Statistics *O.E. Barndorff-Nielson* (1989)
n *D.R. Cox and E.J. Snell* (1989)
e Data *N.G. Becker* (1989)
-Over Trials *B. Jones and* (1989)
n *J.S. Maritz and T. Lwin* (1989)
ated Distributions *K.-T. Fang* g (1989)
l edition *P. McCullagh and* 989)
l. John (1987)
nometrics *C.F. Manski* (1988)
sion *A.J. Miller* (1990)
Crowder and D.J. Hand (1990)
Probabilities *P. Walley* (1990)
astie and R.J. Tibshirani (1990)
tes in Quality Control
d X. Wu (1991)
2nd edition *B.S. Everitt* (1992)
ation: A State-space Approach 93)
tatistics *M.K. Murray* 993)
on *M.H.A. Davis* (1993)
and Probalistic Aspects
elson,
ndall (1993)
Statistics *K.-T. Fang* 93)
tics *D.R. Cox*
'son (1994)
C.D. Daykin, T. Pentikäinen 993)
s in Medicine *J. Aitchison* 994)
Geisser (1993)
artar and M. Lock (1994)
otstrap *B. Efron* 993)
neralized Linear Models
man (1994)
x and M. Cox (1994)

ilable from the publisher)

Analysis of Survival Data

D. R. COX

Nuffield College, Oxford, UK

D. OAKES

Department of Statistics,
University of Rochester, USA
formerly TUC Centenary Institute of Occupational Health
London School of Hygiene and Tropical Medicine, UK

CHAPMAN & HALL

London · Weinheim · New York · Tokyo · Melbourne · Madras

Published by Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, UK

Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, UK

Chapman & Hall GmbH, Pappelallee 3, 69469 Weinheim, Germany

Chapman & Hall USA., One Penn Plaza, 41st Floor, New York, NY10119, USA

Chapman & Hall Japan, ITP - Japan, Kyowa Building, 3F, 2-2-1 Hirakawacho, Chiyoda-ku, Tokyo 102, Japan

Chapman & Hall Australia, Thomas Nelson Australia, 102 Dodds Street, South Melbourne, Victoria 3205, Australia

Chapman & Hall India, R. Seshadri, 32 Second Main Road, CIT East, Madras 600 035, India

First edition 1984
Reprinted 1985, 1988, 1990, 1992, 1994 (twice), 1996

Printed in Great Britain at the University Press, Cambridge

ISBN 0 412 24490 X

A Catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Cox, D.R. (David Roxbee)
Analysis of survival data.
(Monographs on statistics and applied probability)
Bibliography:p.
Includes indexes.
1. Failure time data analysis.I. Oakes, D.
II. Title III. Series
QA276.C665 1984 519.5 83-20882
ISBN 0-412-22490-X

Contents

Preface

The statistical analysis of the duration of life has a long history. The recent surge of interest in the topic, with its emphasis on the examination of the effect of explanatory variables, stems mainly from medical statistics but also to some extent from industrial life-testing. In fact the applications range much more widely, certainly from physics to econometrics. The essential element is the presence of a nonnegative response with appreciable dispersion and often with right censoring.

The object of the present book is to give a concise account of the analysis of survival data. We have written both for the applied statistician encountering problems of this type and also for a wider statistical audience wanting an introduction to the field.

To keep the book reasonably short we have omitted both some of the very special methods associated with the fitting of particular distributions and also the mathematically interesting topic of the application of martingale theory and weak convergence to the rigorous development of asymptotic theory. We have also firmly resisted the temptation to extend the discussion to the statistical analysis of point processes, i.e. systems in which several point events may be experienced by each individual.

We thank warmly Ms P. J. Solomon for comments on a preliminary version.

London, March 1983

D. R. Cox
D. Oakes

CHAPTER 1

The scope of survival analysis

1.1 Introduction

In survival analysis, interest centres on a group or groups of individuals for each of whom (or which) there is defined a point event, often called failure, occurring after a length of time called the failure time. Failure can occur at most once on any individual.

Examples of failure times include the lifetimes of machine components in industrial reliability, the durations of strikes or periods of unemployment in economics, the times taken by subjects to complete specified tasks in psychological experimentation, the lengths of tracks on a photographic plate in particle physics and the survival times of patients in a clinical trial.

To determine failure time precisely, there are three requirements: a time origin must be unambiguously defined, a scale for measuring the passage of time must be agreed and finally the meaning of failure must be entirely clear. We discuss these requirements in a little more detail in Section 1.2.

Sometimes we are concerned solely with the distribution of failure times in a single group. More often, we may wish to compare the failure times in two or more groups to see, for example, whether the failure times of individuals are systematically longer in the second group than in the first. Alternatively, values may be available for each individual of explanatory variables, thought to be related to survival. The lifetime of a machine component may be influenced by the stress exerted on it, or by the working temperature. White blood count is known to influence prognosis in leukaemia. In clinical practice, it is quite common for information on 100 or more variables to be routinely collected on each patient, giving the statistician the unenviable task of summarizing the joint effect of these variables on survival.

Survival analysis is properly thought of as a univariate rather than a multivariate technique because there is only a single response

variable, failure time, even though there may be many explanatory variables. Some special problems involving a multivariate response are, however, discussed in Chapter 10.

1.2 The definition of failure times

We now comment briefly on the requirements for measuring failure time.

The time origin should be precisely defined for each individual. It is also desirable that, subject to any known differences on explanatory variables, all individuals should be as comparable as possible at their time origin. In a randomized clinical trial, the date of randomization satisfies both criteria, and would be the normal choice. While it might be more biologically meaningful to measure time from the first instant at which the patient's symptoms met certain criteria of severity, the difficulty of determining and the possibility of bias in such values would normally exclude their use as time origin. Such information might, however, be useful as an explanatory variable.

The time origin need not be and usually is not at the same calendar time for each individual. Most clinical trials have staggered entry, so that patients enter over a substantial time period. Each patient's failure time is usually measured from his own date of entry. Fig. 1.1 illustrates the calculation.

The evaluation of screening programmes for the detection of breast cancer provides an instructive example of the difficulties in the choice of origin. The aim of screening, of course, is to detect the disease at an earlier stage in its development than would otherwise be possible. Even in the absence of effective treatment, patients with disease detected at screening would be expected to survive longer after diagnosis than patients whose disease is detected without the aid of screening. This bias seriously complicates any comparison of the failure times of the two groups. Perhaps the only satisfactory way to evaluate the effect of screening in reducing mortality is to compare the total mortality rate in a population offered screening with that in a population where no screening programme is available.

The time origin need not always be at the point at which an individual enters the study, but if it is not, special methods are needed. For example, in epidemiological studies of the effects on mortality of occupational exposure to agents such as asbestos, the natural measure of time is age, since this is such a strong determinant of

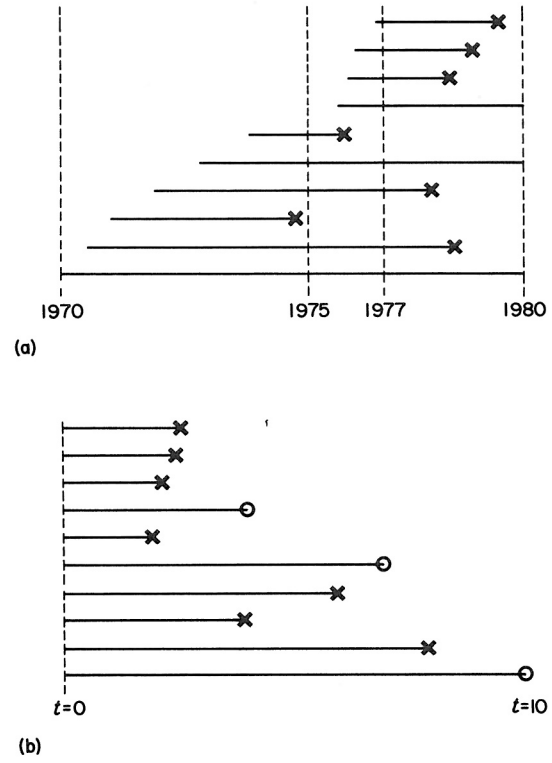

Fig. 1.1. Experience of ten individuals with staggered entry and follow-up until 1980: ×, death; ○, censoring. (a) Real time; (b) time, t, from entry into study.

mortality. However, observation on each individual commences only when he starts work in a job which involves exposure to asbestos. Likewise, in industrial reliability studies, some components may already have been in use for some period before observation begins. We refer to such data as 'left-truncated' and the appropriate methods are discussed in Chapter 11.

Often the 'scale' for measuring time is clock time (real time), although other possibilities certainly arise, such as the use of operating time of a system, mileage of a car, or some measure of cumulative load encountered. Indeed, in many industrial reliability applications, time is most appropriately measured by cumulative usage, in some sense. Or failures may consist of flaws in textile yarn,

when failure 'time' would be the length measured up to the first flaw. There are interesting applications in geometrical probability, where the failure time denotes the length of a line segment contained in a convex body. About the only universal requirement for failure times is that they are nonnegative.

One reason for the choice of a timescale is direct meaningfulness for the individual concerned, justifying the use of real time in investigating survival in a medical context. Another consideration is that two individuals treated identically should, other things being equal, be in a similar state after the lapse of equal 'times'; this is the basis for the use of cumulative load encountered in an engineering context. If two or more different ways of measuring time are available, it may be possible, having selected the most appropriate timescale, to use the other 'times' as explanatory variables.

Finally, the meaning of the point event of failure must be defined precisely. In medical work, failure could mean death, death from a specific cause (e.g. lung cancer), the first recurrence of a disease after treatment, or the incidence of a new disease. In some applications there is little or no arbitrariness in the definition of failure. In others, for example in some industrial contexts, failure is defined as the first instance at which performance, measured in some quantitative way, falls below an acceptable level, defined perhaps by a specification. Then there will be some arbitrariness in the definition of failure and it will be for consideration whether to concentrate on failure time or whether to analyse the whole performance measure as a function of time.

1.3 Censoring

A special source of difficulty in the analysis of survival data is the possibility that some individuals may not be observed for the full time to failure. At the close of a life-testing experiment in industrial reliability, not all components may have failed. Some patients (many, it is to be hoped) will survive to the end of a clinical trial. A patient who has died from heart disease cannot go on to die from lung cancer.

An individual who is observed, failure-free, for 10 days and then withdrawn from study has a failure time which must exceed 10 days. Such incomplete observation of the failure time is called censoring. Note that, like failure, censoring is a point event and that the period of observation for censored individuals must be recorded.

We suppose that, in the absence of censoring, the ith individual in a sample of n has failure time T_i, a random variable. We suppose also that there is a period of observation c_i such that observation on that individual ceases at c_i if failure has not occurred by then. Then the observations consist of $X_i = \min(T_i, c_i)$, together with the indicator variable $V_i = 1$ if $T_i \leqslant c_i$ (uncensored), $V_i = 0$ if $T_i > c_i$ (censored). We refer to the c_i of individuals who in fact are observed to fail as unrealized censoring times, as contrasted with the realized censoring times of the censored individuals. The term potential censoring time is usual when c_i is considered without regard to whether censoring or failure occurs.

In some applications, all the c_i will be known, as for example if the only cause of censoring is the planned ending of follow-up at a predetermined time. Another example is so-called Type I censoring, in which all the c_i are equal, $c_i = c$, a constant under the control of the investigator. In Type II censoring, observation ceases after a predetermined number d of failures, so that c becomes a random variable. Type II censoring is a useful technique for economical use of effort in industrial life-testing. Other forms of so-called random censorship are possible. A crucial condition is that, conditionally on the values of any explanatory variables, the prognosis for any individual who has survived to c_i should not be affected if the individual is censored at c_i. That is, an individual who is censored at c should be representative of all those subjects with the same values of the explanatory variables who survive to c.

The simplest way to ensure this is to take the c_i to be in principle predetermined constants, and this viewpoint will be adopted throughout most of this book. Note, however, that often the c_i will not be known to the investigator in advance, and that the unrealized c_i corresponding to observed failures may never become known. The above condition is also satisfied if the potential censoring times are random variables c_i, which are independent of the T_i. Type II censoring is an example of a more general scheme in which, loosely speaking, censoring can depend on the past history, but not the future, of the whole process. We may call this evolutionary censoring.

1.4 Other methods of analysis

Besides the techniques to be discussed in this book, a number of other approaches have been used to analyse survival data. Perhaps the

simplest method, much used by clinicians, is to dichotomize according to survival or nonsurvival at a critical period such as five years. Comparisons of the five-year survival rates of subjects in various groups can be made using techniques for binary data. Although this approach is often quite satisfactory, it has two major disadvantages. Concentration on a single point of the survival curve necessarily wastes some information. More seriously, calculation of survival rates as simple proportions is directly possible only when no individuals are censored during the critical period. This restriction can lead to some absurdities; see Exercise 1.1.

With survival dichotomized as above, and with quantitative explanatory variables, discriminant analysis has sometimes been used to identify variables that are related to survival, although such use of discriminant analysis is better regarded as an approach to binary logistic regression. Discriminant analysis, can, however, be a useful way of sifting through a large set of variables to determine a few variables or combinations of variables which can then be considered in more detailed analyses. By itself, discriminant analysis provides little insight into the way the explanatory variables affect survival.

Reduction to a binary response is most useful when the survival of each individual is easily classified as either very short or very long. When the potential censoring times are related to the explanatory variables, discriminant analysis will give biased results. Note also that the inclusion of the actual failure time as an explanatory variable in a discriminant analysis would be a serious error, as the failure time is part of the response, not part of the factors influencing response.

In the absence of censoring, the dependence of failure time on the explanatory variables can be explored through multiple regression. Because failure times are never negative and often have highly skewed distributions, preliminary transformations of the data such as the logarithm or reciprocal are often used. The log transformation is closely related to the accelerated life model, discussed in Chapter 5. Either transformation may give undue weight to very short failure times, which will have high negative logarithms and high positive reciprocals.

1.5 Some examples

We now describe in outline three examples that will be referred to a number of times throughout the book. Other examples will be

introduced at the appropriate point in the development. Some of the examples, especially the first, have been widely used in the literature to illustrate alternative techniques.

Example 1.1 Leukaemia: comparison of two groups

Table 1.1 (Gehan, 1965, after Freireich *et al.*) shows times of remission (i.e. freedom from symptoms in a precisely defined sense) of leukaemia patients, some patients being treated with the drug 6-mercaptopurine (6-MP), the others serving as a control. Treatment allocation was randomized. Note the great dispersion and also that censoring is common in the treated group and absent in the control group. It is important to have methods of analysis that are effective in the presence of such unbalanced censoring. In fact, the trial was designed in matched pairs with one member of the pair being withdrawn from study when, or soon after, the other member comes out of remission. This is an aspect we shall ignore.

Example 1.2 Failure times and white blood count, WBC

Table 1.2 shows, for two groups of leukaemia patients, failure time (time to death) in weeks and white blood count, WBC (Feigl and Zelen, 1965). The formal difference from Example 1.1 lies partly in the presence of a continuous explanatory variable, WBC, and partly in that the division into groups is based on an (uncontrolled) measurement for each individual rather than on a randomized treatment allocation.

Example 1.3 Failure times of springs

Table 1.3 illustrates an application from industrial life-testing kindly supplied by Mr W. Armstrong. Springs are tested under cycles of repeated loading and failure time is the number of cycles to failure, it being convenient to take 10^3 cycles as the unit of 'time'. Here 60 springs were allocated, 10 to each of six different stress levels. At the lower stress levels, where failure time is long, some springs are censored, i.e. testing is abandoned before failure has occurred.

Table 1.1 *Times of remission (weeks) of leukaemia patients (Gehan, 1965, from Freireich et al.)*

Sample 0 (drug 6-MP)	6*	6,	6,	6,	7,	9*,	10*,	10,	11*,	13,	16,	17*,	19*,	20*,	22,	23,	25*,	32*,	32*,	34*,	35*	
Sample 1 (control)	1,	1,	2,	2,	3,	4,	4,	5,	5,	8,	8,	8,	8,	11,	11,	12,	12,	15,	17,	22,	23	

* Censored

Table 1.3 *Cycles to failure (in units of 10^3 cycles) of springs*

Stress (N/mm^2)												
950	225	171	198	189	189	135	162	135	117	162		
900	216	162	153	216	225	216	306	225	243	189		
850	324	321	432	252	279	414	396	379	351	333		
800	627	1 051	1 434	2 020	525	402	463	431	365	715		
750	3 402	9 417	1 802	4 326	11 520*	7 152	2 969	3 012	1 550	11 211		
700	12 510*	12 505*	3 027	12 505*	6 253	8 011	7 795	11 604*	11 604*	12 470*		

* Censored

Table 1.2 *Failure time and white blood count (Feigl and Zelen, 1965)*

(*AG positive*), $N = 17$		(*AG negative*), $N = 16$	
White blood count, WBC	*Failure time (weeks)*	*White blood count, WBC*	*Failure time (weeks)*
2 300	65	4 400	56
750	156	3 000	65
4 300	100	4 000	17
2 600	134	1 500	7
6 000	16	9 000	16
10 500	108	5 300	22
10 000	121	10 000	3
17 000	4	19 000	4
5 400	39	27 000	2
7 000	143	28 000	3
9 400	56	31 000	8
32 000	26	26 000	4
35 000	22	21 000	3
100 000	1	79 000	30
100 000	1	100 000	4
52 000	5	100 000	43
100 000	65		

1.6 Computing

Some of the simpler techniques to be described in this book can be applied to modest sets of data using a programmable (or even nonprogrammable) pocket calculator. If large amounts of data are involved or if some of the more elaborate methods of analysis are contemplated, use of the computer is essential and, under the working conditions of most statisticians, the writing of special programs is impossible on other than a very small scale. Therefore, the availability of packaged programs is crucial.

All aspects of computing change so rapidly that a very detailed discussion is not appropriate in a book like this. There follow a few notes on the position at the time of writing, 1983.

The packages GLIM (Release 4), BMDP and SAS contain programs for many of the analyses described in this book. Points to watch in the choice of program include the facilities available for checking the

model, e.g. through empirical survival curves, residual plots and user-defined time-dependent covariates, and the ease with which dummy variables, interactions, etc., may be incorporated in the model.

A logistic regression program written by P.G. Smith (see Breslow and Day, 1980) can be used to fit the multiplicative hazards model with time-dependent covariates for small data sets.

GLIM can also be used to fit some parametric models; see e.g. Aitkin and Clayton (1980) for a discussion of the Weibull distribution, in the presence of censoring. In general, methods based on the likelihood require a function maximization routine. A variety of such routines, some using derivatives of the function to be maximized, may be found in the NAG package. Ill conditioning can easily occur, particularly in attempts to discriminate between different parametric forms for the survival distribution. Routines for calculation of the complete and incomplete gamma function and its derivatives are sometimes needed; see, for example, Moore (1982), Bernardo (1976) and Schneider (1978).

Bibliographic notes, 1

A number of books on survival analysis have appeared recently. Mann *et al.* (1974), Gross and Clark (1975) and Lawless (1982) concentrate largely on fully parametric methods for particular distributions. Kalbfleisch and Prentice (1980) give a very detailed account of the multiplicative hazards model. Miller (1981) describes nonparametric and semiparametric methods. For applications in industrial reliability see Barlow and Proschan (1965, 1975), Nelson (1982) and DePriest and Launer (1983). Elandt-Johnson and Johnson (1980) describe applications in actuarial science and demography. Miké and Stanley (1982) have edited a collection of papers on medical statistics including discussion of survival data.

Armitage (1959) compared the efficiency of a number of simple methods of analysis, including the use of the proportion surviving for some specified time. Expository papers by Peto *et al.* (1976, 1977) describe the applications of some of the simpler methods for the analysis of clinical trials. Recent review papers include Prentice and Kalbfleisch (1979), Lagakos (1979) and Oakes (1981). For the mathematical theory of screening, see Prorok (1976), Shahani and Crease (1977) and Zelen and Feinleib (1969) and for an account of a large randomized trial of screening for breast cancer, see Shapiro

(1977). Three recent papers illustrating the use of survival analysis in occupational epidemiology are Liddell *et al.* (1977), Darby and Reissland (1981) and Breslow *et al.* (1983).

Further results and exercises, 1

1.1. (a) From Fig. 1.1(a) calculate the censoring times of all individuals via Fig. 1.1(b). Note that this can be done only if it is assumed that failure times can be censored solely by the conclusion of the study.

(b) The reduced sample estimator of the probability of surviving five years is the proportion, among subjects with potential censoring times exceeding five years, whose failure time is observed to exceed five years. Show that this estimator is unbiased.

(c) Show that in Fig. 1.1 the reduced sample estimators of the probabilities of surviving three years and five years are respectively 6/10 and 4/6. Comment.

(d) Show that if the third individual in Fig. 1.1(a) had actually entered two years earlier, but died at the same time, so that his survival would have been improved, the reduced sample estimate of the five-year survival rate for the entire group would be worsened, at 4/7 instead of 4/6.

1.2. Suppose that T_1, T_2 and T_3 are independent and identically distributed with a continuous distribution, and are subject to censoring times c_1, c_2 and c_3. Let $Y_i = T_i$ if $T_i \leqslant c_i$, $Y_i = \infty$ otherwise, so that Y_i may be thought of as the largest possible value of T_i consistent with the observed data. Let $X_i = \min(c_i, T_i)$. Then, on the basis of what is observed, T_1 is known to be less than or equal to T_2 if and only if $Y_1 \leqslant X_2$. Show that, whatever the values of c_1, c_2 and c_3,

(a) $\text{pr}(Y_1 \leqslant X_2) = \text{pr}(Y_2 \leqslant X_1)$,
(b) $\text{pr}(Y_1 \leqslant X_2, X_3) = \text{pr}(Y_2 \leqslant X_1, X_3) = \text{pr}(Y_3 \leqslant X_1, X_2)$,
(c) $\text{pr}(Y_1, Y_2 \leqslant X_3) = \text{pr}(Y_1 < Y_3 \leqslant X_2) + \text{pr}(Y_2 < Y_3 \leqslant X_1)$.

[Breslow, 1970]

1.3. Suppose that data are available on a reasonably homogeneous group of patients with renal failure. All patients are initially on dialysis and the time at which they start this treatment is the time origin for each patient. All patients are observed until death. Depending on the availability of suitable donor kidneys, some

patients in due course receive a kidney transplant. It is required to compare the survival under dialysis and after transplant. Criticize qualitatively the following two procedures:

(a) form two groups of patients, those never transplanted and those receiving a transplant. Compare the two distributions of time from entry to death, regardless of the time of transplant, i.e. time on dialysis of the transplanted patients is 'credited' to transplant;

(b) for the transplanted patients take a new time origin at the instant of transplant and compare the distributions of time to death for the 'dialysis only' group with that of time from transplant to death for the transplanted group.

Consider further possible procedures without the disadvantages of (a) and (b). What further difficulties are likely to arise in interpreting such data?

CHAPTER 2

Distributions of failure time

2.1 Introduction

In this chapter we consider a homogeneous population of individuals, each having a 'failure time'. That is, we deal with a single nonnegative random variable, T. In particular, an origin and scale for measuring time are assumed to be clearly defined. We examine the general specification of the distribution of T and then consider various special distributions that are useful.

We write

$$\mathscr{F}_T(t) = \text{pr}(T \geqslant t) \tag{2.1}$$

for the survivor function of T, omitting the suffix T when the random variable involved is clear from the context. Mostly we deal with continuous distributions having a probability density function

$$f_T(t) = -\mathscr{F}'_T(t) = \lim_{\Delta \to 0+} \frac{\text{pr}(t \leqslant T < t + \Delta)}{\Delta}, \tag{2.2}$$

so that

$$\mathscr{F}_T(t) = \int_t^\infty f_T(u)\, du.$$

Discrete and mixed discrete–continuous distributions can usually be handled formally by assigning to the probability density a component $f_j \delta(t - a_j)$ for an atom f_j at a_j, where $\delta(\,.\,)$ denotes the Dirac delta function. In the general case, the probability of survival beyond time t is the right-hand limit $\mathscr{F}(t + 0)$. Note that an unusual convention has been adopted in the definition (2.1) leading to the left continuity of the cumulative distribution function, rather than to the right continuity flowing from the standard definition. Our object is to simplify slightly some subsequent formulae involving the hazard function.

Particular forms of distribution may be useful either because they

are suggested by some theoretical argument or because they provide flexible empirical representations, preferably with relatively simple statistical analysis.

2.2 Hazard function

The functions $\mathscr{F}_T(\,.\,)$ and $f_T(\,.\,)$ provide two mathematically equivalent ways of specifying the distribution of a continuous nonnegative random variable, and there are of course many other equivalent functions. One with special value in the present context is the hazard function, or age-specific failure rate, defined by

$$h_T(t) = \lim_{\Delta \to 0+} \frac{\mathrm{pr}(t \leqslant T < t+\Delta \mid t \leqslant T)}{\Delta}. \tag{2.3}$$

By the definition of conditional probability, we have, omitting the suffix T, that

$$h(t) = f(t)/\mathscr{F}(t). \tag{2.4}$$

If there is an atom f_j of probability at time a_j, $h(t)$ contains a component $h_j\delta(t-a_j)$, where

$$h_j = f_j/\mathscr{F}(a_j), \tag{2.5}$$

and for a purely discrete distribution with atoms $\{f_j\}$ at points $\{a_j\}$, $a_1 < a_2 < \ldots$,

$$h(t) = \sum h_j \delta(t-a_j),$$

where

$$\begin{aligned} h_j &= f_j/\mathscr{F}(a_j) \\ &= f_j/(f_j + f_{j+1} + \ldots). \end{aligned} \tag{2.6}$$

For continuous distributions, by (2.4) and (2.2),

$$\begin{aligned} h(t) &= -\mathscr{F}'(t)/\mathscr{F}(t) \\ &= -d \log \mathscr{F}(t)/dt, \end{aligned}$$

so that, because $\mathscr{F}(0) = 1$,

$$\begin{aligned} \mathscr{F}(t) &= \exp\left(-\int_0^t h(u)\,du \right) \\ &= \exp[-H(t)], \end{aligned} \tag{2.7}$$

say, where $H(\,.\,)$ is called the integrated hazard. Further,

$$f(t) = h(t)\exp[-H(t)]. \tag{2.8}$$

If and only if $h(\,.\,)$ is constant, with value ρ say, the distribution is exponential,

$$\mathscr{F}(t) = e^{-\rho t}, \qquad f(t) = \rho e^{-\rho t}. \tag{2.9}$$

For discrete distributions, it follows on applying (2.6) recursively, or by a direct application of the product law of probabilities, that

$$\mathscr{F}(t) = \prod_{a_j < t} (1 - h_j); \tag{2.10}$$

to have $T \geqslant t$ it is necessary and sufficient to survive all points of support before t.

To define an integrated hazard in the discrete case the most fruitful convention is to take

$$H(t) = \sum_{a_j < t} \log(1 - h_j), \tag{2.11}$$

so that (2.7) still holds:

$$\mathscr{F}(t) = \exp[-H(t)].$$

If the h_j are small

$$H(t) \simeq \sum_{a_j < t} h_j \tag{2.12}$$

and the right-hand side could be taken as an alternative definition. For mixed discrete–continuous distributions, we write

$$\mathscr{F}(t) = \mathscr{P}_0^t[1 - h(u)\,du],$$

where the so-called product integral on the right-hand side is defined analogously to a Riemann integral. Divide $(0, t)$ into a large number of small intervals $[0 = x_0, x_1), [x_1, x_2), \ldots, [x_{n-1}, t = x_n)$, let $\xi_j \in [x_j, x_{j+1})$ and consider the limit $n \to \infty$, with $\max(x_{j+1} - x_j) \to 0$, of

$$\prod [1 - h(\xi_j)(x_{j+1} - x_j)],$$

where $h(\xi_j)(x_{j+1} - x_j)$ is taken to be h_k if $[x_j, x_{j+1})$ contains a point of support a_k, say.

There are a number of reasons why consideration of the hazard function may be a good idea:

(i) it may be physically enlightening to consider the immediate 'risk' attaching to an individual known to be alive at age t;

(ii) comparisons of groups of individuals are sometimes most incisively made via the hazard;

(iii) hazard-based models are often convenient when there is censoring or there are several types of failure;

(iv) comparison with an exponential distribution is particularly simple in terms of the hazard;

(v) the hazard is the special form for the 'single failure' system of the complete intensity function for more elaborate point processes, i.e. systems in which several point events can occur for each individual.

2.3 Some special distributions

We now consider in outline some of the special distributions that are useful for survival data. The simpler analytical expressions of this section are summarized in Table 2.1. Of course, any distribution over nonnegative values is a possible candidate; further, any distribution, even with support including negative real values, is a possible distribution for log T.

The distributions to be discussed are all continuous. They can be classified in various ways, one being by their relation to the exponential distribution, in particular by whether they are over- or underdispersed relative to the exponential distribution.

Greek letters are used to denote adjustable parameters; ρ always has the dimensions of the reciprocal of time and can be interpreted as a rate, whereas κ and τ are dimensionless parameters. The precise interpretation of ρ, κ and τ is, however, different for the different families.

(i) *Exponential distribution*

The exponential distribution of parameter ρ and mean $1/\rho$ has

$$\mathscr{F}(t) = e^{-\rho t}, \qquad f(t) = \rho e^{-\rho t}, \qquad h(t) = \rho, \qquad H(t) = \rho t.$$

The constant hazard reflects the property of the distribution reasonably called lack of memory. For any $t_0 > 0$, the conditional distribution of $T - t_0$, given $T > t_0$, is the same as the unconditional distribution of T.

The coefficient of variation, i.e. the ratio of standard deviation to

Table 2.1 *Properties of some special distributions*

	Survivor function	*Density function*	*Hazard*	*No. of parameters*
(i) Exponential	$e^{-\rho t}$	$\rho e^{-\rho t}$	ρ	1
(ii) Gamma	incomplete gamma function	$\dfrac{\rho(\rho t)^{\kappa-1}e^{-\rho t}}{\Gamma(\kappa)}$	—	2
(iii) Weibull	$\exp[-(\rho t)^{\kappa}]$	$\kappa\rho(\rho t)^{\kappa-1}\exp[-(\rho t)^{\kappa}]$	$\kappa\rho(\rho t)^{\kappa-1}$	2
(iv) Gompertz–Makeham	—	—	$\rho_0 + \rho_1 e^{\rho_2 t}$	3
(v) Compound exponential	$\dfrac{(\kappa/\rho_0)^{\kappa}}{(t+\kappa/\rho_0)^{\kappa}}$	$\dfrac{\kappa(\kappa/\rho_0)^{\kappa}}{(t+\kappa/\rho_0)^{\kappa+1}}$	$\dfrac{\kappa}{t+\kappa/\rho_0}$	2
(vi) Orthogonal polynomial	$e^{-\rho t}[1+\kappa_1\rho t+\kappa_2\rho t(\rho t-2)]$	$\rho e^{-\rho t}[1+\kappa_1 L_1(\rho t)+\kappa_2 L_2(\rho t)]$	—	2
(vii) Log normal	—	—	nonmonotonic	2
(viii) Log logistic	$[1+(t\rho)^{\kappa}]^{-1}$	$\kappa\rho^{\kappa}t^{\kappa-1}[1+(t\rho)^{\kappa}]^{-2}$	$\dfrac{\kappa t^{\kappa-1}\rho^{\kappa}}{[1+(t\rho)^{\kappa}]}$	2
(ix) Generalized F	—	—	—	4
(x) Inverse Gaussian	—	—	—	2
(xi) Translation	—	—	—	1 extra for origin
(xii) Scale family	$\mathscr{G}(\rho t)$	$\rho g(\rho t)$	$\rho h^{(0)}(\rho t)$	1 extra for scale
(xiii) Proportional hazard family	$[\mathscr{L}(t)]^{\psi}$	$\psi[\mathscr{L}(t)]^{\psi-1}l(t)$	$\psi h^{(0)}(t)$	1 extra for proportionality

mean, is unity and this forms a reference standard for judging relative dispersion.

The exponential distribution was widely used in early work on reliability of, for example, electronic components and to a more limited extent in medical studies. That the distribution has only one adjustable parameter often means, however, that methods based on it are rather sensitive to even modest departures, for example in the tail, and the emphasis in recent work has been on methods that make less stringent assumptions about distributional form. Various idealized models lead to the exponential distribution. For example, suppose that extreme 'loads' occur in the environment in a Poisson process and that failure occurs the first time such an extreme 'load' is encountered. Again, suppose that there are many independent modes of failure, so that the observed failure time is the smallest of a large number of independent nonnegative random variables. Then, under suitable restrictions on the component random variables, the distribution of observed failure time is approximately exponential.

Note that by (2.7) if the random variable T has an arbitrary continuous distribution, then $H(T)$ has an exponential distribution with unit parameter.

We consider next a number of two-parameter families of distributions reducing to the exponential distribution by choice of one of the parameters.

(ii) *Gamma distribution*

The gamma family has density

$$\rho(\rho t)^{\kappa-1}e^{-\rho t}/\Gamma(\kappa), \tag{2.13}$$

where $\kappa > 0$ is an additional parameter often called the index. The mean is κ/ρ and the coefficient of variation $1/\sqrt{\kappa}$. While for many statistical purposes the gamma family is the most important family of continuous distributions taking positive values, for the present purpose the usefulness is limited by the relative clumsiness of the survivor function, an incomplete gamma integral.

The special case $\kappa = 2$ may be called the two-hit model, as it corresponds to the distribution of the time to the second point in a Poisson process of rate ρ. Other integer values of κ have an analogous interpretation.

(iii) *Weibull distribution*

The Weibull distribution with scale parameter ρ and index κ has

$$\mathscr{F}(t) = \exp[-(\rho t)^{\kappa}],$$
$$f(t) = \kappa\rho(\rho t)^{\kappa-1}\exp[-(\rho t)^{\kappa}], \qquad h(t) = \kappa\rho(\rho t)^{\kappa-1}. \tag{2.14}$$

Because $H(t) = (\rho t)^{\kappa}$, it follows that T^{κ} has an exponential distribution of parameter ρ^{κ}.

The Weibull distribution arises theoretically as a limit law for the smallest of a large number of independent nonnegative random variables, thus generalizing the result already noted for the exponential distribution; see Exercise 2.7. The convenience of the Weibull distribution for empirical work stems from the simplicity of the three functions in (2.14).

(iv) *Gompertz–Makeham distribution*

A simple form of hazard function is

$$\rho_0 + \rho_1 e^{\rho_2 t} \tag{2.15}$$

with $\rho_0 = 0$ as a special case, the Gompertz form. The associated survivor function and density follow from (2.7) and (2.8).

(v) *Compound exponential distribution*

Suppose that for each individual survival time is exponentially distributed but that the rate varies randomly between individuals. To represent this let P be a random variable with density $f_P(\,.\,)$ and suppose that the conditional density of T given $P = \rho$ is

$$f_{T|P}(t|\rho) = \rho e^{-\rho t}.$$

Then the unconditional density of T is

$$f_T(t) = \int_0^{\infty} \rho e^{-\rho t} f_P(\rho)\, d\rho.$$

A convenient choice for $f_P(\,.\,)$ is the gamma density of mean ρ_0 and index κ

$$f_P(\rho) = \frac{(\kappa/\rho_0)(\kappa\rho/\rho_0)^{\kappa-1}e^{-\kappa\rho/\rho_0}}{\Gamma(\kappa)}$$

leading to the Pareto distribution

$$f_T(t) = \frac{\kappa(\kappa/\rho_0)^\kappa}{(t+\kappa/\rho_0)^{\kappa+1}}. \tag{2.16}$$

Clearly the survivor function and hazard are respectively

$$\frac{(\kappa/\rho_0)^\kappa}{(t+\kappa/\rho_0)^\kappa}, \qquad \frac{\kappa}{(t+\kappa/\rho_0)}.$$

As follows directly from its mode of construction, this distribution is overdispersed relative to the exponential distribution, to which it tends as $\kappa \to \infty$. When κ is small (2.16) has a very long tail; the rth moment exists only if $\kappa > r$.

(vi) *Expansion in orthogonal polynomials*

One way of representing distributions close to a particular simple form is via an expansion in terms of orthogonal polynomials associated with the simple form. The orthogonal polynomials associated with the exponential distribution are the Laguerre polynomials and the simplest such polynomial form uses just the first- and second-degree polynomials

$$L_1(x) = x - 1, \qquad L_2(x) = x^2 - 4x + 2,$$

taking as the density

$$\rho e^{-\rho t}[1 + \kappa_1 L_1(\rho t) + \kappa_2 L_2(\rho t)]. \tag{2.17}$$

This has mean and variance $(1+\kappa_1)/\rho$ and $(1 + 2\kappa_1 - \kappa_1^2 + 4\kappa_2)/\rho^2$, respectively.

From one point of view (2.17) is just a mixture of an exponential distribution and gamma distributions with $\kappa = 2, 3$, all however with the same rate parameter ρ. The main role of the present expansion is in a theoretical context; when the limiting distribution for a problem is exponential, an asymptotic expansion in the form (2.17) will often result.

(vii) *Log normal distribution*

As noted previously, possible distributions for T can be obtained by specifying for $\log T$ any convenient family of distributions on the real line. The simplest possibility is to take $\log T$ normally distributed with mean $\log \rho^{-1}$ and variance τ^2, leading to the log normal family for T with density

$$\frac{1}{\sqrt{(2\pi)\tau t}} \exp\left(-\frac{[\log(t\rho)]^2}{2\tau^2}\right). \tag{2.18}$$

The exponential distribution is not a special case, although a substantial amount of data is needed to discriminate empirically between an exponential distribution and a log normal distribution with $\tau \simeq 0.8$.

The hazard associated with (2.18) is nonmonotonic, although whether the maximum occurs in a range of appreciable probability depends on the value of τ. A disadvantage of the distribution for some purposes is the sensitivity of the resulting methods of statistical analysis to the small failure times.

(viii) *Log logistic distribution*

The continuous logistic density with location v and scale parameter τ, having density

$$\frac{\tau^{-1}\exp[(x-v)/\tau]}{\{1+\exp[(x-v)/\tau]\}^2},$$

is very similar to a normal distribution. If this form is taken for $\log T$, we obtain analogously to the log normal family, the log logistic family. It is convenient to write $v = -\log \rho$, $\kappa = 1/\tau$, so that the survivor function, density and hazard become respectively

$$\mathscr{F}(t) = \frac{1}{1+\exp[(\log t - v)/\tau]} = \frac{1}{1+(t\rho)^\kappa},$$

$$f(t) = \frac{\kappa t^{\kappa-1}\rho^\kappa}{[1+(t\rho)^\kappa]^2}, \tag{2.19}$$

$$h(t) = \frac{\kappa t^{\kappa-1}\rho^\kappa}{1+(t\rho)^\kappa}. \tag{2.20}$$

An advantage of this family over the log normal is the relatively simple explicit form achieved for $\mathscr{F}(t)$, $f(t)$ and $h(t)$. If $\kappa > 1$ the hazard has a single maximum; if $\kappa < 1$ the hazard is decreasing.

For the rth moment to exist, we need $\kappa > r$.

(*ix*) *Comprehensive family*

There are obvious technical advantages to combining distinct families of distributions into a single comprehensive family. In principle such families can always be constructed, indeed in many ways, but the result is usually too complicated to be very useful.

An occasionally useful general family in the present instance is obtained by taking T to be a multiple of the κ_1th power of a random variable $F_{(\kappa_2,\kappa_3)}$ having the standard (central) variance ratio distribution with (κ_2, κ_3) degrees of freedom. Thus

$$T = \rho^{-1} F^{\kappa_1}_{(\kappa_2,\kappa_3)}. \tag{2.21}$$

The three-parameter generalized gamma family is obtained with $\kappa_3 \to \infty$. In general, by choice of the dimensionless parameters $(\kappa_1, \kappa_2, \kappa_3)$, many of the distributions listed above can be obtained. Equation (2.21) can conveniently be written in terms of $\log T$.

(*x*) *Inverse Gaussian distribution*

One approximate stochastic model describes failure as the first passage time of a stochastic process representing 'wear' to a fixed barrier. If the underlying process is Brownian motion with positive drift v and variance per unit time σ^2, the first passage time to a barrier at a has the inverse Gaussian distribution with density

$$\frac{a}{\sigma(2\pi t^3)^{1/2}} \exp\left(-\frac{(a-vt)^2}{2\sigma^2 t} \right). \tag{2.22}$$

This can be reparameterized in various forms, for example as

$$\left(\frac{\kappa/\rho}{2\pi t^3} \right)^{1/2} \exp\left(-\frac{\kappa\rho(t-1/\rho)^2}{2t} \right),$$

with mean $1/\rho$ and coefficient of variation $1/\sqrt{\kappa}$, or as

$$\left(\frac{\psi}{\pi t^3} \right)^{1/2} \exp\left(-\phi t - \frac{\psi}{t} + 2\sqrt{(\phi\psi)} \right),$$

stressing the exponential family structure of the distribution.

The survivor function has the relatively complicated form

$$1 - \Phi\left[\left(\frac{\kappa}{\rho t} \right)^{1/2} (-1 + \rho t) \right] - e^{2\kappa} \Phi\left[-\left(\frac{\kappa}{\rho t} \right)^{1/2} (1 + \rho t) \right], \tag{2.23}$$

where $\Phi(\,.\,)$ is the standardized normal integral.

Although the inverse Gaussian distribution has some attractive theoretical properties and provides a reasonably flexible two-parameter family of distributions, the complexity of the survivor function makes it relatively inconvenient for handling censored data.

(*xi*) *Translation*

All the above distributions have the positive real numbers for their support. By introducing an additional parameter δ and translating the distribution, all can be converted into distributions on (δ, ∞). For instance, the translated exponential distribution has density

$$\rho e^{-\rho(t-\delta)} \qquad (t > \delta).$$

Usually we would require $\delta \geqslant 0$, although in some contexts the possibility could be contemplated of a distribution starting before the formal time origin used to define the random variable T.

(*xii*) *Scale family*

Several of the families outlined above are such that the random variable ρT has a fixed distribution, or at least a distribution involving only dimensionless shape parameters. Thus for the exponential family, ρT has the unit exponential density e^{-t}. If $\mathscr{G}(t)$, $g(t)$ and $h^{(g)}(t)$ denote respectively a survivor function, density and hazard over non-negative values, the corresponding functions

$$\begin{aligned} &\mathscr{G}(t;\rho) = \mathscr{G}(\rho t), \\ &g(t;\rho) = \rho g(\rho t), \qquad h^{(g)}(t;\rho) = \rho h^{(g)}(\rho t) \end{aligned} \tag{2.24}$$

define the scale family generated by $\mathscr{G}(\,.\,)$.

For the gamma family, for instance, $\mathscr{G}(\,.\,)$ depends also on κ.

Note that if $U = e^Z$ is a random variable with the density $g(\,.\,)$, then

$$\log T = -\log\rho + Z, \tag{2.25}$$

having thus the form of a regression model.

(*xiii*) *Lehmann family*

Another useful general family is generated from the survivor function, density and hazard $\mathscr{L}(t)$, $l(t)$ and $h^{(l)}(t)$ by considering

$$\begin{aligned} &\mathscr{L}(t;\psi) = [\mathscr{L}(t)]^{\psi}, \\ &l(t;\psi) = \psi[\mathscr{L}(t)]^{\psi-1} l(t), \qquad h^{(l)}(t;\psi) = \psi h^{(l)}(t). \end{aligned} \tag{2.26}$$

This is called the Lehmann or proportional hazards family based on $\mathscr{L}(t)$.

The families (2.24) and (2.26) are equivalent if and only if $h(t) \propto t^{\gamma}$, for some γ, so that both represent Weibull distributions; see Section 5.3 (ii).

(*xiv*) *Qualitative analytical restrictions*

A final possibility is to restrict the distribution by a qualitative analytical requirement. For example, we may require only that the hazard is nondecreasing, giving the so-called IFR (increasing failure rate) family. The analogous family, DFR, in which the hazard is nonincreasing, is rather less important.

2.4 Comparison of distributions

The families of distributions outlined above can be judged by

(i) their technical convenience for statistical inference;

(ii) the availability of explicit reasonably simple forms for survivor function, density and hazard;

(iii) the capability of representing both over- and underdispersion relative to the exponential distribution;

(iv) the qualitative shape (monotonicity, log concavity, etc.) of the hazard;

(v) the behaviour of the survivor function for small times;

(vi) the behaviour of the survivor function for large times, as judged either directly or by suitable dimensionless ratios of cumulants or moments;

(vi) any connection with a special stochastic model of failure.

Table 2.2 summarizes some of these properties in concise form. In many applications there will be insufficient data to choose between different forms by empirical analysis and then it is legitimate to make the choice on grounds of convenience; points (i) and (ii) are fairly closely related, especially when censored data are to be analysed. Behaviour for small t will be critical for some industrial applications, for instance where guarantee periods are involved, but for most medical applications the upper tail referring to relatively long survival times will be of more interest.

There are several ways of comparing different families either to

Table 2.2 *Some properties useful in assessing distributional form*

	$\log h(t)$	$H(t)$	$\log H(t)$	*Coefficient of variation*
Is it	constant?	linear in t?	—	1?
	exponential	exponential		exponential
Is it	linear in t?	—	linear in t?	<1?
	Gompertz ($\rho_0 = 0$)		Gompertz ($\rho_0 = 0$)	Gamma ($\kappa > 1$), Weibull ($\kappa > 1$)
				Log normal ($\tau < 0.83$),
				Log logistic ($\tau < 0.118$)
Is it	linear in $\log t$?	—	linear in $\log t$?	—
	Weibull		Weibull	
Is it	nonmonotonic?	—	asymptotically	>1?
	Log normal		linear in t?	Gamma ($\kappa < 1$), Weibull ($\kappa < 1$)
	Log logistic		Distribution with	Log normal ($\tau > 0.83$),
			exponential tail	Log logistic ($\tau > 0.118$)
				Compound exponential

highlight their differences or as a basis for an empirical analysis. On the whole, direct consideration of the density is not very effective and we concentrate here on plotting or tabulating

(a) the hazard or log hazard versus t or log t;
(b) the integrated hazard, or log survivor function, or some transform, versus t or log t;
(c) the value of the coefficient of variation, $\gamma = \sigma/\mu$;
(d) the standardized third moment $\gamma_3 = \mu_3/\sigma^3$ versus the coefficient of variation $\gamma = \sigma/\mu$, where μ, σ and μ_3 are the mean, standard deviation and third central moment of T.

Properties based on the hazard or integrated hazard lead directly to methods for analysing data that will be applicable in the presence of censoring. The integrated hazard, or log survivor function, has the advantage of indicating directly the behaviour of the upper tail of the distribution and of leading to a reasonably smooth plot when applied to empirical data. Hazard has the advantage of leading to empirical plots with points with independent errors which are, however, therefore inevitably less smooth than those of integrated hazard! It is advisable that in plots of the hazard the abscissa should be calibrated not only by t, or some function of t, but by $\mathscr{F}(t)$, so that the ranges of most concern are clear.

A value of the coefficient of variation less than one immediately excludes those distributions capable of representing only over-dispersion relative to the exponential distribution.

The graph (d) of moment ratios is particularly useful within the scale family, Section 2.3(xii), in which there is a single shape parameter. The dimensionless ratios are independent of ρ and hence the family is characterized by a curve. Fig. 2.1 shows these curves for gamma, Weibull, log normal and log logistic distributions. A high value of μ_3/σ^3, the third moment ratio, implies a relatively long tail.

The third moment has a large sampling error for long-tailed distributions and it would be possible to consider instead plots based on the standardized moments about the origin of order 3/2 or 1/2, for instance. That is with $E(T^r) = \mu'_r$, we could plot $\mu'_{3/2}/\sigma^{3/2}$ versus σ/μ or $\mu'_{1/2}/\sigma^{1/2}$ versus σ/μ; to emphasize the lower tail, we could consider $\mu'_{-1/2}\sigma^{1/2}$. The relative sensitivity of these and many other broadly similar plots is unclear.

Similar quantities for guiding the choice of models can also be calculated from log T. From censored data, comparison via the

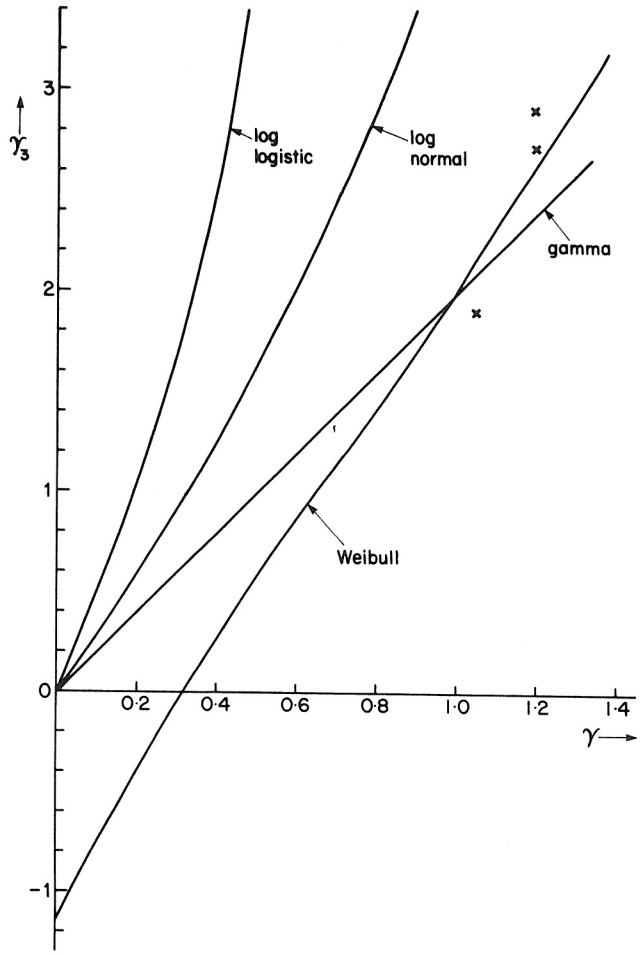

Fig. 2.1. Standardized third moment, γ_3, versus coefficient of variation, γ, for gamma, Weibull, log normal and log logistic families. Exponential distribution is at point (1, 2). ×, Boag's (1949) cancer data.

hazard or log hazard is probably the most widely useful approach, because moments cannot be calculated from censored data without strong assumptions.

Boag (1949) gave three sets of hospital cancer data and showed that the log normal distribution gives a rather better fit than the

exponential. The data are quite heavily grouped and there is a little censoring. Nevertheless the coefficient of variation and standardized third moment can be estimated, attaching reasonable extrapolated values for the censored observations. The results are plotted in Fig. 2.1 and lie fairly close to the Weibull curve, showing more skewness than the gamma and less skewness than the log normal.

Finally, recall, as noted in Section 2.3(i), that one interpretation of the integrated hazard $H(\,.\,)$ is that $H(T)$ has the unit exponential distribution, i.e. $H(\,.\,)$ specifies the transformation of the timescale necessary to induce a unit exponential distribution.

Bibliographic notes, 2

The use of hazard to describe distributions of survival time has a long history in the actuarial literature. For a modern account of survival data with some emphasis on actuarial techniques, see Elandt-Johnson and Johnson (1980) and for accounts emphasizing the fitting of special distributions, see Gross and Clark (1975) and Lawless (1982).

The exponential distribution was probably studied first in connection with the kinetic theory of gases (Clausius, 1858). It plays a central role in the theory of point processes (Cox and Isham, 1980; Cox and Lewis, 1966). The Weibull distribution was introduced by Fisher and Tippett (1928) in connection with extreme value distributions; Weibull (1939a, b) studied it in an investigation of the strength of materials. Several of the other distributions are quite widely used in other statistical contexts. For the generalized F distribution, see Kalbfleisch and Prentice (1980, p. 28). Properties of the inverse Gaussian distribution are reviewed by Folks and Chhikara (1978) and Jørgensen (1982). For a summary of the properties of the main univariate continuous distributions, see Johnson and Kotz (1970). Vaupel *et al.* (1979) and Hougaard (1984) have examined the effect of heterogeneity between individuals via a notion of frailty.

For an account of distributions characterized by a descriptive property of the hazard, see Barlow and Proschan (1975).

Further results and exercises, 2

2.1. Suppose that a continuous random variable T is converted into a discrete random variable by grouping. Suppose that $[t-a, t)$, $[t, t+a)$ are two adjacent groups. Examine how the values of $\mathscr{F}(t-a)$, $\mathscr{F}(t)$ and $\mathscr{F}(t+a)$ can best be used to determine $h(t)$, the underlying continuous hazard at t.

2.2. Let $T_1, \ldots, T_n$ be independent continuous nonnegative random variables with hazard functions $h_1(\,.\,), \ldots, h_n(\,.\,)$. Prove that $T = \min(T_1, \ldots, T_n)$ has hazard function $\sum h_j(t)$.

2.3. Let $T_1, \ldots, T_n$ be independent random variables with Weibull distributions with rate parameters $\rho_1, \ldots, \rho_n$ and common index κ. Prove that $T = \min(T_1, \ldots, T_n)$ also has a Weibull distribution of index κ.

2.4. In a compound exponential distribution, let the rate be represented by the random variable P. Prove that

$$E(T) = E(1/P),$$
$$\operatorname{var}(T) = 2E(1/P^2) - [E(1/P)]^2.$$

Check the results from the case where P has a gamma distribution.

2.5. Verify that special cases of the generalized F distribution of Section 2.3(ix) are achieved as follows:

$\kappa_1 = 1$, $\kappa_2 = 2$, $\kappa_3 \to \infty$, exponential;
$\kappa_1 = 1$, κ_2 arbitrary, $\kappa_3 \to \infty$, gamma;
κ_1 arbitrary, $\kappa_2 = 2$, $\kappa_3 \to \infty$, Weibull;
$\kappa_2, \kappa_3 \to \infty$, log normal;
κ_1 arbitrary, $\kappa_2 = \kappa_3 = 2$, log logistic.

For the last two cases examine the moment generating function.

2.6. Prove that for the compound exponential distribution of Section 2.3(v) both density and survivor function are completely monotonic, whatever the mixing density $f_P(\,.\,)$. List some consequences.

[Widder, 1946, Chapter 4]

2.7. Suppose that $V_1, \ldots, V_m$ are independent and identically distributed continuous nonnegative random variables such that as $v \to 0$ the density and survivor function are asymptotically $av^{\kappa-1}$ and $1 - av^\kappa/\kappa$ respectively, where $a > 0$ and $\kappa > 0$. If $W = \min(V_1, \ldots, V_m)$ and $T = (a/\kappa)^{1/\kappa} m^{1/\kappa} W$, prove that, as $m \to \infty$, T has as its limiting distribution the Weibull distribution of index κ, the exponential distribution being the limiting distribution in the special case $\kappa = 1$.

2.8. If in the representation of Exercise 2.7 the density of the V_i's is, as $v \to 0$, asymptotically $av^{\kappa-1}(1 + b_1 v + b_2 v^2 + \cdots)$, obtain expansions for the asymptotic survivor function and density of T as $m \to \infty$, showing in particular that the survivor function of T has the form

$$\begin{cases} \exp(-t^\kappa)(1 - \frac{1}{2}t^{2\kappa}/m + \ldots) & (\kappa < 1), \\ \exp(-t^\kappa)(1 - ct^{\kappa+1}/m^{1/\kappa} + \ldots) & (\kappa \geqslant 1), \end{cases}$$
$$c = b_1 \kappa^{1+1/\kappa}(\kappa+1)^{-1} a^{-1/\kappa} + \tfrac{1}{2}\delta_{\kappa 1},$$

with $\delta_{\kappa 1} = 1$ $(\kappa = 1)$, $\delta_{\kappa 1} = 0$ $(\kappa \neq 1)$.

Examine more general expansions for the originating density such as $at^{\kappa-1}(1 + b_1 t^\eta + \ldots)$.

2.9. If T has the Weibull distribution with parameters ρ and κ, prove that $U = \log T$ has the Gumbel distribution with survivor function $\exp(-\rho^\kappa e^{\kappa u})$ and density $\kappa\rho^\kappa \exp(\kappa u - \rho^\kappa e^{\kappa u})$. Write this in scale and location form by reparameterization. Obtain the Gumbel distribution also as the limiting distribution of $\rho\kappa T - \kappa$ as $\kappa \to \infty$.

2.10. Suppose that an individual selected at random has hazard $\rho^2 t + V$, where V is an unobserved random variable having a gamma distribution. Prove that the unconditional hazard has the form $\rho^2 t + \xi/(1 + \eta t)$, where ξ and η are parameters determined by the gamma density. Show that this can take the 'bath-tub' form with a local minimum.

[Borgefors and Hjorth, 1981]

2.11. Show that if the hazard function has the form

$$\kappa\rho(\rho t)^{\kappa-1}\exp[(\rho t)^\kappa]$$

the survivor function is

$$\exp\{-[\exp((\rho t)^\kappa) - 1]\}.$$

[Dhillon, 1979, 1981]

2.12. Prove that the square of the coefficient of variation of the log logistic distribution is $(\kappa/\pi)\tan(\pi/\kappa) - 1$, for $\kappa > 2$.

2.13. Show that the Gompertz–Makeham distribution with hazard

$$h(t) = \rho_0 + \rho_1 e^{\rho_2 t}$$

can be obtained as a compound exponential distribution provided $\rho_2 \leqslant 0$, and determine the distribution of the mixing random variable P.

2.14. Show that for the log normal distribution, the curve in Fig. 2.1 has equation

$$y = 3x + x^3,$$

where y is the standardized third moment and x is the coefficient of variation.

CHAPTER 3

Parametric statistical analysis: single sample

3.1 Introduction

In Chapter 2 we described some parametric families of survival distributions and gave some criteria for the appropriate choice of family in applications. We now suppose that a specific family has been selected, so that the distribution is known up to a vector parameter ϕ and that there is available for inference about ϕ a single sample of failure times, possibly subject to censoring. Often we may write $\phi^{\mathrm{T}} = (\omega^{\mathrm{T}}, \lambda^{\mathrm{T}})$, where ω is a parameter of particular interest and λ a nuisance parameter.

Here and throughout the book, we concentrate on methods based on the likelihood function. Iterative numerical solution of the likelihood equations is nearly always involved, and, as mentioned in Chapter 1, the availability of suitably flexible computer programs is crucial. After deriving the general form of the likelihood function for a censored sample, we briefly review methods of inference based on large-sample maximum likelihood theory. The exponential and Weibull distributions are considered in more detail as illustrations of the general approach. Unusually, for the exponential distribution, some 'exact' sampling theory is available.

3.2 The likelihood function

We consider first the case where the survival distribution is continuous. A subject observed to fail at t contributes a term $f(t;\phi)$ to the likelihood, the density of failure at t. The contribution from a subject whose survival time is censored at c is $\mathscr{F}(c;\phi)$, the probability of survival beyond c. The full likelihood from n independent subjects, indexed by i, is then

$$\text{lik} = \prod_{\mathrm{u}} f(t_i;\phi) \prod_{\mathrm{c}} \mathscr{F}(c_i;\phi), \tag{3.1}$$

where the two products are taken over uncensored and censored subjects respectively. The log likelihood is

$$l = \sum_{\mathrm{u}} \log f(t_i;\phi) + \sum_{\mathrm{c}} \log \mathscr{F}(c_i;\phi), \tag{3.2}$$

with a similar convention for the summations.

In terms of the observed failure or censoring time $x_i = \min(t_i, c_i)$, this becomes

$$l = \sum_{\mathrm{u}} \log f(x_i;\phi) + \sum_{\mathrm{c}} \log \mathscr{F}(x_i;\phi).$$

Since $f(t) = h(t)\mathscr{F}(t)$, this may be written

$$l = \sum_{\mathrm{u}} \log h(x_i;\phi) + \sum \log \mathscr{F}(x_i;\phi).$$

The log survivor function is minus the integrated hazard, so that

$$l = \sum_{\mathrm{u}} \log h(x_i;\phi) - \sum H(x_i;\phi).$$

Finally, on setting $r(u) = \text{card}\{i : x_i \geqslant u\}$, the number of subjects still in view at time u, we note that l may be written

$$l = \sum_{\mathrm{u}} \log h(x_i;\phi) - \int_0^\infty r(u)h(u;\phi)\,du. \tag{3.3}$$

Of course, the integral is only formally over an infinite range, because $r(u)$ will be zero beyond the last observed survival or censoring time. The integrand may be interpreted as the total hazard operating at time u. These expressions for l emphasize the fundamental role played by the hazard function in the development.

Suppose now that the survival distribution is discrete, with atoms $f_j(\phi)$ at preassigned points $a_j (a_1 < a_2 < \cdots)$. We shall assume that an individual censored at c could have been observed to fail at c. With this convention, the contribution to the likelihood from a subject observed to fail at a_j is $f_j(\phi)$, and from a subject censored at c is

$$\text{pr}(T > c) = \mathscr{F}(c+;\phi) = 1 - \sum_{j:a_j \leqslant c} f_j(\phi).$$

In terms of the discrete hazard function $h_j(\phi)$ given by (2.6) we have as in (2.10)

$$f_j(\phi) = h_j(\phi) \prod_{k<j} [1 - h_k(\phi)],$$

$$\mathscr{F}(c+;\phi) = \prod_{j:a_j \leqslant c} [1 - h_j(\phi)].$$

Each term is a product over the atoms $\{a_j\}$ of the survival distribution.

To derive the full likelihood from a sample of n observations, we first collect all the terms corresponding to the atom a_j. If there are d_j failures among the $r_j = r(a_j)$ individuals in view at a_j, the contribution to the total likelihood is

$$[h_j(\phi)]^{d_j}[1-h_j(\phi)]^{r_j-d_j}.$$

The total log likelihood is then

$$\sum_j \{d_j \log h_j(\phi) + (r_j - d_j)\log[1-h_j(\phi)]\}. \tag{3.4}$$

Note that this is the same as would be obtained from a series of independent binomial terms, with r_j trials and probability of 'success' $h_j(\phi)$.

In practice, truly discrete survival distributions are rarely encountered. Ties in reported values are usually due to the grouping of data from an underlying continuous distribution. For most purposes, and especially for the single-sample problem, the consequent small inaccuracies in the data can generally be ignored. An exception to this rule is the fitting of a log normal distribution to data with many values close to zero. As noted in Chapter 2, the fitted parameters are sensitive to the very short survival times.

The exact likelihood for grouped data can be derived: it involves integrals of the density function over the grouping intervals.

3.3 Likelihood theory: general considerations

Various approaches are possible to the extraction of information about ϕ from the log likelihood function (3.2) or (3.4). If a prior distribution is available for the unknown parameter the usual calculations of Bayesian theory lead to the posterior distribution of the parameter of interest. Note that in the case $\phi^{\mathrm{T}} = (\omega^{\mathrm{T}}, \lambda^{\mathrm{T}})$, a joint prior distribution is needed over the parameter of interest ω and the nuisance parameter λ. If a sampling theory approach is used, it may be possible to develop 'exact' confidence intervals and tests, perhaps eliminating the nuisance parameter by a conditioning argument.

More commonly, the asymptotic considerations of maximum likelihood theory are used. Three broad types of asymptotic procedure, based on likelihood, are available for testing the null hypothesis $\omega = \omega_0$ and hence for deriving a confidence set for ω as the

collection of parameter values not 'rejected' at the level in question. These types are as follows:

(a) First, there is the direct use of the likelihood ratio statistic

$$W(\omega_0) = W = 2[l(\hat{\omega},\hat{\lambda}) - l(\omega_0, \hat{\lambda}_{\omega_0})], \tag{3.5}$$

where $(\hat{\omega},\hat{\lambda})$ is the joint maximum likelihood estimate of (ω,λ) and $\hat{\lambda}_{\omega_0}$ is the maximum likelihood estimate of λ when $\omega = \omega_0$. The function $l(\omega,\hat{\lambda}_\omega)$ is sometimes called the profile log likelihood for ω. Under the null hypothesis $\omega = \omega_0$, $W(\omega_0)$ has, approximately, a chi-squared distribution with $p_\omega = \dim(\omega)$ degrees of freedom. The corresponding $1-\alpha$ confidence region is

$$\{\omega : W(\omega) \leqslant c^*_{p_\omega,\alpha}\}, \tag{3.6}$$

where $c^*_{p,\alpha}$ is the upper α point of the chi-squared distribution with p degrees of freedom. If the asymptotic distribution were exact, we would have $E[W(\omega_0);\omega_0] = p_\omega$. Often it is possible to find an expansion

$$E[W(\omega_0);\omega_0] = p_\omega\left[1 + \frac{c}{n} + o\left(\frac{1}{n}\right)\right].$$

Then $(1 + c/n)$, with if necessary c estimated consistently, is called a Bartlett correction factor and improved properties are obtained by replacing W by

$$W' = W/(1 + c/n)$$

in (3.5) and (3.6). It is rarely feasible, however, to carry out such calculations in the presence of censoring.

(b) Secondly, we may make direct use of the maximum likelihood estimate $\hat{\omega}$. The observed information matrix is the matrix of minus the second derivatives of l with respect to (ω,λ), evaluated at $(\hat{\omega},\hat{\lambda})$. Write $v_{\omega\omega}(\hat{\omega},\hat{\lambda})$ for the leading submatrix of the inverse of the observed information matrix; it can be regarded as the estimated covariance matrix of $\hat{\omega}$. Then we may use instead of (3.5) the Wald statistic

$$W_e(\omega_0) = (\hat{\omega} - \omega_0)^{\mathrm{T}} v^{-1}_{\omega\omega}(\hat{\omega},\hat{\lambda})(\hat{\omega} - \omega_0), \tag{3.7}$$

again with an approximate chi-squared distribution with p_ω degrees of freedom under the null hypothesis. Equation (3.7) leads directly to an elliptical confidence region for ω, centred on $\hat{\omega}$. There are alternative ways of estimating the covariance matrix, for example via

expected rather than observed second derivatives of the log likelihood. If ω is a scalar parameter, there results the symmetric $1-2\alpha$ confidence interval

$$\hat{\omega} - k_\alpha^* v_{\omega\omega}^{1/2}(\hat{\omega},\hat{\lambda}), \qquad \hat{\omega} + k_\alpha^* v_{\omega\omega}^{1/2}(\hat{\omega},\hat{\lambda}),$$

where $\Phi(-k_\alpha^*) = \alpha$.

(c) A third possibility is to use the gradient of the log likelihood at ω_0, replacing λ by $\hat{\lambda}_{\omega_0}$, i.e. to calculate

$$U_{\omega_0} = \left[\frac{\partial}{\partial\omega} l(\omega,\lambda)\right]_{\omega=\omega_0,\,\lambda=\hat{\lambda}_{\omega_0}}. \tag{3.8}$$

This $p_\omega \times 1$ vector is, when $\omega = \omega_0$, approximately normally distributed with zero mean and covariance matrix $v_{\omega\omega}^{-1}(\omega_0,\hat{\lambda}_{\omega_0})$. The test statistic based on U_{ω_0} is

$$W_U(\omega_0) = U_{\omega_0}^{\mathrm{T}} v_{\omega\omega}(\omega_0,\hat{\lambda}_{\omega_0}) U_{\omega_0}. \tag{3.9}$$

Again there are alternative ways of evaluating the covariance matrix, and the distribution under the null hypothesis is approximately chi-squared with p_ω degrees of freedom.

We note in passing that the estimated covariance matrix, evaluated at $\omega = \omega_0$, $\lambda = \hat{\lambda}_0$, of

$$\left[\frac{\partial}{\partial\omega} l(\omega,\lambda)\right]$$

is $I_{\omega\omega}(\omega_0,\hat{\lambda}_{\omega_0})$, the leading submatrix of the observed information matrix. In general, $I_{\omega\omega}(\omega_0,\hat{\lambda}_{\omega_0}) \neq v_{\omega\omega}^{-1}(\omega_0,\hat{\lambda}_{\omega_0})$; the difference represents the gain in information about ω provided by knowledge of λ.

The three procedures (a)–(c) will very often give virtually identical answers. Procedure (b), the direct use of the maximum likelihood estimate, has advantages in simple presentation of conclusions, but the disadvantage is that it is not invariant under reparameterization and that it may yield absurd answers if the likelihood is of unusual shape, e.g. is multimodal or zero for certain values of ϕ. Procedure (c) has some computational advantage in that only maximization at $\phi = \phi_0$ is required, so that the adequacy of a basic model specified by λ can be tested by augmentation in various directions without re-maximization. In cases of doubt, procedure (a), the direct use of maximized log likelihoods is recommended. It is invariant under reparameterization and the shape of the resulting confidence region is settled by the data. These qualitative arguments are reinforced by recent work on higher-order asymptotic theory, in particular

involving conditioning on approximately ancillary statistics (Efron and Hinkley, 1978; Cox, 1980; Barndorff-Nielsen, 1980, 1983; Barndorff-Nielsen and Cox, 1984).

The asymptotic theory of maximum likelihood estimation on which the normal and chi-squared approximations are based does require the satisfaction of some 'regularity conditions' concerning the smoothness of the likelihood function. In particular, the theory does not hold for threshold parameters. With observations that are not independent or identically distributed, it is necessary, roughly speaking, that the proportion of the total information in the sample contributed by any single observation should converge to zero as the sample size increases. Of course, even if the problem is such that the conditions for asymptotic normality hold, the theory may give a poor approximation to the small-sample results.

If there is serious doubt about the applicability of asymptotic distribution theory, a Bartlett correction factor can be calculated or computer simulation used to examine the distribution on the null hypothesis of any appropriate statistic.

3.4 Exponentially distributed failure times

The exponential distribution, with rate parameter ρ, $\mathscr{F}(t) = e^{-\rho t}$, has constant hazard function $h(t,\phi) = h(t,\rho) = \rho$. The log likelihood for the single unknown parameter ρ is thus

$$l = \textstyle\sum_{\mathrm{u}} \log\rho - \rho\sum x_i = d\log\rho - \rho\sum x_i, \tag{3.10}$$

say, and the total number d of failures and the total $\sum x_i$ of the censored and uncensored failure times form a minimal sufficient statistic for ρ. Note that unless d or $\sum x_i$ or some function of them is fixed by design, we have a two-dimensional statistic for a one-dimensional parameter, showing an example of a so-called curved exponential family. Often $\sum x_i$ is called the total time at risk.

The derivatives of l are

$$U_\rho = \partial l/\partial\rho = d/\rho - \textstyle\sum x_i, \tag{3.11}$$

$$I = -\partial^2 l/\partial\rho^2 = d/\rho^2. \tag{3.12}$$

The maximum likelihood estimator $\hat{\rho}$ of ρ is the solution of $U_\rho = 0$, namely

$$\hat{\rho} = d/\textstyle\sum x_i, \tag{3.13}$$

the total number of failures divided by the total time at risk. Censored failure times contribute to the denominator but not to the numerator of this ratio.

When there is no censoring, the log likelihood becomes

$$l = n \log \rho - \rho \sum x_i,$$

and the curved exponential family collapses to a full one-dimensional family with the single minimal sufficient statistic $\sum x_i$ for ρ. Here, exact inference for ρ is possible, because $\sum x_i$, the sum of n independent exponentially distributed random variables with the same parameter ρ, has a gamma distribution with index n and scale parameter ρ. Thus $2n\rho/\hat{\rho}$ has a chi-squared distribution with $2n$ degrees of freedom. Interval estimates and hypothesis tests for ρ follow immediately. In particular, a $1-\alpha$ confidence interval for ρ is

$$\frac{\hat{\rho} c^*_{2n,1-\frac{1}{2}\alpha}}{2n} < \rho < \frac{\hat{\rho} c^*_{2n,\frac{1}{2}\alpha}}{2n},$$

where $c^*_{p,\alpha}$ is the upper α point of the chi-squared distribution with p degrees of freedom.

Example 3.1

Consider the leukaemia data of Freireich *et al.* given in Table 1.1. For the control group, with no censoring, $n = 21$ and $\sum x_i = 182$. If an exponential distribution is assumed

$$\hat{\rho} = 21/182 = 0.115,$$

and an exact 95% confidence interval for ρ is (0.071, 0.170), since the 0.025 and 0.975 points of the chi-squared distribution with 42 degrees of freedom are respectively 26.0 and 61.8.

The exact theory holds also with Type II censoring, that is when observation ceases after a predetermined number of failures, d. This is easily seen by noting that if $t_{(i)}$ denotes the ith ordered failure time ($i = 0, 1, \ldots, d$: $t_{(0)} = 0$), then $t_{(i)} - t_{(i-1)}$ has an exponential distribution with parameter $(n-i+1)\rho$ and that

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{d} (n-i+1)(t_{(i)} - t_{(i-1)}),$$

with this censoring mechanism.

With other censoring patterns, the exact sampling distribution of $\hat{\rho}$ is difficult to derive. It has been tabulated for the special case of Type I censoring, when observation on all individuals ceases at a predetermined time c. However, good approximate procedures for general censoring patterns can be obtained by treating $2d\rho/\hat{\rho}$ as a chi-squared variable on $2d$ degrees of freedom, ignoring the fact that d is now a random variable. The resulting confidence intervals are very similar to those obtained from the likelihood ratio.

Example 3.1 (continued)

For the treated group (6-MP), $\sum x_i = 359$, $d = 9$. The maximum likelihood estimator is

$$\hat{\rho} = d/\sum x_i = 9/359 = 0.025.$$

The log likelihood function is plotted in Fig. 3.1 and shows a noticeable lack of symmetry. The 95% confidence interval for ρ from the likelihood ratio is (0.0120, 0.0452). The interval obtained from the upper and lower 0.025 points of the chi-squared distribution with 18

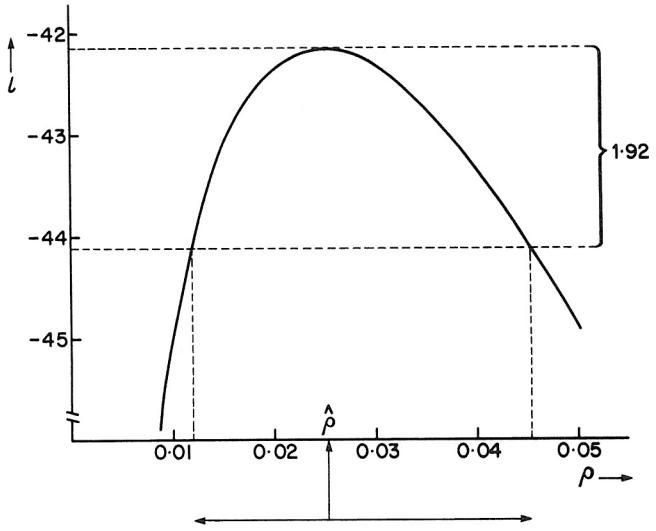

Fig. 3.1. Leukaemia data, 6-MP group. Log likelihood function, l, for exponential parameter, ρ. 95% confidence interval derived from chi-squared distribution. Maximum likelihood estimate, $\hat{\rho}$.

degrees of freedom is (0.0115, 0.0439). The standard error of the maximum likelihood estimate is

$$\left(\left[-\frac{\partial^2 l}{\partial \rho^2}\right]_{\hat{\rho}}\right)^{-1/2} = \left(\frac{\hat{\rho}^2}{d}\right)^{1/2} = 0.00836.$$

This gives a symmetric 95% confidence interval, based on a normal approximation to the distribution of $\hat{\rho}$, of (0.0087, 0.0414). In view of the shape of the likelihood function, this symmetric interval would be an inappropriate choice here.

3.5 Proportional hazards family

The likelihood equations take a simple form for the Lehmann family $h(t;\rho) = \rho h_0(t)$, where the hazard function $h_0(t)$ is assumed known. In fact, if

$$H_0(t) = \int_0^t h_0(u)\,du$$

denotes the integrated hazard corresponding to $h_0(t)$, the random variables $T_i' = H_0(T_i)$ will have exponential distributions with parameter ρ, so that the methods of the previous section apply.

Alternatively, we may proceed directly from the log likelihood. For a general censored sample, equation (3.3) gives

$$l = d\log\rho + \textstyle\sum_{\mathrm{u}} \log h_0(x_i) + \rho e, \tag{3.14}$$

where

$$e = \int_0^\infty r(u)h_0(u)\,du = \sum_{i=1}^{n} H_0(x_i).$$

It is easily seen that, whatever the censoring mechanism, the random variable ρe has the same expectation as the number d of observed failures; see Exercise 3.2. The derivative of l is

$$dl/d\rho = d/\rho - e, \tag{3.15}$$

leading to the simple form

$$\hat{\rho} = d/e \tag{3.16}$$

for the maximum likelihood estimator. In epidemiological applications, the $h_0(t)$ may represent known age-specific mortality rates for a given 'standard population', and d the number of deaths observed in a 'study population' of interest. The ratio d/e, possibly expressed as a

percentage, is called the standardized mortality ratio (SMR). It is usually necessary to allow for dependence of the hazard function on calendar time as well as age, in the calculations of e.

The $1-\alpha$ 'limits on the expectation' of a Poisson variable with observed value k, are

$$\theta_L(k) = \bar{\theta}(k, \tfrac{1}{2}\alpha), \qquad \theta_U(k) = \bar{\theta}(k+1, 1-\tfrac{1}{2}\alpha),$$

where $\bar{\theta}(k, \alpha)$ is the root of

$$\sum_{i=k}^{\infty} \frac{\theta^i}{i!}\exp(-\theta) = \alpha. \tag{3.17}$$

A $1-\alpha$ confidence interval for ρ can be calculated from these limits as $(\theta_L(d)/e, \theta_U(d)/e)$, i.e. treating d as if it had a Poisson distribution with mean ρe, with e nonrandom. It can be shown that the lower limit of this interval is the same as that obtained by taking $2\rho e$ to have a chi-squared distribution with $2d$ degrees of freedom. For the upper limit, the degrees of freedom, however, must be taken as $2d+2$.

3.6 Likelihood estimation for the Weibull distribution

We now consider maximum likelihood estimation for the parameters (κ, ρ), both assumed unknown, of the Weibull distribution, with hazard $h(t) = \kappa\rho(\rho t)^{\kappa-1}$. From (3.2) the log likelihood from a censored sample is

$$l = d\log\kappa + \kappa d\log\rho + (\kappa-1)\textstyle\sum_{\mathrm{u}} \log x_i - \rho^\kappa \sum x_i^\kappa.$$

Even in the absence of censoring, there is no fixed dimensional sufficient statistic for (ρ, κ); the Weibull is not an exponential family. The first derivatives are

$$U_\rho = \frac{\partial l}{\partial \rho} = \frac{\kappa d}{\rho} - \kappa\rho^{\kappa-1}\textstyle\sum x_i^\kappa, \tag{3.18}$$

$$U_\kappa = \frac{\partial l}{\partial \kappa} = \frac{d}{\kappa} + d\log\rho + \textstyle\sum_{\mathrm{u}} \log x_i - \rho^\kappa \sum x_i^\kappa \log(\rho x_i). \tag{3.19}$$

If κ is specified, the maximum likelihood estimator $\hat{\rho}_\kappa$ of ρ can be found explicitly by solving $U_\rho = 0$ as

$$\hat{\rho} = (d/\textstyle\sum x_i^\kappa)^{1/\kappa}, \tag{3.20}$$

a result which could be derived immediately from the fact that T^κ has an exponential distribution with parameter ρ^κ. Substitution into the

equation $U_\kappa = 0$ yields the simpler form

$$0 = \frac{d}{\kappa} + \sum_u \log x_i - d\frac{\sum x_i^\kappa \log x_i}{\sum x_i^\kappa} \tag{3.21}$$

for the maximum likelihood estimator $\hat{\kappa}$. Equation (3.21) does not contain ρ and can be solved by a one-dimensional iterative scheme in κ.

The second derivatives of l are

$$-I_{\rho\rho} = \frac{\partial^2 l}{\partial \rho^2} = -\frac{\kappa d}{\rho^2} - \kappa(\kappa - 1)\rho^{\kappa-2}\sum x_i^\kappa, \tag{3.22}$$

$$-I_{\rho\kappa} = \frac{\partial^2 l}{\partial \rho \partial \kappa} = \frac{d}{\rho} - \rho^{\kappa-1}(1 + \kappa \log \rho)\sum x_i^\kappa - \kappa\rho^{\kappa-1}\sum x_i^\kappa \log x_i, \tag{3.23}$$

$$-I_{\kappa\kappa} = \frac{\partial^2 l}{\partial \kappa^2} = -\frac{d}{\kappa^2} - \rho^\kappa \sum x_i^\kappa [\log(\rho x_i)]^2. \tag{3.24}$$

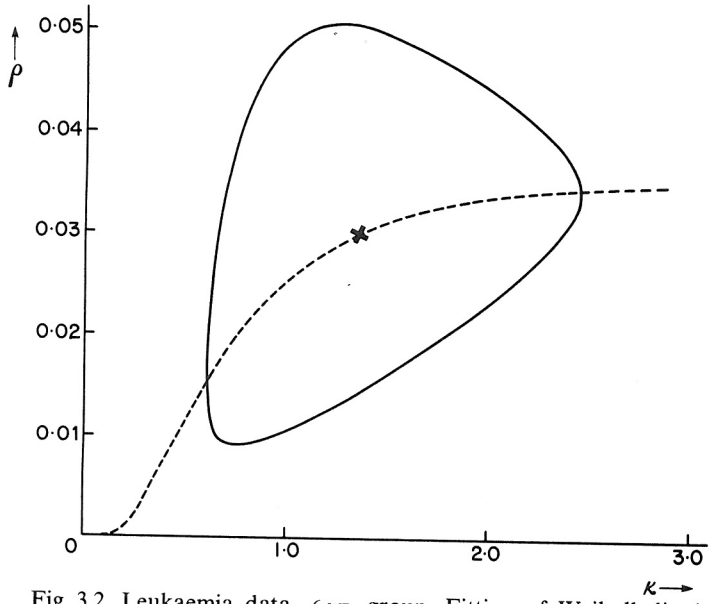

Fig. 3.2. Leukaemia data, 6-MP group. Fitting of Weibull distribution: ———, boundary of 95% confidence region based on likelihood ratio statistic; – – – –, maximum likelihood estimate $\hat{\rho}_\kappa$ of ρ for given κ; ×, maximum likelihood estimate $(\hat{\kappa}, \hat{\rho})$.

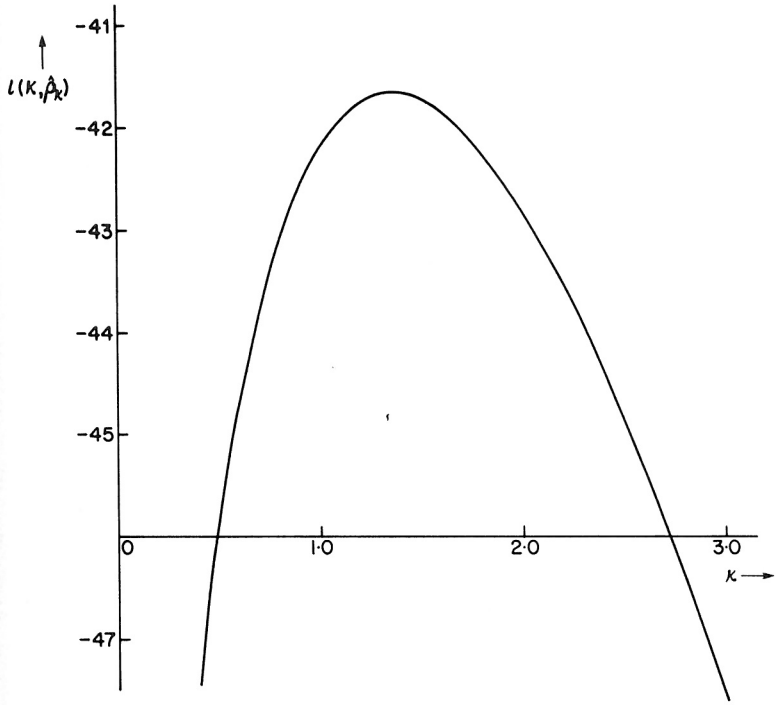

Fig. 3.3. Leukaemia data, 6-MP group. Log likelihood profile for Weibull parameter, κ, i.e. $l(\kappa, \hat{\rho}_\kappa)$ versus κ.

Example 3.2

For the leukaemia data (treated group) the joint maximum likelihood estimator of (κ, ρ) is $\hat{\kappa} = 1.35$, $\hat{\rho} = 0.030$. Fig. 3.2 shows the joint 95% confidence region for (κ, ρ) obtained from the likelihood ratio statistic W. For given κ the maximum likelihood estimator $\hat{\rho}_\kappa$ of ρ is also shown as a function of κ. Fig. 3.3 shows the likelihood profile $l(\kappa, \hat{\rho}_\kappa)$. The 95% confidence interval for κ based on W is (0.72, 2.20).

3.7 A test for exponentiality

We now derive the score test of the hypothesis $\kappa = 1$ corresponding to exponentiality. This will be a useful test against alternative hy-

potheses which specify monotone hazard functions.

The maximum likelihood estimator $\hat{\rho}_\kappa$ of ρ when $\kappa = \kappa_0 = 1$ is just

$$\hat{\rho}_{\kappa_0} = d/\sum x_i.$$

The score function is, from (3.19),

$$U_{\kappa_0} = \left[\frac{\partial}{\partial\kappa} l(\kappa, \rho)\right]_{\kappa_0, \rho = \hat{\rho}_{\kappa_0}}$$
$$= d + \sum_{\mathrm{u}} \log x_i - d\frac{\sum x_i \log x_i}{\sum x_i}.$$

The observed information matrix at $(\kappa_0, \hat{\rho}_{\kappa_0})$ has elements

$$I_{\kappa\kappa} = d + \sum(\hat{\rho}_{\kappa_0} x_i)[\log(\hat{\rho}_{\kappa_0} x_i)]^2,$$
$$I_{\kappa\rho} = \sum x_i \log(\hat{\rho}_{\kappa_0} x_i),$$
$$I_{\rho\rho} = d/\hat{\rho}_{\kappa_0}^2.$$

The inverse matrix v has leading element

$$v_{\kappa\kappa} = (I_{\kappa\kappa} - I_{\kappa\rho}^2/I_{\rho\rho})^{-1},$$

and the approximate chi-squared statistic can be constructed as at (3.9). When, as here, $p_\omega = 1$, the signed statistic

$$U_{\kappa_0}(v_{\kappa\kappa})^{1/2},$$

approximately a standard normal deviate on the null hypothesis, is to be preferred, as it indicates the direction of the departure from the null hypothesis.

Example 3.3

For the leukaemia data,

$$d = 9, \quad \sum x_i = 359, \quad \sum_{\mathrm{u}} \log x_i = 21.19,$$
$$\sum x_i \log x_i = 1077.3, \quad \sum x_i (\log x_i)^2 = 3334.8.$$

Thus, $\hat{\rho}_{\kappa_0} = 0.02507$ and $U_{\kappa_0} = 3.18$. The elements of the observed information matrix at $\kappa = \kappa_0$ are

$$I_{\kappa\kappa} = 15.79, \quad I_{\kappa\rho} = -246.0, \quad I_{\rho\rho} = 14320,$$

and $v_{\kappa\kappa} = 0.0865$. The standard normal deviate for the test of $\kappa = \kappa_0$ is $3.18 \times (0.0865)^{1/2} = 0.935$, indicating consistency with the null hypothesis of exponentiality. This agrees qualitatively with the conclusions from the likelihood ratio statistic.

In view of the small number of failures, it is not surprising that the null hypothesis cannot be rejected. The evidence, weak as it is, in favour of a monotone increasing rather than a monotone decreasing hazard, is probably due to the apparent threshold at six weeks.

Bibliographic notes, 3

Estimation from censored samples when the number of failures is predetermined and the distribution exponential was considered by Sukhatme (1937) and Epstein and Sobel (1953). Bartholomew (1957) derived asymptotic methods for the exponential distribution with fixed censoring times and also (Bartholomew, 1963) gave the exact distribution of the maximum likelihood estimator for Type I censoring. Cox (1953) suggested a chi-squared approximation in the context of a single Poisson process observed for a fixed time.

The computation and interpretation of standardized mortality ratios is discussed in most texts on medical statistics; see, for example, Bradford Hill (1977). Breslow (1977) gave the likelihood derivation. Limits on the expectation of a Poisson variable are tabulated by Fisher and Yates (1963) and Pearson and Hartley (1966). The explicit connection with chi-squared was noted by Fisher (1935) but was known much earlier, in a different context, to A. K. Erlang.

Likelihood estimation in the Weibull distribution was discussed in detail by Pike (1966) and Peto and Lee (1973). The reliability literature (Mann *et al.*, 1974) contains many alternative procedures. These are often based on order statistics and are thus applicable in the presence of a threshold parameter, as well as of censoring.

For a general introduction to large-sample likelihood theory, see Rao (1973, Chapter 6) and Cox and Hinkley (1974, Chapter 9).

Further results and exercises, 3

3.1. Suppose that failure times are exponentially distributed with

parameter ρ, and that ρ has a prior distribution of the gamma form $f_P(\rho)$ of Section 2.3. Show that the posterior distribution of ρ given the number of observed failures d and total time at risk $\sum x_i$ is also gamma, with parameters $\kappa_1 = \kappa + d$, $\rho_1 = (\kappa + d)\rho_0/(\kappa + \rho_0 \sum x_i)$. Consider the extension to the proportional hazards family, with $h_0(t)$ assumed known.

3.2. Show that, if T has a continuous cumulative distribution function F with integrated hazard $H(\,.\,)$, and if $X = \min(T, c)$, then $E[H(X)] = F(c)$.

3.3. Show that, if there are at least two distinct uncensored failure times, and each $x_i > 0$, equation (3.21) always has a unique root in $\kappa > 0$.

3.4. Show that in the absence of censoring the expected information matrix for the Weibull distribution has elements

$$-E\left(\frac{\partial^2 l}{\partial \rho^2}\right) = \frac{n\kappa^2}{\rho^2}, \qquad -E\left(\frac{\partial^2 l}{\partial \rho \partial \kappa}\right) = \frac{n\psi(2)}{\rho},$$

$$-E\left(\frac{\partial^2 l}{\partial \kappa^2}\right) = \frac{n}{\kappa^2}\{1 + \psi'(2) + [\psi(2)]^2\},$$

where

$$\psi(\alpha) = \frac{\partial}{\partial \alpha} \log \Gamma(\alpha)$$

is the digamma function.

3.5. Show that in the absence of censoring, the gamma distribution (2.13) has minimal sufficient statistic $(\sum t_i, \sum \log t_i)$ for ρ and κ, and that the maximum likelihood estimator of κ with unknown ρ is the solution of

$$\psi(\kappa) - \log \kappa - \log R = 0,$$

where R is the ratio of the geometric to the arithmetic mean of the sample.

3.6. By considering (i) the number of events and (ii) the time to the kth event in a Poisson process of rate θ observed over the unit interval,

show that

$$\bar{\theta}(k, \alpha) = \tfrac{1}{2} c^*_{2k,\alpha}.$$

3.7. For n uncensored individuals with exponentially distributed failure times and log likelihood function $n \log \rho - \rho \sum x_i$, prove that the likelihood ratio statistic W of (3.5) for testing the null hypothesis $\rho = \rho_0$ is

$$W = 2(n \log n - n \log \sum x_i - n - n \log \rho_0 + \rho_0 \sum x_i).$$

Prove that the expected value under the null hypothesis is

$$E(W) = 1 + (6n)^{-1} + O(n^{-2}),$$

so that the statistic with a Bartlett correction factor is

$$W' = W/[1 + (6n)^{-1}].$$

Examine numerically the relation between confidence limits from W, from W', and from the 'exact' solution.

CHAPTER 4

Single-sample nonparametric methods

4.1 Introduction

The methods of the previous chapter all require specification of the functional form of the distribution that failure time would have in the absence of censoring. We now discuss nonparametric techniques which require no such assumptions. As well as providing flexible alternatives to the parametric techniques, they are useful in connection with graphical assessments of goodness of fit for complex models. The term 'life table' is often used for a nonparametric estimate of a survivor function from censored data.

4.2 Product-limit estimator

We begin by assuming that the possibly improper distribution is discrete, with atoms f_j at finitely many specified points $a_1 < a_2 < \cdots < a_g$. In practice, these points are often taken to be equally spaced, $a_j = j$ in suitable time units, but this is not necessary. As described in Section 2.2, the survivor function $\mathscr{F}(t)$ may be expressed in terms of the discrete hazard function h_j as

$$\mathscr{F}(t) = \prod_{a_j < t} (1 - h_j) = \prod^{(t)} (1 - h_j),$$

where $\prod^{(t)}$, and subsequently $\sum^{(t)}$, denote product and sum over j, $a_j < t$. Thus, in terms of the h_j, the f_j may be written in the form

$$\begin{aligned} f_1 &= h_1, \quad f_2 = (1-h_1)h_2, \quad \ldots, \\ f_j &= (1-h_1)(1-h_2)\ldots(1-h_{j-1})h_j, \quad \ldots, \\ f_g &= (1-h_1)(1-h_2)\ldots(1-h_{g-1})h_g. \end{aligned} \tag{4.1}$$

The constraints $f_j \geqslant 0$, $\sum f_j \leqslant 1$ become, simply, $0 \leqslant h_j \leqslant 1$.

A nonparametric estimator of the survivor function is

$$\hat{\mathscr{F}}(t) = \prod^{(t)} (1 - \hat{h}_j), \tag{4.2}$$

where the $\hat{h}_j$ are the maximum likelihood estimators of the h_j. From equation (3.4), the log likelihood in terms of the h_j is

$$\sum_j [d_j \log h_j + (r_j - d_j)\log(1 - h_j)], \tag{4.3}$$

where r_j is the number of individuals in view at a_j, and d_j is the number who fail at a_j. It is conventional to include in r_j any individuals who are censored at a_j. Other conventions are possible and would lead to slightly different results in the sequel. Any difficulty that this causes in practice can usually be resolved by obtaining the data recorded on a finer scale.

The log likelihood (4.3) is exactly that for g independent binomials, with respectively r_j trials, d_j failures, and probability of failure h_j. It is particularly easy to maximize here, as the parameter vector is $\{h_j\}$ itself. Thus, $\hat{h}_j$ is the solution of

$$\frac{\partial l}{\partial h_j} = \frac{d_j}{h_j} - \frac{r_j - d_j}{1 - h_j} = 0,$$

i.e. $\hat{h}_j = d_j/r_j$. The corresponding estimator $\hat{\mathscr{F}}$ of the survivor function is

$$\hat{\mathscr{F}}(t) = \prod^{(t)} \left(1 - \frac{d_j}{r_j}\right), \tag{4.4}$$

obtained by substituting in (4.2).

Table 4.1 *Calculation of the product-limit estimator for the leukaemia data of Table* 1.1, 6-MP *group*

a_j	r_j	d_j	$1 - \frac{d_j}{r_j}$	$\prod_{l \leqslant j}\left(1 - \frac{d_l}{r_l}\right) = \hat{\mathscr{F}}(a_j +)$
6	21	3	0.8571	0.8571
7	17	1	0.9412	0.8067
10	15	1	0.9333	0.7529
13	12	1	0.9167	0.6902
16	11	1	0.9091	0.6275
22	7	1	0.8571	0.5378
23	6	1	0.8333	0.4482

Any term in the product which has $d_j = 0$ can be omitted without affecting (4.4). The estimate $\hat{\mathscr{F}}(t)$ is, therefore, formally independent of the selection of points a_j for which the observed number of failures is zero. Thus, $\hat{\mathscr{F}}(t)$ is a function of the data only. It can, in fact, be shown to maximize the (generalized) likelihood over the space of all distributions, although this property has only limited direct statistical implications. Usually, $\hat{\mathscr{F}}(t)$ is called the Kaplan–Meier, or product-limit estimator. Table 4.1 shows an example of its calculation, for the leukaemia data of Table 1.1.

4.3 Greenwood's formula

If the possible failure times $a_1, a_2, \ldots, a_g$ are fixed, and the censoring mechanism allows the numbers of failures d_j at each a_j to increase at the same rate as the total sample size n, then the standard large-sample theory for maximum likelihood estimators applies, and the methods outlined in Chapter 3 may be used to make inferences about the $\hat{h}_j$ or functions of them such as $\hat{\mathscr{F}}(t)$.

Thus, asymptotically, $\sqrt{n}(\hat{h}_j - h_j)$ will have a multivariate normal distribution with mean zero and a covariance matrix which can be estimated by the inverse of the observed information matrix. Here

$$\left[\frac{\partial^2 l}{\partial h_j \partial h_k}\right]_{\hat{h}} = \begin{cases} -\dfrac{r_j}{\hat{h}_j(1-\hat{h}_j)} & (j=k) \\ 0 & (j \neq k), \end{cases}$$

the same as would be obtained for k independent binomials. Since

$$\log \hat{\mathscr{F}}(t) = \sum^{(t)} \log(1-\hat{h}_j),$$

and we have just seen that the $\hat{h}_j$ are asymptotically independent, the asymptotic variance of $\log \hat{\mathscr{F}}(t)$ and hence of $\hat{\mathscr{F}}(t)$ can easily be found, for any fixed t. Thus

$$\begin{aligned} \operatorname{var}[\log \hat{\mathscr{F}}(t)] &\approx \sum^{(t)} \operatorname{var}[\log(1-\hat{h}_j)] \\ &\approx \sum^{(t)} \left(\frac{1}{1-\hat{h}_j}\right)^2 \operatorname{var}(\hat{h}_j) \\ &\approx \sum^{(t)} \left(\frac{1}{1-\hat{h}_j}\right)^2 \frac{\hat{h}_j(1-\hat{h}_j)}{r_j} \\ &= \sum^{(t)} \frac{d_j}{r_j(r_j-d_j)}, \end{aligned}$$

and

$$\operatorname{var}[\hat{\mathscr{F}}(t)] = [\hat{\mathscr{F}}(t)]^2 \sum^{(t)} \frac{d_j}{r_j(r_j-d_j)}. \tag{4.5}$$

This is known as Greenwood's formula.

Confidence limits can now be obtained via the normal approximation based either on $\hat{\mathscr{F}}(t)$ or on $\log \hat{\mathscr{F}}(t)$. Limits such as

$$\hat{\mathscr{F}}(t_0) \pm k_\alpha^* \{\operatorname{var}[\hat{\mathscr{F}}(t_0)]\}^{1/2},$$

where $\Phi(-k_\alpha^*) = \alpha$, refer to a prespecified t_0. A larger multiplier would be needed for a simultaneous confidence band for the function $\mathscr{F}(t)$ over some interval.

Some authors have suggested that (4.5) may be unstable in the tail of the distribution and have proposed an alternative, simpler estimate, namely,

$$\operatorname{var}[\hat{\mathscr{F}}(t)] = \frac{[\hat{\mathscr{F}}(t)]^2[1-\hat{\mathscr{F}}(t)]}{r(t)}. \tag{4.6}$$

A rationale for (4.6) is as follows. Given the values of $n, r(t)$ and $\hat{\mathscr{F}}(t)$, it is plausible that the least informative configuration of the data is when all the censoring in $(0, t)$ occurs at the origin, so that the censored observations contribute no information to the estimation of $\mathscr{F}(t)$. In that case, the number of uncensored observations would be $r(t)/\hat{\mathscr{F}}(t) = n_0$, and (4.6) is obtained as the variance of a single binomial proportion $\hat{\mathscr{F}}(t)$ based on n_0 trials.

As explained in Section 3.3, the dependence on the function of $\mathscr{F}(t_0)$ chosen as the basis of the normal approximation can be avoided by the use of likelihood based confidence intervals. These can be derived as follows, working from the binomial log likelihood (4.3). The maximized log likelihood when the h_j are unconstrained in the unit cube is

$$\sum \{d_j \log(d_j/r_j) + (r_j - d_j)\log[(r_j - d_j)/r_j]\}.$$

Now suppose that $\theta = \mathscr{F}(t)$ is regarded as the parameter of interest:

$$\sum^{(t)} \log(1-h_j) = \log\theta.$$

To test the null hypothesis $\theta = \theta_0$, we introduce a Lagrange multiplier ζ_0 and maximize instead of (4.3)

$$\sum d_j \log h_j + \sum (r_j - d_j)\log(1-h_j) + \sum^{(t)} \zeta_0 \log(1-h_j).$$

The maximizing values are $\tilde{h}_j$, with

$$\tilde{h}_j = \hat{h}_j \qquad (a_j \geqslant t),$$
$$\tilde{h}_j = d_j/(r_j + \zeta_0) \qquad (a_j < t),$$

where ζ_0 is determined by

$$\sum^{(t)} \log(1 - \tilde{h}_j) = \log \theta_0. \tag{4.7}$$

The statistic for testing the null hypothesis is

$$W(\theta_0) = 2\sum^{(t)} \{r_j \log[(r_j + \zeta_0)/r_j] + (r_j - d_j)\log[(r_j - d_j)/(r_j + \zeta_0 - d_j)]\}. \tag{4.8}$$

A $1 - \alpha$ confidence region for $\theta = \mathscr{F}(t)$ is now formed by taking

$$\{\theta; W(\theta) \leqslant c^*_{1\alpha}\},$$

where $c^*_{1\alpha}$ is the upper α point of the chi-squared distribution with one degree of freedom. This is best achieved by taking ζ as a new (data-dependent) parameter, obtaining from W confidence regions for ζ and then recalibrating the ζ scale in terms of θ by (4.7).

The relation with Greenwood's formula is made explicit by the expansions

$$W(\theta_0) = \zeta_0^2 \sum^{(t)} \frac{d_j}{r_j(r_j - d_j)} + O_p(1/\sqrt{n})$$

$$= [\log \hat{\mathscr{F}}(t_0) - \log \theta_0]^2 \left(-\sum^{(t)} \frac{d_j}{r_j(r_j - d_j)} \right)^{-1} + O_p(1/\sqrt{n}).$$

The difference between confidence limits derived this way and those obtained from Greenwood's formula is most pronounced in the tails of the distribution, where asymmetric limits are most natural. Thus in the data analysed in Table 4.1, the first value of the survivor function, estimated as 0.8571, has a standard error from Greenwood's formula of 0.0764, and calculation of limits via a normal approximation is hazardous, and impossible at the more extreme levels. The values of ζ corresponding to $\alpha = 0.95$ are 59 and -11.9 and the resulting limits for $\mathscr{F}(t)$ are 0.9625 and 0.6703. At the final value recorded, with estimated survivor function 0.4482, with standard error 0.1346, the values of ζ are 12.3 and -3.75, with limits for $\mathscr{F}(t)$ of 0.6965 and 0.2028, almost symmetrical and close to the values based on Greenwood's formula.

4.4 Actuarial estimator

In practice, the distribution may be continuous rather than discrete, as has been assumed so far in this chapter. In this section, we consider the estimation of a continuous distribution with piecewise-constant hazard rate, that is for $j = 1, \ldots, g$,

$$h_T(t) = \frac{f(t)}{\mathscr{F}(t)} = \rho_j \qquad (a_{j-1} \leqslant t < a_j),$$

where the a_j are prespecified, with the convention that $a_0 = 0$. This representation is not especially plausible, but it does allow the estimated hazard rate to reflect the behaviour of the data in a way that is not possible under strong parametric assumptions, while avoiding the analytic complexities of fully nonparametric estimation of continuous distributions. The use of splines would allow additional smoothness conditions to be introduced into $h(t)$, at some cost in computational complexity, and with the disadvantage of possibly allowing $h(t)$ to be negative over part of its range.

We shall assume that censoring is also governed by a random mechanism, with its own piecewise-constant hazard

$$h_c(t) = \lambda_j \qquad (a_{j-1} \leqslant t < a_j),$$

and we set $b_j = a_j - a_{j-1}$, the interval width. We consider maximum likelihood estimation of the parameters ρ_j, the λ_j being regarded as nuisance parameters, first when all the survival times and censoring times are reported exactly, and secondly when they are given in grouped form. The latter case, where only the numbers d_j of failures and m_j of censorings in each interval $[a_{j-1}, a_j)$ are recorded, out of the r_{j-1} subjects entering that interval, is the more commonly encountered.

In the first case, the log likelihood in the ρ_j may be derived without reference to the λ_j by the methods of Chapter 3. In fact, from (3.10),

$$l(\rho) = \sum_{j=1}^{g} (d_j \log \rho_j - u_j \rho_j),$$

where

$$u_j = \sum_{i=1}^{n} I(x_i; a_{j-1}, a_j),$$

is the total time at risk in the interval (a_{j-1}, a_j). Here the function I is

defined as

$$I(x;a_{j-1},a_j)=\begin{cases}0 & (x<a_{j-1}),\\ x-a_{j-1} & (a_{j-1}\leqslant x<a_j),\\ a_j & (x\geqslant a_j).\end{cases}$$

The maximum likelihood estimator of ρ_j, obtained by solving $\partial l/\partial\rho_j=0$, is

$$\hat{\rho}_j=d_j/u_j, \tag{4.9}$$

the total number of failures in $[a_{j-1}, a_j)$ divided by the total time at risk in that interval, a straightforward extension of the result (3.13) for a single exponential parameter ρ. Since $\partial^2 l/\partial\rho_j\partial\rho_k\equiv 0\ (j\neq k)$, the $\hat{\rho}_j$ are asymptotically independent.

In the second case, we must consider, separately for each interval $(a_{j-1}, a_j]$, the contribution to the joint likelihood lik(ρ, λ):

(i) from the $r_{j-1}-d_j-m_j$ subjects who survive uncensored throughout the interval;
(ii) from the d_j subjects who fail during the interval;
(iii) from the m_j subjects censored during the interval;

all conditioned on survival to the start of the interval. For clarity, we temporarily drop the subscripts $j-1$ and j. The conditional probabilities for the three events are

(i) $\exp[-b(\rho+\lambda)]$,

(ii) $$\int_0^b \rho e^{-\rho v}e^{-\lambda v}dv=\frac{\rho}{\rho+\lambda}\{1-\exp[-b(\rho+\lambda)]\},$$

(iii) $$\int_0^b \lambda e^{-\lambda v}e^{-\rho v}dv=\frac{\lambda}{\rho+\lambda}\{1-\exp[-b(\rho+\lambda)]\}.$$

The contribution to the total log likelihood arising from the interval $(a_{j-1}, a_j]$ is

$$l_j(\rho_j,\lambda_j)=-(r-d-m)b(\rho+\lambda)+d\log\left(\frac{\rho}{\rho+\lambda}\right)+m\log\left(\frac{\lambda}{\rho+\lambda}\right)+(d+m)\log\{1-\exp[-b(\rho+\lambda)]\},$$

where again the subscripts $j-1$ and j have been dropped from the right-hand side of the equation. As before, no other interval contributes to the log likelihood in (ρ_j, λ_j).

The maximum likelihood estimates $\hat{\rho}_j$ and $\hat{\lambda}_j$ can be obtained explicitly as the solutions of

$$\frac{\partial l}{\partial\rho}=-(r-d-m)b+\frac{d}{\rho}-\frac{d+m}{\rho+\lambda}+\frac{(d+m)b\exp[-b(\rho+\lambda)]}{1-\exp[-b(\rho+\lambda)]}=0,$$

$$\frac{\partial l}{\partial\lambda}=-(r-d-m)b+\frac{m}{\lambda}-\frac{d+m}{\rho+\lambda}+\frac{(d+m)b\exp[-b(\rho+\lambda)]}{1-\exp[-b(\rho+\lambda)]}=0, \tag{4.10}$$

for subtraction of the two equations gives

$$d/\rho=m/\lambda=(d+m)/(\rho+\lambda),$$

and substitution gives

$$\hat{\rho}=-\frac{d}{b(d+m)}\log\left(\frac{r-d-m}{r}\right),\qquad \hat{\lambda}=-\frac{m}{b(d+m)}\log\left(\frac{r-d-m}{r}\right). \tag{4.11}$$

If the interval width b is small, then $(d+m)/r$ will also be small and the logarithm may be expanded in series, giving

$$b\hat{\rho}=\frac{d}{r}+\frac{1}{2}\frac{d(d+m)}{r^2}+O\left(\frac{d+m}{r}\right)^3$$

$$=\frac{d}{r-\frac{1}{2}(d+m)}\left[1+O\left(\frac{d+m}{r}\right)^2\right].$$

The estimator

$$\hat{\rho}_j=\frac{d_j}{b_j[r_{j-1}-\frac{1}{2}(d_j+m_j)]} \tag{4.12}$$

is traditionally used to estimate the hazard rate ρ_j in the interval $[a_{j-1}, a_j)$. Comparison with (4.9) shows that the use of this estimator is in a sense equivalent to assuming that the deaths and censorings occur uniformly throughout the interval, when the denominator of (4.12) would equal u_j, the total time at risk in the interval.

Of generally greater interest than ρ_j itself, is the conditional probability $\exp(-b_j\rho_j)$ of survival throughout the interval in the absence of censoring. To the same order of approximation, the estimated probability of failure during the interval is

$$1-\exp(-b_j\rho_j)\approx d_j/r'_j=\tilde{q}_j,$$

say, where r'_j, called the adjusted number at risk in $[a_{j-1}, a_j)$, is given by

$$r'_j = r_{j-1} - \tfrac{1}{2}m_j.$$

The actuarial estimator of $\mathscr{F}$ is obtained by combining the $\bar{q}_j$. It has the form

$$\tilde{\mathscr{F}}(a_j) = \prod_{k \leqslant j}\left(1 - \frac{d_k}{r'_k}\right). \tag{4.13}$$

This estimator differs from the product-limit estimator of Section 4.2 evaluated at $t = a_j+$ only in the 'half-period' correction of replacing r_j by r'_j. Usually the two estimates differ little, unless the data are heavily tied. Greenwood's formula for the estimated variance of $\tilde{\mathscr{F}}(a_j)$ becomes

$$\operatorname{var}[\tilde{\mathscr{F}}(a_j)] = [\tilde{\mathscr{F}}(a_j)]^2 \sum_{k=1}^{j} \frac{d_k}{r'_k(r'_k - d_k)}.$$

4.5 Cumulative hazard estimators: goodness of fit

As mentioned in Section 2.4, plots of the hazard or cumulative hazard function are often useful in assessing the fit of a parametric family of survival distributions to a given set of data. Although minus the logarithm of the Kaplan–Meier estimator could be used to estimate the cumulative hazard, it is more usual to take

$$\hat{H}(t) = \sum^{(t)} d_j/r_j, \tag{4.14}$$

as suggested by equation (2.12). Notice that if there are no ties and no censoring, so that $\{a_1, a_2, \ldots, a_g\}$ denote the ordered failure times, then

$$\hat{H}(a_k) = e_{nk} = \sum_{j=1}^{k} \frac{1}{n+1-j},$$

the expected value of the kth-order statistic in a unit exponential sample.

The estimated variance of $\hat{H}(t) = -\log \tilde{\mathscr{F}}(t)$ was obtained en route to Greenwood's formula as

$$\operatorname{var}[\hat{H}(t)] = \sum^{(t)} \frac{d_j}{r_j(r_j - d_j)}. \tag{4.15}$$

Cumulative hazard plots have the disadvantage of tending to place too much visual emphasis on the behaviour in the tail of the

distribution, where, as is clear from (4.15), the estimate is most unstable. Moreover, the sampling errors in $\hat{H}(t)$ are not independent.

With small amounts of data, a visual appraisal of random error may be desired, by inspection of the variation among points with independent errors, preferably of constant known variance. One approximate way of achieving this is as follows. Choose a small integer k, for example $k = 4$. Let $S_1^{(k)}, S_2^{(k)}, \ldots$, be the total time at risk up to the kth failure, between the kth and $2k$th failure, etc. If ρ_j denotes the average total hazard during the period defining $S_j^{(k)}$, i.e. summed over all subjects at risk in that period, then $2\rho_j S_j^{(k)}$ is distributed as chi-squared with $2k$ degrees of freedom. It follows from standard properties of the log chi-squared distribution that

$$Z_j^{(k)} = -\log\left(\frac{S_j^{(k)}}{k}\right) - \frac{1}{2k - \frac{1}{3}}$$

has mean approximately $\log \rho_j$ and variance approximately $(k - \frac{1}{2})^{-1}$. Further, the $Z_j^{(k)}$ are approximately mutually independent. A natural plot is thus of $Z_j^{(k)}$ versus $\bar{t}_j^{(k)}$ or $\log \bar{t}_j^{(k)}$, where $\bar{t}_j^{(k)}$ is the time at the centre of the relevant interval.

4.6 Bayesian nonparametric methods

In the absence of censoring, the natural prior distribution for describing uncertainty about the atoms f_j of the discrete possibly improper survival distribution (2.6) is the Dirichlet distribution, with density proportional to

$$f_1^{\alpha_1 - 1} f_2^{\alpha_2 - 1} \ldots f_g^{\alpha_g - 1}(1 - f_1 - f_2 - \ldots - f_g)^{\alpha_{g+1} - 1} \tag{4.16}$$

for $\alpha_j \geqslant 0$ $(j = 1, 2, \ldots, g+1)$. This is the conjugate prior for the multinomial likelihood, in the sense that the posterior distribution for $\{f_j\}$, given that, of a total of n failures, d_j are observed to occur at $a_j (\sum d_j \leqslant n)$, is of the same form with parameters $\alpha'_j = \alpha_j + d_j (j = 1, \ldots, g), \alpha'_{g+1} = \alpha_{g+1} + n - \sum d_j$.

When some of the failure times are censored, however, the posterior distribution of f is no longer of the Dirichlet form. As usual, it is more convenient to transform to the (discrete) hazard function, via the transformation (4.1). The Jacobian of this transformation is

$$(1 - h_1)^{g-1}(1 - h_2)^{g-2} \ldots (1 - h_{g-1})^1,$$

and, as is evident from the definitions of f and h,

$$1 - f_1 - f_2 - \ldots - f_g = (1 - h_1)(1 - h_2)\ldots(1 - h_g),$$

both expressions equalling $\mathrm{pr}(T > a_g)$. Expressed in terms of $\{h_j\}$ the prior (4.16) becomes, apart from a proportionality constant,

$$\prod_{j=1}^{g} h_j^{\alpha_j - 1}(1 - h_j)^{\gamma_j - 1}, \tag{4.17}$$

where

$$\gamma_j = \sum_{k=j+1}^{g+1} \alpha_k. \tag{4.18}$$

The expression (4.17) has the form of a product of independent beta distributions. The degenerate form of (4.16), with $\alpha_j = 0$ for all j, corresponds to the degenerate form of (4.17), with $\alpha_j = \gamma_j = 0$, for all j. In general, however, exchangeability of the f_j, i.e. $\alpha_1 = \alpha_2 = \ldots = \alpha_g$, is not equivalent to exchangeability of the h_j.

As noted earlier, the likelihood for $\{h_j\}$, from right-censored data in which d_j failures are observed among the r_j subjects at risk at a_j, is proportional to

$$\prod_{j=1}^{g} h_j^{d_j}(1 - h_j)^{r_j - d_j}, \tag{4.19}$$

equivalent to that from a product of independent binomials. From (4.17) and (4.19) the posterior distribution of $\{h_j\}$, given the data, is also a product of independent beta distributions, with indices $\alpha_j' = \alpha_j + d_j$, $\gamma_j' = \gamma_j + r_j - d_j$. This will not in general correspond to a Dirichlet distribution over the $\{f_j\}$.

The posterior distribution of the survivor function $\mathscr{F}(a_j) = (1 - h_1)(1 - h_2)\ldots(1 - h_{j-1})$ can in theory be derived, but it does not take a simple form. The posterior moments of $\mathscr{F}(a_j)$ are easily obtained (Exercise 4.7). For the degenerate 'ignorance' prior with $\alpha_j = \gamma_j = 0$, the product-limit estimator is recovered as the posterior mean of $\mathscr{F}$.

Bibliographic notes, 4

Life tables have been used by demographers and actuaries for many years to describe and compare patterns of human mortality, often via the so-called expectation of life. The product-limit estimator appears first to have been proposed by Böhmer (1912) but the actuarial estimator itself is much older. Formula (4.5) is due to Greenwood (1926). Kaplan and Meier (1958) derived the product-limit estimator from maximum likelihood arguments; for further discussion see Johansen (1978). A key reference is Efron (1967); see Chapter 11 of this book. Anderson and Senthilselvan (1980) discuss the use of splines in the estimation of hazard functions.

At the price of some rather artificial restrictions on the form of the survival distribution, we have been able to appeal to standard large-sample likelihood theory to justify the methods. These restrictions can largely be dispensed with at the cost of increasing the complexity of the theory. The notion of weak convergence (Billingsley, 1968, 1971) provides the appropriate mathematical tool. For applications to censored data, see Breslow and Crowley (1974) and Meier (1975). Gillespie and Fisher (1979) and Hall and Wellner (1980) use the theory to derive simultaneous confidence bands. See Aalen (1976, 1978) and Gill (1980) for an approach using martingale theory. Reid (1981a) discusses influence functions for censored data, and Efron (1981) and Reid (1981b) describe some numerical methods for obtaining interval estimates of the median of a survival distribution from censored data.

A comment similar to that of the preceding paragraph applies to the Bayesian theory. Ferguson (1973) introduced Dirichlet processes, for the purpose of deriving nonparametric Bayesian estimates of a distribution function from uncensored data. This avoids the need to specify the atoms of the distribution in advance. Extensions to censored survival times are discussed by Susarla and Van Ryzin (1976), Doksum (1974), Cornfield and Detre (1977), Kalbfleisch and Mackay (1978), Kalbfleisch (1978) and Ferguson and Phadia (1979). Burridge (1981b) describes an empirical Bayes approach.

The cumulative hazard plot is due to Nelson (1969, 1972). See Cox (1979) for discussion of graphical methods for assessment of fit.

Further results and exercises, 4

4.1. Verify that in the absence of censoring, the product-limit estimator reduces to the empirical distribution function

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} I(x_i, t),$$

where $I(x, t) = 1$ if $x < t$, $I(x, t) = 0$ if $x \geqslant t$. Show also that Greenwood's formula reduces to the usual binomial variance estimate in this case.

4.2. Suppose that the censoring times are random with survivor function $K(t)$. Let $Z(t) = \sqrt{n}[\hat{\mathscr{F}}_n(t) - \mathscr{F}(t)]$, where $\hat{\mathscr{F}}_n(t)$ is the product-limit estimator of the survivor function based on n observations. Show informally that as $n \to \infty$, the covariance function $\mathrm{cov}[Z_n(s), Z_n(t)]$ $(s \leqslant t)$ converges to

$$-\mathscr{F}_0(s)\mathscr{F}_0(t)\int_{u=0}^{s} \frac{d\mathscr{F}(u)}{[\mathscr{F}(u)]^2 K(u)}.$$

[For a rigorous proof, see Breslow and Crowley, 1974.]

4.3. The restricted mean of F is estimated from the Kaplan–Meier estimator as

$$\hat{\mu}_L = \int_0^L \hat{\mathscr{F}}(t)\,dt.$$

Show from the preceding exercise that the variance of $\hat{\mu}_L$ may be estimated by

$$\mathrm{var}(\hat{\mu}_L) = \frac{1}{n}\int_0^L \frac{1}{[\hat{\mathscr{F}}(s)]^2 K(s)}\left(\int_s^L \hat{\mathscr{F}}(u)\,du\right)^2 |d\hat{\mathscr{F}}(s)|.$$

What happens as $L \to \infty$? Discuss how the variance of $\hat{\mu}_L$ and of its limit (the so-called expectation of life) may be estimated.
[Irwin, 1949; Kaplan and Meier, 1958; Meier, 1975; Reid, 1981a]

4.4. In Section 4.4, show that the exact conditional distribution of d_j given r_{j-1} and $d_j + m_j$, is binomial, with index $d_j + m_j$ and parameter $\rho/(\rho + \lambda)$.

4.5. Examine the extent to which the arguments of Section 4.4 hold if it is assumed that the hazard functions $h_T(t)$ and $h_c(t)$ in (a_{j-1}, a_j) take the form

$$h_T(t) = \rho_j h_0(t), \qquad h_c(t) = \lambda_j h_0(t),$$

for some unknown function $h_0(t)$.

4.6. Compare the second derivatives of the log likelihoods for the two cases of Section 4.4. Find an expression for the expected loss of information about ρ_j through the grouping, and evaluate this for representative values of $b\rho_j, b\lambda_j$.

[Pierce *et al.*, 1979]

4.7. Show that if h has a beta distribution, with density proportional

to $h^{\alpha-1}(1-h)^{\gamma-1}$, then its expected value is $\alpha/(\alpha+\gamma)$ if $\alpha > 0$, $\gamma > 0$, zero if $\alpha = 0, \gamma > 0$, and unity if $\alpha > 0, \gamma = 0$. Hence derive the posterior mean estimate of the survivor function in Section 4.6.

4.8. Prove via the argument used to derive Greenwood's formula that the approximate variance of the logit transform, $\log\{\hat{\mathscr{F}}(t)/[1-\hat{\mathscr{F}}(t)]\}$, is

$$[1-\hat{\mathscr{F}}(t)]^{-2}\sum^{(t)} \frac{d_j}{r_j(r_j - d_j)}.$$

Suggest how this could be used for the analysis by empirically weighted least squares of a linear logistic model for the survivor function at a single specified point t_0, the data being grouped into sets with equal or nearly equal values of the explanatory variables.

4.9. A smoothed estimate of the hazard function can be produced from uncensored data by finding a 'kernel estimate' of the density and

(i) dividing by the sample survivor function;
(ii) dividing by the integral of the smoothed density or by some other smoothed survivor function.

Compare these and comment on the advisability of a kernel that is different in different parts of the range.

How would these procedures and discussion be adapted to censored data? Under what circumstances is it desirable to use graphical and numerical procedures capable of producing smooth answers out of the most limited data as contrasted with procedures in which the intrinsic variability is shown explicitly?

CHAPTER 5

Dependence on explanatory variables: model formulation

5.1 Introduction

In the previous chapters we have discussed models and analysis for relatively simple problems involving a single distribution. When, as is often the case, two or more sets of data have to be compared, this is sometimes best done by estimating survivor functions for each set of data separately and then making a qualitative comparison, either directly or via summary statistics. More sensitive or more complicated comparisons are, however, best handled by comprehensive models in which the effect of the explanatory variables is typically represented by unknown parameters.

In the present chapter, we review some of the many possible models that may be used to represent the effect on failure time of explanatory variables. For this we suppose that for each individual there is defined a $q \times 1$ vector z of explanatory variables. The components of z may represent various features thought to affect failure time, such as

(i) treatments;
(ii) intrinsic properties of the individuals;
(iii) exogenous variables.

Further components of z may be synthesized to examine interaction effects, in a way that is broadly familiar from multiple regression analysis.

The explanatory variables may be classified also in other ways, in particular as for each individual constant or time-dependent. Some of the ideas that follow do not apply to time-dependent explanatory variables; further, for many of the statistical techniques, computation is much harder for time-dependent explanatory variables. Nevertheless, for a variety of reasons that will appear later, it is

important to accommodate time-dependent explanatory variables in the discussion.

A few outline examples of explanatory variables are as follows.

In a simple comparison of two treatments, for instance of a 'new' treatment with a 'control', we consider a binary explanatory variable equal to one for individuals receiving the treatment, and equal to zero for those receiving the control. If the treatment is specified by a dose (or stress) level, the corresponding explanatory variable is dose or log dose. If the treatment is factorial, several explanatory variables will be required with synthesized product variables to represent interactions where appropriate. In most cases such variables will be constant for each individual, but time-dependent treatment variables can arise in two rather different ways.

First, especially in some industrial reliability contexts, a time-varying stress may be applied, or some cumulative measure of total load may be judged relevant. In defining the components of the vector z, it may then be convenient to introduce functions of the whole previous history of the dose or stress process.

Secondly, it may happen that a treatment under study is not applied until some time after the time origin. Then a suitable explanatory variable may be a time-dependent binary variable that jumps from 0 to 1 at the point of application of the treatment.

Explanatory variables measuring intrinsic properties of individuals include, in a medical context, such demographic variables as sex, age on entry, and variables describing medical history before admission to the study. Other variables may define qualitative groupings of the individuals.

Considerable care is needed in introducing as explanatory variables time-dependent variables that may be influenced by the treatment variables under investigation. For example, consider the comparison of the effect on survival of two alternative treatments for the control of hypertension, blood pressure being an explanatory variable. The use of blood pressure before treatment assignment as a fixed explanatory variable is a standard device for precision improvement and interaction detection. The inclusion of blood pressure monitored after treatment assignment as an explanatory variable would address the question as to whether any difference in survival between the treatments is explained by the effect on blood pressure control. Time-dependent explanatory variables are discussed in a little more detail in Chapter 8.

Finally exogenous variables define, in particular, environmental features of the problem and may be needed also to represent groupings of the individuals corresponding to observers, sets of apparatus, etc.

It is often convenient to define the vector z of explanatory variables so that $z = 0$ corresponds to some meaningful 'standard' set of conditions, for example a control treatment. Frequently models can conveniently be developed in two parts:

(a) a model for the distribution of failure time when $z = 0$;
(b) a representation of the change induced by a nonzero z, often in terms of some parametric form.

In the description of particular models that follows, it is often convenient to start with the simplest case of the comparison of two treatments, corresponding to a single binary explanatory variable, the generalization usually being obvious. Throughout $\psi(z)$ denotes a function linking z to survival: increasing $\psi(z)$ always corresponds to increasing risk, i.e. to decreasing failure time. The symbol β is reserved for a parameter vector characterizing $\psi(z)$. Note that the functions $\psi(z)$ in two different models are not in general quantitatively comparable.

5.2 Accelerated life model

(i) *Simple form*

Suppose that there are two treatments represented by values 0 and 1 of the explanatory variable z. Let the survivor function at $z = 0$ be $\mathscr{F}_0(t)$; in the accelerated life model there is a constant ψ such that the survivor function at $z = 1$, written variously $\mathscr{F}_1(t)$ or $\mathscr{F}(t;1)$, is

$$\mathscr{F}_1(t) = \mathscr{F}_0(\psi t), \tag{5.1}$$

so that

$$f_1(t) = \psi f_0(\psi t), \qquad h_1(t) = \psi h_0(\psi t). \tag{5.2}$$

A stronger version is that any individual having survival time t under $z = 0$ would have survival time t/ψ under $z = 1$, i.e. the corresponding random variables are related by $T_1 = T_0/\psi$.

More generally, with an arbitrary constant vector z of explanatory variables, suppose that there is a function $\psi(z)$ such that the survivor

function, density and hazard are respectively

$$\begin{aligned} \mathscr{F}(t;z) &= \mathscr{F}_0[t\psi(z)], \\ f(t;z) &= f_0[t\psi(z)]\psi(z), \\ h(t;z) &= h_0[t\psi(z)]\psi(z). \end{aligned} \tag{5.3}$$

If $\mathscr{F}_0(\,.\,)$ refers to the standard conditions $z = 0$, then $\psi(0) = 1$.

A representation in terms of random variables is

$$T = T_0/\psi(z), \tag{5.4}$$

where T_0 has survivor function $\mathscr{F}_0(\,.\,)$. If $\mu_0 = E(\log T_0)$, we can write this as

$$\log T = \mu_0 - \log\psi(z) + \varepsilon, \tag{5.5}$$

where ε is a random variable of zero mean and with a distribution not depending on z.

In problems with a limited number of distinct values of z, it may be unnecessary to specify $\psi(\,.\,)$ further. In other contexts, a parametric form for $\psi(\,.\,)$ may be needed; we then write $\psi(z;\beta)$. Since $\psi(z;\beta) \geqslant 0$, $\psi(0;\beta) = 1$, a natural candidate is

$$\psi(z;\beta) = e^{\beta^{\mathrm{T}}z}, \tag{5.6}$$

where now the parameter vector β is $q \times 1$. Then (5.5) can be written

$$\log T = \mu_0 - \beta^{\mathrm{T}}z + \varepsilon, \tag{5.7}$$

a linear regression model. Note that for the comparison of two groups, with a single binary explanatory variable, we get (5.1) with $\psi = e^{\beta}$.

(ii) *Some consequences useful for model checking*

The central property of the accelerated life model can be re-expressed in various ways that can be used as a basis for testing the adequacy of the model. Thus from (5.5) the distributions of $\log T$ at various values of z differ only by translation. In particular var($\log T$) is constant.

Alternatively in the two-sample problem we can compare quantiles. We define $t_0^{(a)}, t_1^{(a)}$, for $0 < a < 1$, by

$$\begin{aligned} a &= \mathscr{F}_0(t_0^{(a)}) & t_0^{(a)} &= \mathscr{F}_0^{-1}(a), \\ a &= \mathscr{F}_1(t_1^{(a)}), & t_1^{(a)} &= \mathscr{F}_1^{-1}(a), \end{aligned} \tag{5.8}$$

Table 5.1 *Spring data of Table* 1.3; *T is thousands of cycles to failure. Mean and standard deviation of log T*

Stress (N/mm^2)	*Log T*	
	Mean	SD
950	5.11	0.214
900	5.35	0.198
850	5.84	0.172
800	6.52	0.580
750	8.41	0.735

so that under (5.1)

$$t_1^{(a)} = t_0^{(a)}/\psi,$$

i.e. the so-called Q–Q plots (quantile–quantile plots) are straight lines through the origin. We have for simplicity assumed in (5.8) that $\mathscr{F}_0(\,.\,)$ is strictly decreasing, so that the quantiles are uniquely defined.

Quite often a simple analysis will show the accelerated life model to be inadequate. For instance, with the spring data of Table 1.3, inspection suggests that the relative dispersion increases substantially as stress decreases and mean failure time increases. This is confirmed by calculating the mean and standard deviation of log T at each stress level. For this the lowest stress level at which the censoring is severe has been omitted and for the next lowest stress the single censored value has been treated as a failure, in this instance tending to underestimate the mean and standard deviation of log T. The results are summarized in Table 5.1.

The accelerated life model may hold at the three highest stress levels, but there is overwhelming evidence against it as a representation for the whole stress range. Further interpretation would be aided by information on the mode of failure, in particular as to whether this is different at the lowest stress levels. At 750 N/mm^2 the distribution is close to an exponential with a nonzero starting value.

(iii) *Time-dependent explanatory variables*

Suppose now that the explanatory variable z is time-dependent, $\{z(t)\}$, say. First, it will usually be good to define $z(t)$ so that the hazard at any particular time t depends only on the explanatory variable at that time. This may involve introducing as components of $z(t)$ integrals, sums, derivatives and differences of the explanatory variables as originally recorded.

The essence of the accelerated life model is that 'time' is contracted or expanded relative to that at $z = 0$. This suggests that for an individual characterized by $z(t)$, time $t^{(z)}$, say, evolves relative to the time $t^{(0)}$ for that individual had he been at $z = 0$ in accordance with

$$dt^{(z)}/dt^{(0)} = 1/\psi[z(t^{(z)})],$$

i.e.
$$t^{(0)} = \int_0^{t^{(z)}} \psi[z(u)]\,du = \Psi(t^{(z)}), \tag{5.9}$$

say, so that the failure times are related, instead of by (5.4), by

$$T = \Psi^{-1}(T_0).$$

Note, however, that the result of applying two such transformations to T_0 will in general depend on the order in which they are applied, so that linear combinations of such time-dependent explanatory variables will not obey the commutativity relations of ordinary arithmetic.

Hence survivor function, density and hazard are

$$\begin{aligned} \mathscr{F}[t;\{z(\,.\,)\}] &= \mathscr{F}_0[\Psi(t)], \\ f[t;\{z(\,.\,)\}] &= \psi[z(t)]f_0[\Psi(t)], \\ h[t;\{z(\,.\,)\}] &= \psi[z(t)]h_0[\Psi(t)]. \end{aligned} \tag{5.10}$$

(iv) *Generality of the time-dependent model*

The accelerated life model with time-varying explanatory variables has rarely been used in applications, so far as we know, although it would be appropriate for systems subject to nonconstant treatment variables, e.g. 'stress'. There is another sense, however, in which the use of time-varying explanatory variables converts the very special model into a very general one. Consider for simplicity the comparison of two groups and suppose that instead of a simple binary explanatory variable we introduce

$$z = \begin{cases} 0 & \text{group 0,} \\ \xi(t) & \text{group 1,} \end{cases}$$

where $\xi(t)$ is a function to be chosen and we take $\psi(z) = e^z$. Then by

(5.10) the survivor function in group 1 is

$$\mathscr{F}_0[\Lambda(t)],$$

where

$$\Lambda(t) = \int_0^t e^{\xi(u)}\,du.$$

Thus a given survivor function $\mathscr{F}_1(t)$ is reproduced by taking

$$e^{\xi(t)} = \frac{d}{dt}\mathscr{F}_0^{-1}\mathscr{F}_1(t), \tag{5.11}$$

it being assumed that the support of $\mathscr{F}_0(\,.\,)$ contains that of $\mathscr{F}_1$.

One way of producing a fairly rich family of models for the two-group problem is thus to write for $j = 0, \ldots, p$

$$z_j = \begin{cases} 0 & \text{group 0,} \\ t^j & \text{group 1,} \end{cases} \tag{5.12}$$

for some suitable value of p and then to take

$$\psi(z) = e^{\beta^{\mathrm{T}} z}, \tag{5.13}$$

where β is a $q \times 1$ parameter vector, $q = p + 1$.

In most instances this extension of the accelerated life model is a formal one without direct physical significance. Note that functions other than powers of t could be used in (5.12) and that the argument extends in principle to problems more complex than the comparison of two groups.

(v) *Several types of failure*

One possible explanation of inconsistency with the accelerated life model is the presence of several types of failure, each following an accelerated life model but with different modifying functions ψ. As z varies, the balance between the types of failure changes. Of course, if the types of failure are observed, we can construct a more detailed model; see Chapter 9. If the distinct types of failure are not observable, it may sometimes be fruitful to hypothesize a small number of failure types, to attempt to deduce something about their properties by examining simple models and then to aim at further data to see whether the hypothesized failure types have physical identity.

Suppose then that there are l failure types, and that the observed failure time T can be represented as

$$T = \min(T_1, \ldots, T_l), \tag{5.14}$$

where at $z = 0$ the T_j are independent random variables with survivor functions $\mathscr{F}_{0j}(\,.\,)$, possibly improper. Consider for simplicity the case of a single binary variable z and suppose that at $z = 1$ the survivor function of T_j is $\mathscr{F}_{0j}(\psi_j t)$. Then

$$\mathscr{F}_0(t) = \textstyle\prod_j \mathscr{F}_{0j}(t), \quad \mathscr{F}_1(t) = \prod_j \mathscr{F}_{0j}(\psi_j t)$$

and it follows easily that

$$h_0(t) = \textstyle\sum_j h_{0j}(t), \qquad h_1(t) = \sum_j \psi_j h_{0j}(\psi_j t). \tag{5.15}$$

(vi) *Parametric version*

So far the survivor function $\mathscr{F}_0(\,.\,)$ at $z = 0$ has been unspecified. If now we take $\mathscr{F}_0(\,.\,)$ to be a member of any of the parametric families discussed in Chapter 2, we obtain a special family of accelerated life models. If, further, $\psi(\,.\,)$ is specified parametrically, we have a fully parametric model. In particular, if the survivor function $\mathscr{F}_0(\,.\,)$ is log normal and $\psi(z;\beta) = e^{\beta^{\mathrm{T}} z}$, the linear model (5.7) for log T is a normal-theory one and all the usual least-squares methods are available, provided that there is no censoring.

One important special case arises when $\mathscr{F}_0(\,.\,)$ is a Weibull distribution, (2.14), with parameters (ρ_0, κ), say. Then with constant explanatory variables, it is clear that T, for specified z, also has a Weibull distribution, with parameters $(\rho_0\psi(z), \kappa)$. A special case of this is the exponential distribution, $\kappa = 1$.

The most important special case, however, is probably the log logistic, which is introduced by a rather different route in the next subsection.

(vii) *Log logistic accelerated life model*

If attention is concentrated on a particular time t_0, failure or nonfailure by time t_0 can be treated as a binary response. It is then natural to consider a linear logistic model in which

$$\log\{\mathscr{F}(t_0; z)/[1 - \mathscr{F}(t_0; z)]\} = \beta^{\mathrm{T}} z + \alpha(t_0),$$

where $\alpha(t_0)$ refers to the baseline $z = 0$.

Now suppose that such a model is required to hold for all t_0. We could, of course, make $\tilde{\beta}$ as well as $\alpha(t)$ depend on t, but the simplest representation arises if $\tilde{\beta}$ is independent of t; we then require that $\alpha(t) \to \infty$ as $t \to 0$ and $\alpha(t) \to -\infty$ as $t \to \infty$. This can most simply be achieved by taking $\alpha(t)$ to be proportional to $-\log t$. If we write $\alpha(t) = -\kappa \log(t\rho)$ and $\tilde{\beta} = -\kappa\beta$, then

$$\mathscr{F}(t;z) = \frac{1}{1+(\rho t e^{\beta^{\mathrm{T}} z})^{\kappa}}. \tag{5.16}$$

This is of precisely the accelerated life form with baseline survivor function

$$\frac{1}{1+(t\rho)^{\kappa}},$$

the log logistic distribution; see Section 2.3(viii).

This representation can be extended in various ways as a basis for testing goodness of fit; one such, as noted above, is to allow β to depend on t.

5.3 Proportional hazards model

(i) *Simple form*

A second broad family of models that has been widely used in the analysis of survival data is best specified via the hazard function. For a constant vector z of explanatory variables suppose that the hazard is

$$h(t;z) = \psi(z) h_0(t). \tag{5.17}$$

Here $h_0(\,.\,)$ is the hazard for an individual under the standard conditions, $z = 0$, and we require $\psi(0) = 1$. The survivor function and density are thus

$$[\mathscr{F}_0(t)]^{\psi(z)}, \qquad \psi(z)[\mathscr{F}_0(t)]^{\psi(z)-1} f_0(t).$$

Thus the survivor functions form the Lehmann family generated from $\mathscr{F}_0(\,.\,)$. We call (5.17) the (simple) proportional hazards model.

Note that the function $\psi(z)$, while fulfilling the same role as the $\psi(z)$ of Section 5.2, does not have precisely the same interpretation. The function $\psi(z)$ can be parameterized, as in the previous discussion, as $\psi(z;\beta)$ and in particular the most important special case is again

$$\psi(z;\beta) = e^{\beta^{\mathrm{T}} z}. \tag{5.18}$$

The reasons for considering this model are that

(a) there is a simple easily understood interpretation to the idea that the effect of, say, a treatment is to multiply the hazard by a constant factor;

(b) there is in some fields empirical evidence to support the assumption of proportionality of hazards in distinct treatment groups;

(c) censoring and the occurrence of several types of failure are relatively easily accommodated within this formulation and in particular the technical problems of statistical inference when $h_0(t)$ is arbitrary have a simple solution.

(ii) *Relation with accelerated life model*

For constant explanatory variables the question naturally arises as to when the proportional hazards model (5.17) is also an accelerated life model. For this we need there to exist a function $\chi(z)$ such that

$$[\mathscr{F}_0(t)]^{\psi(z)} = \mathscr{F}_0[t\chi(z)]. \tag{5.19}$$

Write

$$\mathscr{G}_0(\tau) = \log[-\log \mathscr{F}_0(e^{\tau})].$$

Then

$$\log\psi(z) + \mathscr{G}_0(\tau) = \mathscr{G}_0[\tau + \lambda(z)],$$

where $\lambda(z) = \kappa^{-1}\log\chi(z)$. For this to hold for all τ and for some nonzero $\lambda(z)$, i.e. nonunit $\chi(z)$, we need

$$\mathscr{G}_0(\tau) = \kappa\tau + \alpha, \qquad \lambda(z) = \log\psi(z),$$

where α, κ are constants. Thus, on writing $\rho = e^{\alpha/\kappa}$, we have that

$$\mathscr{F}_0(t) = \exp[-(\rho t)^{\kappa}].$$

That is, the Weibull distribution is the only initial distribution for which, with constant explanatory variables, the accelerated life and proportional hazards models coincide.

It follows directly from the definition of the Weibull distribution that the accelerated life model with 'scale' parameters $\psi_{\mathrm{AL}}(z)$ has survivor function and hazard

$$\exp\{-[\rho\psi_{\mathrm{AL}}(z)t]^{\kappa}\}, \qquad \kappa[\rho\psi_{\mathrm{AL}}(z)]^{\kappa} t^{\kappa-1},$$

i.e. is a proportional hazards model defined by

$$\psi_{PH}(z) = [\psi_{AL}(z)]^{\kappa}.$$

In particular if $\psi_{AL}(z) = \exp(\beta_{AL}^{T} z)$, then $\psi_{PH}(z) = \exp(\beta_{PH}^{T} z)$ with $\beta_{PH} = \kappa\beta_{AL}$.

The distinction between the proportional hazards and the accelerated life models is perhaps best seen from an artificial special case.

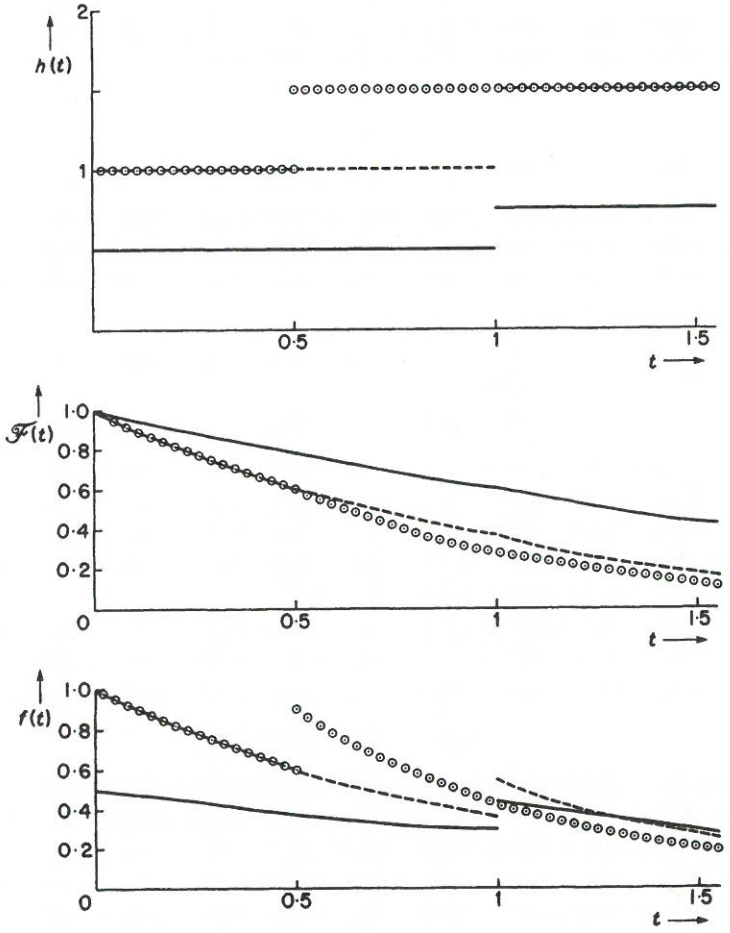

Fig. 5.1. Hazard, survivor function and density: ———, baseline; ○○○○, accelerated by a factor of 2; – – –, with hazard doubled.

Fig. 5.1 shows a baseline hazard (and survivor function and density) of step function form and the corresponding functions first accelerated by a factor of 2 and then with the hazard multiplied by 2.

(iii) Time-dependent explanatory variables

The specification (5.17) of the simple proportional hazards model

$$h(t;z) = \psi(z)h_0(t)$$

extends immediately when the explanatory variable is time-dependent. Many of the remarks of Sections 5.2(iii) and (iv) concerning time-dependent variables and the accelerated life model are relevant here also. Thus we aim to define $z(t)$ so that $h(t;z)$ depends only on $z(t)$ and not on $z(t')$, $t' \neq t$. Also time-dependent $z(t)$ may be either values of relevant subject matter variables that change with time or may be derived variables included to test the applicability of or to generalize the special model (i). Thus in the comparison of two groups, the specification (5.12) and (5.13) gives hazards in the two groups respectively of

$$h_0(t), \qquad \exp\left(\sum_{j=0}^{p} \beta_j t^j\right) h_0(t) \tag{5.20}$$

and so, at least for sufficiently large p, any two distributions with common support can be represented.

(iv) Parametric version

In the discussion so far, while the dependence on the explanatory variable has sometimes been parameterized, for instance in the form (5.18), the hazard at $z = 0, h_0(t)$, has been left arbitrary. Fully parametric models can be obtained by inserting for $h_0(t)$ the hazard for one of the families of distributions discussed in Chapter 2.

The most important special cases are probably the Weibull, including the exponential as a particular form, the Gompertz–Makeham and the log logistic.

5.4 Nonmultiplicative hazard-based model

In the simple proportional hazards model (5.17),

$$h(t;z) = \psi(z)h_0(t),$$

with constant explanatory variables, the hazard functions at different levels of z are proportional. It is possible, however, that the hazard functions are parallel rather than proportional or that, more generally, they can be made parallel by a nonlinear transformation of h. Thus we may consider the representation

$$h(t;z) = \phi(z) + h_0(t) \tag{5.21}$$

or

$$h^{(\lambda)}(t;z) = \phi(z) + h_0^{(\lambda)}(t), \tag{5.22}$$

where, for example,

$$h^{(\lambda)}(t;z) = \begin{cases} [h(t;z)]^{\lambda} & (\lambda \neq 0), \\ \log h(t;z) & (\lambda = 0). \end{cases} \tag{5.23}$$

In (5.21) and (5.22), $\phi(0) = 0$ and $\phi(z)$ is constrained so that the right-hand side is nonnegative. Parametric versions are obtained by, for instance, writing $\phi(z) = \beta^{\mathrm{T}} z$ and by taking one of the parametric families of Chapter 2 for $h_0(t)$. With $\lambda = 0$ we recover the proportional hazards model.

The choice between these models will normally be an empirical matter, involving either the formal estimation of λ in (5.23) or, more commonly, the inspection of estimated hazard functions.

Some light is thrown on the alternative formulations by (5.14). That is, if

$$T = \min(T_1, \ldots, T_l), \tag{5.24}$$

where $T_1, \ldots, T_l$ are independent random variables, then

$$h(t) = \sum_{j=1}^{l} h_j(t), \tag{5.25}$$

where h_j is the hazard for T_j.

Now suppose that the difference between two treatments lay in the elimination of some of the T_j, i.e. some of the sources of failure. Then the difference between the two hazards would be

$$\sum{}' h_j(t),$$

where the sum is over the eliminated T_j. If the eliminated T_j were to have constant hazards over the range of interest, then an additive model (5.21) would result. If, on the other hand, the eliminated T_j were in some rough sense a large random sample from the l, then a proportional hazards model would result.

5.5 Transferred origin model

A possible relation between the hazard functions in two groups is that one is the translation in time of the other. That is, for some constant Δ the hazard functions are

$$h_0(t), \qquad h_0(t + \Delta),$$

with an obvious extension to $h_0[t + \Delta(z)]$ for the hazard corresponding to explanatory variable z.

Clearly $h_0(t + \Delta) = \psi h_0(t)$ for all t if and only if $h_0(t)$ is exponential, the Gompertz form, so that the transferred origin model is equivalent to the simple proportional hazards model only in this case. The model is directly meaningful only for values of t for which both t and $t + \Delta$ are positive.

5.6 Accelerated onset model

In some contexts the effect of a treatment may be to accelerate (or retard) the onset of failure in some individuals, leaving the remainder unchanged. Thus for accelerated onset an individual who has survived an appreciable time will have the same hazard function in the two groups.

One way of representing this is to postulate a mixture of two (or more) types of individuals with survivor functions $\mathscr{F}_{01}(t)$ and $\mathscr{F}_{02}(t)$, say. The distinction between the types is not directly observable. If the proportional hazards model applies to the first type, we have for the two survivor functions

$$\begin{aligned} \mathscr{F}_0(t) &= \theta \mathscr{F}_{01}(t) + (1 - \theta)\mathscr{F}_{02}(t), \\ \mathscr{F}_1(t) &= \theta [\mathscr{F}_{01}(t)]^{\psi} + (1 - \theta)\mathscr{F}_{02}(t), \end{aligned} \tag{5.26}$$

the corresponding hazard functions being

$$\begin{aligned} h_0(t) &= [\theta f_{01}(t) + (1 - \theta) f_{02}(t)] \\ &\quad \times [\theta \mathscr{F}_{01}(t) + (1 - \theta)\mathscr{F}_{02}(t)]^{-1}, \\ h_1(t) &= \{\theta \psi f_{01}(t)[\mathscr{F}_{01}(t)]^{\psi - 1} + (1 - \theta) f_{02}(t)\} \\ &\quad \times \{\theta [\mathscr{F}_{01}(t)]^{\psi} + (1 - \theta)\mathscr{F}_{02}(t)\}^{-1}. \end{aligned}$$

In (5.26) we require $\mathscr{F}_{01}(t)$ and $[\mathscr{F}_{01}(t)]^{\psi}$ to converge to zero faster than $\mathscr{F}_{02}(t)$, as $t \to \infty$.

While under suitable circumstances such a model could be fitted

following parameterization of the component survivor functions, it is likely that a large amount of high-quality data would be needed unless, perhaps, the component survivor functions are drastically restricted, for example by supposing one or both of them to be exponential. In the latter case the survivor functions are, say,

$$\theta e^{-\rho' t} + (1-\theta)e^{-\rho'' t}, \qquad \theta e^{-\psi\rho' t} + (1-\theta)e^{-\rho'' t}, \tag{5.27}$$

where $0 < \theta < 1$ and if ρ' and ρ'' are appreciably different, plots of log survivor function versus t will expose what is happening.

An alternative rather more empirical approach for accelerated onset is to set out an initial perturbation of the hazard $h_0(t)$ with total integral zero, so that once this perturbation has been survived, the survivor functions are identical. A simple version is to propose hazard functions for the two groups of

$$h_0(t), \qquad h_0(t) + \alpha e^{-\gamma t}(t - 1/\gamma). \tag{5.28}$$

For $\gamma t \gg 1$, hazard and survivor functions are identical. With a system more complex than the comparison of two groups, one or both of α and γ could be taken as functions of z.

Often it would be sensible to combine such an effect of accelerated onset with some modification of the whole hazard, e.g. by considering as the two hazards

$$h_0(t), \qquad \psi h_0(t) + \alpha e^{-\gamma t}(t - 1/\gamma). \tag{5.29}$$

The special cases $\psi = 1, \alpha = 0$; $\psi = 1, \alpha \neq 0$; $\psi \neq 1, \alpha = 0$ all have direct interpretations.

5.7 Treatments with a transient effect

A situation rather similar to that of Section 5.6 arises when a treatment effect, or more generally influence of explanatory variables, is likely to be transient, i.e. applying only for small values of t. In the absence of a specific model based on a theoretical analysis of the system, such transient treatment effects can be represented empirically in various ways. Perhaps the simplest is to take, in the case of two groups, hazards

$$h_0(t), \qquad \exp(\beta_1 + \beta_2 e^{-\gamma t}) h_0(t). \tag{5.30}$$

If $\beta_2 = 0$ we recover the simple proportional hazards model, whereas if $\beta_1 = 0, \gamma > 0$, the hazard for individuals in group 1 eventually reverts

to that for group 0 and, in an appropriate context, individuals who survive until their hazard is very close to $h_0(t)$ can be regarded as 'ultimately cured'.

5.8 Discussion

In the previous sections we have outlined what may seem a bewildering variety both of broad types of model and of minor variants. Each is expressed in terms of a survivor function $\mathscr{F}_0(t)$ holding under some standard conditions $z = 0$ and of a modifying factor specified in various ways. The function $\mathscr{F}_0(t)$ may itself be specified parametrically; see Chapter 2.

If interest lies in the qualitative effect on failure time of various explanatory variables, choice of a model may not be critical. On the other hand, if interest lies in relatively subtle aspects of the dependence, or in discrimination between alternative specifications, a large amount of high-quality data is likely to be necessary. In some cases, especially in the physical sciences, there may be some special theory to guide the choice of model.

Bibliographic notes, 5

The parametric accelerated life model has several links with models widely used in other types of application, especially when $\psi(z;\beta) = e^{\beta^T z}$. For it is then a log linear model, i.e. a linear regression model for log T, although in general with nonnormal error. For other forms of ψ, a nonlinear regression model results. The central assumption (5.4) for random variables is in the spirit of Fraser's (1968, 1979) structural inference.

Q–Q plots are described by Wilk and Gnanadesikan (1968) and used for nonparametric inference in the two-sample problem by Doksum and Sievers (1976).

In association with the exponential distribution, regression models of various forms were considered by Feigl and Zelen (1965), Cox and Snell (1968) and in a rather general log linear form by Glasser (1967). The proportional hazards model in a general form with time-dependent explanatory variables was given by Cox (1972) and the relation with the accelerated life model stated without proof; see Oakes (1981) for a review of subsequent developments. Non-multiplicative hazard-based models are studied by Aranda-Ordaz (1980).

While the qualitative ideas of accelerated onset and transient effects are well known, explicit models do not appear to have been formulated before.

Further results and exercises, 5

5.1. Find from (5.11) the time-dependent accelerated life version of the two-sample problem with $\mathscr{F}_1(t) = \mathscr{F}_0[(\rho t)^\kappa]$. Suggest how to find the function $\xi(t)$ 'nonparametrically' from the two survivor functions. Formulate three-sample, or more generally multi-sample, versions of the same problems.

5.2. Prove that a two-term representation (5.15) of two hazard functions $h_0(t)$ and $h_1(t)$ in terms of two underlying unobserved types of failure following accelerated life models is always possible if h_0 and h_1 are both of the Gompertz–Makeham form.

5.3. Examine the accelerated onset model (5.26) and (5.27) in which both component survivor functions are exponential. How could consistency with (5.26) be checked and parameters estimated graphically?

5.4. Suppose that, in the two-sample problem, the survivor functions satisfy the proportional odds model

$$\frac{\mathscr{F}_1(t)}{1-\mathscr{F}_1(t)} = \psi \frac{\mathscr{F}_0(t)}{1-\mathscr{F}_0(t)}.$$

(a) Show that the ratio of hazard functions $h_1(t)/h_0(t) \to 1$ as $t \to \infty$.

(b) Suppose that there is no censoring, and that survival is dichotomized at a point t_0, so that survivors beyond t_0 are counted as 'positives', failures before t_0 as 'negatives'. Show that the corresponding 2×2 contingency table has odds ratio ψ. How may this result be used (i) to estimate ψ and (ii) to assess graphically the goodness of fit?

(c) Show that if $\mathscr{F}_0(t)$ has the log logistic form then so does $\mathscr{F}_1(t)$.

(d) Investigate the general formulation where ψ is a function of explanatory variables, $\psi = \psi(z)$.

[Plackett, 1965; Clayton, 1974; McCullagh, 1980; Bennett, 1983]

5.5. Consider two groups with hazards $h_0(t), h_1(t)$ and survivor functions $\mathscr{F}_0(t)$, $\mathscr{F}_1(t)$. Examine the connection between the crossings of the two hazard functions and the crossing of the two survivor functions, noting in particular the following:

(a) if the continuous survivor functions cross once, the hazards must cross at least once;

(b) it is possible that $\mathscr{F}_1(t) > \mathscr{F}_0(t)$ for all $t > 0$ and yet for the hazards to cross very often;

(c) if both groups consist of a mixture of 'short' survivors with possibly different distributions in the two groups and of 'long' survivors with the same distribution in the two groups, then the hazards, but not in general the survivor functions, are equal for large t.

5.6. Suppose that in comparing two groups the failure times in one group are almost all less than t' whereas those in the other group almost all exceed t'. What can be concluded about the two hazard functions; can the proportional hazards model be refuted? Is there any special difficulty in checking whether the accelerated life model holds?

5.7. For the accelerated life log logistic model of Section 5.2(vii) and Exercise 5.4, compare in the uncensored case the asymptotic efficiencies of full maximum likelihood estimation of the parameter β and that of maximum likelihood estimation of the logistic binary response model formed by considering only survival or failure at a fixed time t_0.

5.8. Suppose that $h_0(t)$ and $h_1(t)$ are continuous bounded hazard functions satisfying, for some ψ, $0 < \psi < 1$, and all t,

$$h_0(t) > h_1(t) > \psi h_0(\psi t).$$

Show that a representation of the form (5.15) with $l = 2$ in terms of nonnegative component hazard functions $h_{01}(t)$ and $h_{02}(t)$ is possible, and determine these functions in terms of $h_0(t)$ and $h_1(t)$.

5.9. By using an argument like that of Section 5.3(ii), show that the proportional odds model

$$\frac{\mathscr{F}_1(t)}{1-\mathscr{F}_1(t)} = \psi \frac{\mathscr{F}_0(t)}{1-\mathscr{F}_0(t)}$$

is an accelerated life model if and only if $\mathscr{F}_0(t)$ has the log logistic form.

Develop a similar characterization of the log normal distribution in terms of the probit model

$$\Phi^{-1}\{\mathscr{F}_1(t)\} = \psi + \Phi^{-1}\{\mathscr{F}_0(t)\},$$

where $\Phi(\,.\,)$ is the standard normal distribution function.