The Statistical Analysis of Failure Time Data

Second Edition

JOHN D. KALBFLEISCH

ROSS L. PRENTICE

A JOHN WILEY & SONS, INC., PUBLICATION

This page intentionally left blank

The Statistical Analysis of Failure Time Data

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

The Statistical Analysis of Failure Time Data

Second Edition

JOHN D. KALBFLEISCH

ROSS L. PRENTICE

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data Is Available

ISBN 0-471-36357-X

Printed in the United States of America.

10 9 8

To Sharon and Didi

This page intentionally left blank

Contents

Preface xi

1. Introduction 1

1.1 Failure Time Data, 1
1.2 Failure Time Distributions, 6
1.3 Time Origins, Censoring, and Truncation, 12
1.4 Estimation of the Survivor Function, 14
1.5 Comparison of Survival Curves, 20
1.6 Generalizations to Accommodate Delayed Entry, 23
1.7 Counting Process Notation, 24
Bibliographic Notes, 26
Exercises and Complements, 28

2. Failure Time Models 31

2.1 Introduction, 31
2.2 Some Continuous Parametric Failure Time Models, 31
2.3 Regression Models, 40
2.4 Discrete Failure Time Models, 46
Bibliographic Notes, 49
Exercises and Complements, 49

3. Inference in Parametric Models and Related Topics 52

3.1 Introduction, 52
3.2 Censoring Mechanisms, 52
3.3 Censored Samples from an Exponential Distribution, 54
3.4 Large-Sample Likelihood Theory, 57
3.5 Exponential Regression, 65

3.6 Estimation in Log-Linear Regression Models, 68
3.7 Illustrations in More Complex Data Sets, 70
3.8 Discrimination Among Parametric Models, 74
3.9 Inference with Interval Censoring, 78
3.10 Discussion, 83
Bibliographic Notes, 85
Exercises and Complements, 87

4. Relative Risk (Cox) Regression Models **95**

4.1 Introduction, 95
4.2 Estimation of β, 99
4.3 Estimation of the Baseline Hazard or Survivor Function, 114
4.4 Inclusion of Strata, 118
4.5 Illustrations, 119
4.6 Counting Process Formulas, 128
4.7 Related Topics on the Cox Model, 130
4.8 Sampling from Discrete Models, 135
Bibliographic Notes, 142
Exercises and Complements, 144

5. Counting Processes and Asymptotic Theory **148**

5.1 Introduction, 148
5.2 Counting Processes and Intensity Functions, 149
5.3 Martingales, 157
5.4 Vector-Valued Martingales, 164
5.5 Martingale Central Limit Theorem, 165
5.6 Asymptotics Associated with Chapter 1, 167
5.7 Asymptotic Results for the Cox Model, 172
5.8 Asymptotic Results for Parametric Models, 178
5.9 Efficiency of the Cox Model Estimator, 181
5.10 Partial Likelihood Filtration, 188
Bibliographic Notes, 189
Exercises and Complements, 190

6. Likelihood Construction and Further Results **193**

6.1 Introduction, 193
6.2 Likelihood Construction in Parametric Models, 193
6.3 Time-Dependent Covariates and Further Remarks on Likelihood Construction, 196

6.4 Time Dependence in the Relative Risk Model, 200
6.5 Nonnested Conditioning Events, 208
6.6 Residuals and Model Checking for the Cox Model, 210
Bibliographic Notes, 212
Exercises and Complements, 214

7. Rank Regression and the Accelerated Failure Time Model **218**

7.1 Introduction, 218
7.2 Linear Rank Tests, 219
7.3 Development and Properties of Linear Rank Tests, 224
7.4 Estimation in the Accelerated Failure Time Model, 235
7.5 Some Related Regression Models, 241
Bibliographic Notes, 242
Exercises and Complements, 244

8. Competing Risks and Multistate Models **247**

8.1 Introduction, 247
8.2 Competing Risks, 248
8.3 Life-History Processes, 266
Bibliographic Notes, 273
Exercises and Complements, 275

9. Modeling and Analysis of Recurrent Event Data **278**

9.1 Introduction, 278
9.2 Intensity Processes for Recurrent Events, 280
9.3 Overall Intensity Process Modeling and Estimation, 282
9.4 Mean Process Modeling and Estimation, 286
9.5 Conditioning on Aspects of the Counting Process History, 297
Bibliographic Notes, 299
Exercises and Complements, 300

10. Analysis of Correlated Failure Time Data **302**

10.1 Introduction, 302
10.2 Regression Models for Correlated Failure Time Data, 303
10.3 Representation and Estimation of the Bivariate Survivor Function, 308
10.4 Pairwise Dependency Estimation, 311
10.5 Illustration: Australian Twin Data, 313

10.6 Approaches to Nonparametric Estimation of the Bivariate Survivor Function, 315
10.7 Survivor Function Estimation in Higher Dimensions, 322
Bibliographic Notes, 323
Exercises and Complements, 324

11. Additional Failure Time Data Topics **328**

11.1 Introduction, 328
11.2 Stratified Bivariate Failure Time Analysis, 329
11.3 Fixed Study Period Survival Studies, 334
11.4 Cohort Sampling and Case–Control Studies, 337
11.5 Missing Covariate Data, 343
11.6 Mismeasured Covariate Data, 346
11.7 Sequential Testing with Failure Time Endpoints, 348
11.8 Bayesian Analysis of the Proportional Hazards Model, 352
11.9 Some Analyses of a Particular Data Set, 361
Bibliographic Notes, 369
Exercises and Complements, 371

Glossary of Notation **375**

Appendix A: Some Sets of Data **378**

Appendix B: Supporting Technical Material **396**

Bibliography **404**

Author Index **429**

Subject Index **435**

Preface

As in the first edition of this book, the purpose of this revision is the collection and unified presentation of statistical models and methods for the analysis of failure time data. The motivation for this effort continues to derive primarily from biomedical contexts and, to a lesser extent, industrial life-testing purposes.

A voluminous literature on failure time analysis and the closely related event history analysis has developed in the more than 20 years since the publication in 1980 of the first edition of this book. The theoretical underpinnings of the methods described previously have been strengthened in the interim, and many important generalizations and related developments have taken place. Counting process methods and related martingale convergence results have led to precise and general asymptotic results for tests and estimators under key classes of failure time models and important censoring and truncation mechanisms. These developments have also contributed to the formulation of broader classes of models and methods.

An important challenge in developing this revision was to preserve the feature of a fairly elementary and classical likelihood-based presentation of failure time models and methods while integrating the counting process notation and related theory. This we have done by using classical notation and descriptions throughout the first four chapters of the revision while introducing the reader to key estimating functions and estimators in notation involving counting processes and stochastic integration. These chapters deal with survivor function estimation and comparison of survival curves (Chapter 1); statistical models for failure time distributions, including parametric and semiparametric regression models (Chapter 2); testing and estimation in parametric regression models under right censoring and other selected censoring schemes (Chapter 3); and testing and estimation under the semiparametric Cox regression model (Chapter 4). These chapters, along with parts of Chapters 6 to 8, can form the basis for an introductory graduate-level biostatistics or statistics course. We have tried to keep a solid contact with the first edition in many places and, for example, have retained illustrations from that edition where they still seemed to make the relevant points well.

A new Chapter 5 provides a more systematic introduction to counting processes and martingale convergence results and describes how they can be applied to yield

asymptotic results for many of the statistical methods discussed in the first four chapters. The treatment is somewhat less formal than in some more specialized books, but presents the reader with a development and summary of the main ideas and a good basis for further investigation and study.

The remainder of the book uses the notation from counting processes and stochastic integrals where it is helpful, but continues to emphasize the likelihood basis for testing and estimation procedures. Like Chapter 5 in the first edition, Chapter 6 is devoted to general concepts of likelihood and partial likelihood construction, especially in relation to time-dependent and evolving covariate histories. We also provide an example in which martingale methods do not allow the development of asymptotic results because the conditioning events are not nested in time. Like our previous Chapter 6, Chapter 7 is devoted to the semiparametric log-linear or accelerated failure time model. Over the past two decades much effort has been devoted to regression estimation under this model, to the point where it can provide a practical alternative to the Cox model. Like our previous Chapter 7, Chapters 8 through 10 are devoted to aspects of multivariate failure time data analysis, including competing risk and multistate failure time modeling and estimation (Chapter 8), recurrent event modeling and estimation (Chapter 9), and correlated failure time methods (Chapter 10). Aside from a part of Chapter 8, most of the material in these chapters reflects developments since the first edition was published. Martingale convergence results are applicable to some of the estimating functions considered in these chapters, but others rely on empirical process methods. The latter methods can largely subsume the martingale methods, but we have not attempted comprehensive coverage here. Chapter 11 is devoted to more specialized topics. We have retained some of the material from our original Chapter 8 while providing a description of methods for such topics as risk set sampling, missing covariate data, mismeasured covariate data, sequential testing and estimation, and Bayesian methods, mostly in the context of the Cox model. The revision as a whole can serve as the textbook for a more advanced graduate course in biostatistics or statistics.

With the vast literature that has developed on failure time analysis, we have had to be selective in both the scope and depth of our coverage. We have chosen not to provide in-depth coverage of probability theory that is relevant to the asymptotic methods and results discussed, nor, except for some general comments in Appendix B, have we attempted to include a description of how available statistical software packages can or cannot be used to implement the various methods. We have chosen to emphasize some statistical models and approaches that seem to us to be of particular importance, to stress the ideas behind their development and application, and to provide some worked examples that illustrate their use.

To augment the usefulness of this revision as a graduate text, we have included a set of exercises at the end of each chapter. A number of these problems introduce the reader to additional pertinent failure time literature. As before, we have used references sparingly, especially in the early chapters, and bibliographic notes are provided at the close of each chapter. For historical reasons we have retained most of bibliographic notes from the original version, but we have augmented them with important recent references for each failure time topic.

There are a number of books on failure time methods that nicely complement this work and provide more comprehensive coverage of specific topics. For example, Lawless (1982) provides extensive coverage of parametric failure time models and estimation procedures; Cox and Oakes (1984) provide a concise and readable account of a range of failure time data topics; Fleming and Harrington (1991) provide a rigorous presentation of Cox regression methods and selected other failure time topics with considerable attention to model checking procedures; Andersen et al. (1993) give a comprehensive compendium of failure time and event history analysis methods with emphasis on counting processes. Andersen et al. (1993) provide additional material on a number of the topics discussed here. Books by Collett (1994) and Klein and Moeschberger (1997) provide relatively less technical accounts of the methods for key failure time topics. Collett includes a presentation of computer software options. Therneau and Grambsch (2000) discuss the implementation of failure time methods using SAS and S-Plus and provide a number of detailed illustrations with particular attention to model building and testing. Hougaard (2000) presents the first book dedicated to multivariate failure time methods. His book nicely complements our Chapters 8 through 10, with a greater emphasis on random effects or frailty models.

We would like to express our thanks to colleagues and to former and current students who have helped to shape our understanding of failure time analysis issues and methods. Their ideas and efforts have helped to inform this presentation.

JOHN D. KALBFLEISCH
ROSS L. PRENTICE

February 2002

This page intentionally left blank

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 ≥ 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*	1	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

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

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

$$\begin{aligned} \lambda(t) &= f(t)/F(t) \\ &= -d \log F(t)/dt. \end{aligned}$$

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

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

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,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

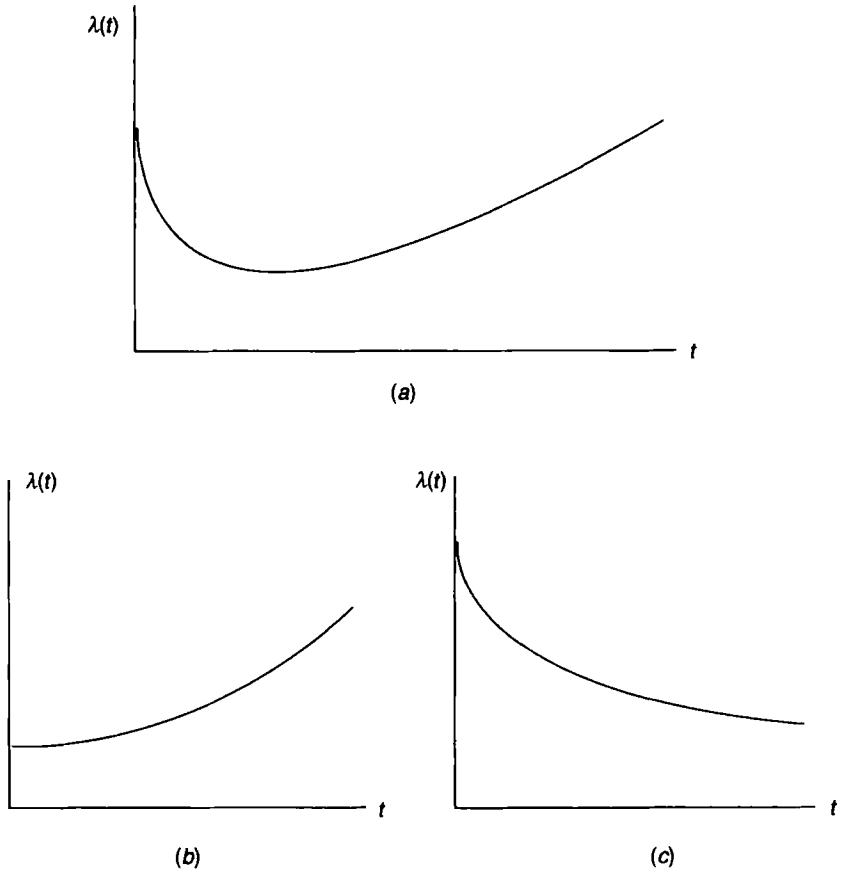

Figure 1.1 Examples of hazard functions: (*a*) hazard for human mortality; (*b*) positive aging; (*c*) negative aging.

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,

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

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*,

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

is a simple estimate of the distribution function $\bar{F}(x) = P(X \leq x)$ 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