on test in the ith group at time 0, and let $N_{il}(t), t > 0$ be the counting process for the number of failures observed in $(0, t]$ for the lth individual in the ith group, $l = 1, \ldots, n_{i0}; i = 0, \ldots, p$. The corresponding at risk processes are $Y_{i\ell}(t)$, and again we assume independent censoring. Let $N_{i.}(t) = \sum_{l=1}^{n_{i0}} N_{il}(t)$ record the number of observed failures in the ith group and $Y_{i.}(t) = \sum Y_{i\ell}(t)$ specify the number at risk at time t. The ith component of the log-rank statistic (1.21) can now be written as

$$w_i = \int_0^\infty dN_{i.}(u) - \frac{Y_{i.}(u)}{Y..(u)}\, dN..(u), \tag{1.29}$$

where $N..(t) = \sum_{i=0}^p N_{i.}(t)$. With some algebra, it can be verified that

$$w_i = \sum_{\ell=0}^p \int_0^\infty \left[\delta_{i\ell} - \frac{Y_{i.}(u)}{Y..(u)} \right] dN_{\ell.}(u), \qquad i = 1, \ldots, p, \tag{1.30}$$

where $\delta_{i\ell} = \mathbf{1}(i = \ell)$. The variance and covariance formulas can also be expressed in counting process notation, as discussed further in Section 5.6.

BIBLIOGRAPHIC NOTES

Some useful references to life-table estimation are those by Berkson and Gage (1952), Cutler and Ederer (1958), Chiang (1960,1968), and Gehan (1969). The Kaplan–Meier or product limit estimator appears first to have been proposed as a limit of the life-table estimator by Böhmer (1912). It was not followed up, however,

and was reintroduced in the important paper by Kaplan and Meier (1958), who showed that the estimate was a nonparametric MLE through an argument similar to that given in Section 1.4.1. Efron (1967) showed that the estimate satisfied a certain self-consistency property and discussed asymptotic properties. Breslow and Crowley (1974) first derived the asymptotic results for the Kaplan–Meier estimator under a random censorship model. More recent references for asymptotic results utilizing counting processes and martingales are reviewed in the notes to Chapter 5. The estimates based on the life table will tend to be slightly biased due to the grouping, and this will also typically be true for the piecewise continuous estimate (1.17). The nonparametric maximum likelihood approach and the self-consistency ideas of Efron (1967) were extended by Turnbull (1974, 1976) to include left and right truncations and interval censoring. Some of this work is reviewed in Section 3.9.1, and additional references and discussion on interval censoring are given in the bibliographic notes for Chapter 3.

The Nelson–Aalen estimate was first proposed by Nelson (1969,1972) as the basis for simple graphical checks for hazard shape in industrial life testing. Its large-sample properties were studied by Breslow and Crowley (1974) and by Aalen (1976). Altshuler (1970) also derived the Nelson–Aalen estimator and gave a related estimate of the survivor function. The product integral was introduced in the statistical literature by Cox (1972) as a compact description of the relationship between the hazard and the survivor function. A useful summary can be found in Dollard and Friedman (1979). See Gill and Johansen (1990) for a comprehensive account of product integration in relation to failure time data.

The adequacy of the asymptotic approximations to the Kaplan–Meier and Nelson–Aalen estimators has received some attention in the literature. It is evident that transformations to improve the asymptotic approximation in the tail is a useful technique and this has been explored by Klein (1991), who suggests a logistic rather than a $\log(-\log)$ transformation. Thomas and Grunkemeier (1975) developed a generalized likelihood ratio test of an hypothesized value for $F(c)$ at a given c (see Exercise 1.8) and argued that a χ^2_1 asymptotic distribution should apply and gave some simulations. This approach has received some attention in the literature, and asymptotic results have been derived by Li (1995a,b), Li et al. (1996), and Murphy (1995) for nonparametric likelihood ratio tests in various contexts. This approach is essentially that of empirical likelihood, and the recent book by Owen (2001) gives references and a good summary of asymptotic results.

We have given a derivation of the log-rank test in Section 1.5 that is essentially the same as that given originally by Mantel (1966). The test has been widely used in the literature, and both it and the weighted log-rank test arise in various contexts. The name log-rank was coined by Peto and Peto (1972) and the motivation of the term is not entirely clear to all — some say to apply it one first logs the data and then ranks them. The weighted log-rank test has been considered by many authors. Tarone and Ware (1977) first considered the general class. Harrington and Fleming (1982) considered a family of weight functions indexed by a parameter ρ that included the Wilcoxon and log-rank tests as special cases. They derive the asymptotic null distribution of the maximum weighted log-rank statistic in the class.

Fleming and Harrington (1991, Chap. 7) give an extensive discussion of log-rank and weighted log-rank procedures and have collected numerous references. References for counting processes and associated asymptotics are collected in Chapter 5.

EXERCISES AND COMPLEMENTS

1.1 Consider the mouse carcinogenesis data of Appendix A (data set V). Compute the product limit (Kaplan–Meier) estimates (1.10) of the survivor function for the endpoint, reticulum cell sarcoma, for the control and germ-free groups by:

(a) Ignoring failures from thymic lymphoma and other causes (i.e., eliminate mice dying by these causes before carrying out calculations).

(b) Regarding failure times from lymphoma or other causes as right censored.

Comment on the relative merits of parts (a) and (b). (*Hint*: Try to understand what is being estimated in both cases.) On the basis of the survivor function plots, does the germ-free environment appear to reduce the risk of reticulum cell sarcoma?

1.2 Plot on a single graph the logarithms of the estimates obtained from the life table (1.16), product limit (1.10), and the continuous (1.17) estimates of the survivor function for the thymic lymphoma data in the germ-free group. Regard failures from reticulum cell sarcoma and other causes as censored. Use grouping intervals of width 50 days for (1.16) and (1.17).

1.3 Show that the Kaplan–Meier estimate reduces to $\hat{F}(t) = $ (no. observations $> t)/n$ when there is no censoring. Show that Greenwood's formula (1.14) reduces in this case to the usual estimate of the variance of a binomial proportion. That is,

$$\widehat{\text{var}}\,[\hat{F}(t)] = n^{-1}\hat{F}(t)[1 - \hat{F}(t)].$$

1.4 Let T be a discrete failure time variable taking values on the points $x_1, x_2, \ldots$ with survivor function $F(t)$. Show that the area under the survivor function, $\int_0^\infty F(t)\,dt = E(T)$. (*Note:* A simple geometric proof of this is obtained by partitioning a plot of the survivor function into rectangles with bases along the vertical axis.)

1.5 Let T be a discrete, continuous, or mixed random variable with survivor function $F(t)$. Show that $E(T) = \int_0^\infty F(t)\,dt$.

1.6 An electronic system is at continuous risk of failure with a constant hazard of λ events per hour. In addition, power surges occur each hour (i.e., at times 1,2,...), and at each power surge there is a 10% chance that the system will

fail immediately. Obtain expressions for the survivor and cumulative hazard functions. Find the mean of T.

1.7 Let the survival time $T > 0$ be an integer-valued random variable with finite mean r_0 and let

$$r_i = E(T - i \mid T > i)$$

be the expected residual life at time i, $i = 1, 2, \ldots$. Show that the survivor function for integer t is

$$F(t) = P(T > t) = \prod_{i=1}^{t} \frac{r_{i-1} - 1}{r_i}.$$

Thus, in the discrete case also, the residual mean lifetime specifies the distribution of T. (*Note*: The geometric argument in Exercise 1.4 can be used to show that $r_j = [1 - F(1) - \cdots - F(j-1)]/F(j)$, for $j = 1, 2, \ldots$.)

1.8 As in Section 1.3, let $t_1 < t_2 < \cdots < t_k$ represent the observed failure times in a sample of size n_0 from a homogeneous population with survivor function $F(t)$. Suppose that d_j items fail at t_j and that n_j items are at risk at t_j^-.

(a) Let b be a prespecified time $(b > t_1)$ and c be a constant $(0 \le c \le 1)$. Show that subject to the constraint $F(b) = c$, the nonparametric maximum likelihood estimate of $F(t)$ is

$$\tilde{F}(t) = \prod_{j \mid t_j \le t} (1 - \tilde{\lambda}_j),$$

where $t_0 = \tilde{\lambda}_0 = 0$ and $\tilde{\lambda}_j = d_j/(n_j + a)$ if $t_j \le b$ and d_j/n_j if $t_j > b$, $j = 1, \ldots, k$. The value a is chosen to satisfy $\tilde{F}(b) = c$. Note that if $b \le t_1$, the constrained estimate is not unique for $t < b$. An arbitrary convention would assign a hazard $1 - c$ at $t = \epsilon$ for some small positive $\epsilon < b$.

(b) Show that the log-likelihood ratio statistic for the hypothesis $F(b) = c$ can be written

$$R = \sum_{i \mid t_i \le b} \left[(n_i - d_i) \log \left(1 + \frac{a}{n_i - d_i} \right) - n_i \log \left(1 + \frac{a}{n_i} \right) \right].$$

(c) Thomas and Grunkemeier (1975) show that the usual asymptotic properties apply and that $-2R$ is asymptotically χ_1^2 under the hypothesis. Use this result to establish a 95% confidence interval for $F(b)$. Compare these results with those obtained in Section 1.3 for the carcinogenesis data (Table 1.1) with $b = 150$.

1.9 Suppose that censored samples are available on two populations with survivor functions $F_1(t)$ and $F_2(t)$. Consider the hypothesis $F_2(b) = F_1(b)$ at some prespecified time b. Extend the results in Exercise 1.8 to obtain the nonparametric

likelihood ratio statistic for this hypothesis. Apply this approach to test for equality of the survivor functions at $b = 250$ for the carcinogenicity data (Table 1.1).

1.10 Show that the mean vector and variance matrix for $(d_{1j}, \ldots, d_{pj})$ in the distribution (1.18) are as asserted.

1.11 Consider again the mouse carcinogenesis data (data set V, Appendix A). Use the log-rank test (1.16) to test the hypothesis that germ-free isolation does not affect overall mortality.

1.12 Suppose that $T_1, \ldots, T_n$ are independent exponential variates with respective failure rates $\lambda_1, \ldots, \lambda_n$. Let $\gamma_1, \ldots, \gamma_m$ be the distinct elements of $\lambda_1, \ldots, \lambda_n$. Let $S = \sum_{i=1}^{n} T_i$.

(a) Show that the survivor function of S may be written as

$$F_S(t) = P(S > t) = \sum_{j=1}^{m} p_j(t) e^{-\gamma_j t},$$

where the p_j's are polynomials in t.

(b) Let $\lambda_S(t)$ be the hazard function of S and show that $\lambda_S(t) \leq \lambda_{\min}$ for all t and that $\lim_{t \to \infty} \lambda_S(t) = \lambda_{\min}$, where $\lambda_{\min} = \min(\gamma_1, \ldots, \gamma_m)$.

1.13 Consider the transfusion-related AIDS data in data set III, Appendix A. As discussed in Section 1.6, these data are subject to right truncation in that a condition for study membership is that diagnosis of AIDS takes place prior to the end of the study period. Let T represent the number of months from transfusion to AIDS diagnosis, and $F(t) = P(T > t)$ be the corresponding (discrete) survivor function. Let t_i be the month of diagnosis, and let a_i be the total months elapsed to the end of the study period for the ith subject, $i = 1, \ldots, n$.

(a) Under what conditions would the likelihood function be of the form $\prod_{i=1}^{n} \{[F(t_i) - F(t_i^-)]/[1 - F(a_i)]\}$.

(b) Explain why F can only be estimated up to a constant of proportionality.

(c) Let $a = \max(a_1, \ldots, a_n)$, and find the maximum likelihood estimate of the conditional survivor function $G(t) = F(t)/[1 - F(a)] = P(T > t \mid T \leq a)$.

(d) What additional information would you need to estimate the median time from transfusion infection to diagnosis with AIDS? (Lagakos et al., 1988; Kalbfleisch and Lawless, 1989)

1.14 Consider the data of Table 1.2. Apply the log-rank test to compare the two treatment groups in the trial. Consider dividing the data into three strata consisting of patients in the age groups ≤ 15, 16–25, and ≥ 26, respectively. Apply a log-rank test separately in each stratum and the stratified log-rank test. Discuss the results.