REGRESSION MODELS AND THE ANALYSIS OF CENSORED SURVIVAL DATA

by

Richard Kay B.Sc., M.Sc.

Thesis submitted for the Ph.D. degree of London University.

1976

London School of Hygiene and Tropical Medicine

ABSTRACT

A problem which frequently arises in the analysis of censored survival data in medical statistics is that of obtaining treatment comparisons while adjusting for and evaluating the effects of many uncontrolled independent variables. Recent interest in this area has centred around the use of non-linear regression models which assume that independent variables affect the hazard function in a multiplicative way. A non-parametric and several parametric models of this type have been proposed in the literature. These models, with extensions which stratify according to the independent variables to incorporate situations where the proportional hazards assumption is violated, are discussed and associated methods of inference presented. Results, in the single independent variable case, concerning the efficiency of inferences based on the non-parametric model when the true model for survival time is of the exponential parametric form are extended to incorporate the within strata models and the case of two independent variables. The effect of censoring on these efficiency results is assessed using compurter simulation. The important question of assessing goodness of fit to the data is considered and finally an example with data arising from a clinical trial is used to illustrate the techniques discussed in the study.

CONTENTS

ACKNOWLEDGEMENTS

I should like to thank my supervisor, Professor P. Armitage for invaluable guidance and encouragement throughout the course of this work.

I am also grateful to Dr N. Tygstrup and Dr. K. Juhl for permission to publish the data contained in the final chapter and for helpful comments during its analysis.

I am very much indebted to the 'Medical Research Council' who financed me during the first two years of this study.

Finally, I would like to thank Mrs K. M. Heaton and other secretaries in the Mathematics department at the University of Salford who so accurately typed the thesis.

Chapter 1

PRELIMINARIES

§1.1: General

A Medical Problem

In medical statistics one is often required to assess the effect of independent (i.e. explanatory) variables such as age, treatment, sex etc., on the time to death, called survival time, of individuals. For ease of exposition the term death will be adopted throughout, although one could equally well consider time to any well-defined event, such as response or relief of symptoms. The methods to be described in this work also have applications outside the field of medicine, for example in industrial life testing experiments. However, for convenience of terminology, we shall discuss the techniques within the medical framework.

For various reasons, data resulting from this type of investigation is frequently incomplete, in the sense that observations on survival time are not known exactly for all individuals. This may be due to limitations on the length of study or death from a cause other than that under investigation and so on. An incomplete observation of this type is termed a loss or censored observation. In §1.2. different types of censoring will be discussed in some detail.

Definitions

It is convenient at this stage to introduce some definitions. Let T be a random variable representing survival time. The survivor function of T, denoted by $\mathcal{F}(t)$, is defined by

$$\bar{F}(t) = p(T \geq t), \qquad t > 0.$$

The distribution function of T is then

$$F(t) = 1 - \bar{F}(t) = p(T < t), \quad t > 0.$$

The hazard function (sometimes called age-specific failure rate or force of mortality) $\lambda(t)$ of T is defined by

$$\lambda(t) = \lim_{dt \to 0+} \frac{p(t \leq T < t + dt \mid T \geq t)}{dt}, \quad t > 0,$$

with corresponding cumulative hazard function given by

$$\Lambda(t) = \int_0^t \lambda(u)du.$$

It follows directly that $\lambda(t) = -\bar{F}'(t)/\bar{F}(t)$ and $\Lambda(t) = -\log\bar{F}(t)$. ($\bar{F}'(t)$ denotes the first derivative of $\bar{F}(t)$ w.r.t. t).

A family of survivor functions $\{\bar{F}_\alpha(t); \alpha \in I\}$ in which any two such functions are connected by the relationship $\bar{F}_1(t) = \{\bar{F}_2(t)\}^\theta$ for some $\theta \in (0,\infty)$, is called a Lehmann family of survivor functions. Alternatively, if $\lambda_1(t)$, $\lambda_2(t)$ are the hazard functions corresponding to $\bar{F}_1(t)$ and $\bar{F}_2(t)$ respectively, this relationship can be written equivalently as $\lambda_1(t) = \theta \lambda_2(t)$.

Throughout this work the assumption that independent variables have multiplicative effects upon the hazard function will usually be made. This assumption is incorporated in most of the models for survival data discussed in the literature.

Example I

Mainly in the earlier part of this work, the following data reported by Freireich et. al. (1963) will be used to illustrate some of the techniques discussed. A trial was conducted to compare the effects of 6-mercaptopurine (6-MP) and a placebo on the maintenance of remissions in acute leukaemia. One year after the start of the study, the lengths of remission, in weeks, were recorded and are given in table 1.1.

Table 1.1. Data from trial comparing 6-MP and a Placebo on the maintenance of remissions in acute leukaemia. Units are weeks.

6-MP :	6	6	6	6◆	7	9◆	10	10◆	11◆	13	16	17◆	19◆	20◆	
	22	23	25◆	32◆	32◆	34◆	35◆								
Placebo :	1	1	2	2	3	4	4	5	5	8	8	8	8	11	11
	12	12	15	17	22	23									

◆ denotes a censored observation.

1.2. Types of censoring

Formal Definition

Let T be a random variable representing survival time. If the only information regarding an observation t on T is that $t \in (x_1, x_2)$, $x_2 > x_1 \geq 0$ then the observation t is said to be interval-censored in (x_1, x_2). If $x_2 = \infty$, t is said to be right-censored at x_1, while if $x_1 = 0$, t is said to be left-censored at x_2. An observation that is not censored is said to be exact or uncensored.

In medical investigations the most common type of censoring is right-censoring, so that this work will be concerned mainly with situations in which observations are subject to right-censoring only, and the term right-censored will be abbreviated to censored for convenience. Any deviation from this will be indicated where appropriate.

A special case of right-censoring occurs when all censored observations are equal to a constant t^* which is greater than the largest exact survival time. This is termed extreme censoring. It can occur, for example, in a study in which all individuals are put on test at the same point in time, and at a time t^* after the start of the study the data is recorded as time to death or right-censored at t^*.

It will be necessary throughout to assume that censoring and death are determined by independent mechanisms. Although this is invariably realistic in medical applications, it may not be so elsewhere. For example, in industrial life-testing a situation may arise in which items are removed from test due to diminishing performance, prior to failure.

Censoring Mechanism

To simplify the mathematical development in certain situations it will be convenient to make assumptions about the underlying mechanism producing censored observations.

A convenient such assumption is the random censorship model introduced by Gilbert (1962) and used subsequently by several authors, notably Breslow (1970) and Breslow and Crowley (1974). Suppose $T_1, \ldots, T_n$ are independent random variables, with T_i representing the survival time for the i^{th} individual having distribution function $F_{T_i}(.)$. Under the random censorship model it is assumed that there exist independent and identically distributed random variables Y_i, $i=1,\ldots,n$ with common distribution function $H_Y(y)$, which represent the periods of observation for the different individuals. $F_Y(y)$ is termed the censoring distribution. Thus one observes

$$T_i^* = \min (T_i, Y_i)$$

$$\delta_i = \begin{cases} 0 & \text{if } T_i^* = Y_i \\ 1 & \text{if } T_i^* = T_i \end{cases}$$

The distribution function of T_i^* is given by

$$F_{T_i^*}(t) = 1 - (1 - F_{T_i}(t))\ (1 - H_Y(t)) \qquad i=1,\ldots,n$$

Alternatively, Mantel and Myers (1971) suggest that for each individual i, there exists a maximum observable time Y_i, the time between entry into the study and termination of the trial for the purposes of data analysis. Thus for individuals who die, $T_i < Y_i$ while for survivors $T_i = Y_i$. Information regarding the values of the Y_i's however may not always be available. This model will be referred to as the fixed observation time model.

<u>Censoring Pattern</u>

To conclude this section the idea of censoring pattern, which has been introduced by Gehan (1965a), will be discussed.

Suppose that, of the n observations on survival time, c are censored and n-c are exact. Let

$$t_{(1)} < t_{(2)} < \ldots < t_{(k)}$$

be the distinct ordered exact observations and put

$$d_i = \begin{cases} 0 & i=0 \\ \text{number of exact observations equal to } t_{(i)} & i=1,2,\ldots,k, \end{cases}$$

$$\ell_i = \text{number of censored observations} \geq t_{(i)} \text{ and } < t_{(i+1)} \quad i=0,1,\ldots,k,$$

$t_{(0)} = 0$ and $t_{(k+1)} = \infty$.

Then $(d_i, \ell_i;\ i=0,\ldots,k)$ is called the observed censoring pattern and may be represented diagrammatically as follows:

Note that $\sum_{i=0}^{k} d_i = n-c$ and $\sum_{i=0}^{k} \ell_i = c$.

II.3. Estimation of a survivor function

Introduction

If the individuals present in a study can reasonably be split into a finite number of relatively homogeneous subsets according to the independent variables, a useful visual indication of their effect on survival experience may be obtained by plotting estimated survivor functions within each subset.

The Product Limit estimate

Kaplan and Meier (1958) have proposed the following method for estimating a survivor function. Suppose that independent observations are taken on the survival time T which has survivor function $\bar{F}(t)$. Let c represent the number of censored observations and let n-c be the number uncensored. If

$$t_{(1)} < t_{(2)} < \ldots < t_{(k)} \qquad k \leq n$$

denote the distant ordered uncensored observations and $\{d_i, l_i, i=0,1,\ldots,k\}$ the observed censoring pattern then the Product Limit (PL) estimate P(t) of $\bar{F}(t)$ is defined by

$$P(t) = \begin{cases} 1 & t < t_{(1)} \\ \prod_{t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) & t \geq t_{(1)} \end{cases}$$

where $n_i = \sum_{j=i}^{k} (l_j + d_j)$ is the number of observations (censored and uncensored) not less than $t_{(i)}$.

Kaplan and Meier obtain P(t) by assuming that λ(t) is zero except at points where deaths occur and show that P(t) is the maximum likelihood estimator of $\bar{F}(t)$ in the family of all possible survivor functions. These authors provide expressions for computing the variance of P(t) and estimating the mean survival time using P(t).

Further theoretical justification of the form of P(t) has been provided by Breslow and Crowley (1974) under the random censorship model. The adaptation of the PL estimate for interval-censored data has been discussed by Peto (1973).

Altshuler's estimate

An alternative method of estimating a survivor function has been proposed by Altshuler (1970) who suggests that a natural estimator of $\Lambda(t)$, the cumulative hazard function, is given by

$$e(t) = \sum_{t_{(i)} \leq t} d_i/m_i$$

The resulting estimator of $\bar{F}(t)$ is then

$$\hat{F}(t) = \exp\{-e(t)\} .$$

Altshuler shows that $-e(t)$ is a consistent estimator of $\log \bar{F}(t)$. Taking the natural logarithm of P(t) it follows that

$$\log P(t) = \sum_{t_{(i)} \leq t} \log (1 - d_i/m_i)$$

$$= - \sum_{t_{(i)} \leq t} \left\{ \frac{d_i}{m_i} + \frac{1}{2} \left(\frac{d_i}{m_i}\right)^2 + \dots \right\}$$

so that for $m_k \cdot d_k \geq 1$,

$$\log P(t) = - \sum_{t_{(i)} \leq t} d_i/m_i = - e(t)$$

Approximating λ(t) as a step function

Several authors (Kalbfleisch and Prentice (1973), Breslow (1974)) have considered the estimation of a survivor function in a more general context, which will be discussed later in §4.3. It is useful however to outline their methods as they apply to a single group of observations.

Kalbfleisch and Prentice begin by approximating λ(t) as a step function

$$\lambda(t) = \lambda_i \qquad t \in [b_{i-1}, b_i) = I_i \quad i=1,\ldots,r$$

where $b_0 = 0 < b_1 < b_2 < \ldots . < b_{r-1} < b_r = \infty$ define a suitable subdivision of the time scale. The survivor function and p.d.f. of T are then given respectively by

$$\bar{F}(t) = \begin{cases} \exp(-\lambda_1 t) & t \in I_1 \\ \exp\{-\lambda_i(t - b_{i-1}) - \sum_{j=1}^{i-1} \lambda_j (b_j - b_{j-1})\} & t \in I_i \end{cases} \quad 2.1.$$

$$i=2,\ldots,r.$$

and

$$f(t) = \begin{cases} \lambda_1 \exp(-\lambda_1 t) & t \in I_1 \\ \lambda_i \exp\{-\lambda_i(t - b_{i-1}) - \sum_{j=1}^{i-1} \lambda_j (b_j - b_{j-1})\} & t \in I_i \end{cases}$$

$$i=2,\ldots,r.$$

If t_{is}, $s=1,2,\ldots$ denote the observations on survival time in I_i, of which p_i are exact and q_i censored, the log likelihood function of $\underline{\lambda}' = (\lambda_1,\ldots,\lambda_r)$ is

$$L(\underline{\lambda}) = \sum_{i=1}^{r} p_i \log \lambda_i - \sum_{i=1}^{r} \lambda_i \sum_{s=1}^{p_i+q_i} (t_{is} - b_{i-1})$$

$$- \sum_{i=1}^{r-1} \lambda_i (b_i - b_{i-1}) \sum_{s=i+1}^{r} (p_s + q_s)$$

from which it follows that, for $i=1,2,\ldots,r$, $\hat{\lambda}_i$ the maximum likelihood estimate of λ_i is given by

$$\hat{\lambda}_i = \begin{cases} p_i \left\{ \sum_{s=1}^{p_i+q_i} (t_{is} - b_{i-1}) + (b_i - b_{i-1}) \sum_{j=i+1}^{r} (p_j + q_j) \right\}^{-1} & i=1,\ldots,r-1 \\ p_r \left\{ \sum_{s=1}^{p_r+q_r} (t_{rs} - b_{r-1}) \right\}^{-1} & i=r \end{cases}$$

$\hat{\mathcal{F}}_1(t)$, the resulting estimated survivor function is obtained on replacing $\underline{\lambda}$ by $\hat{\underline{\lambda}}$ in 1.1. Log $\hat{\mathcal{F}}_1(t)$ is a connected series of straight lines with $\hat{\mathcal{F}}_1(\infty) = 0$ if $\hat{\lambda}_r \neq 0$.

Alternatively, Breslow chooses intervals

$$I_i = [t_{(i-1)}, t_{(i)}) \quad i=1,\ldots,k \quad t_{(o)}=0, \quad t_{(k+1)}=\infty$$

and treats all censorings occurring in I_i as having occurred at $t_{(i-1)}$. Estimation of $\underline{\lambda}$ proceeds as above and

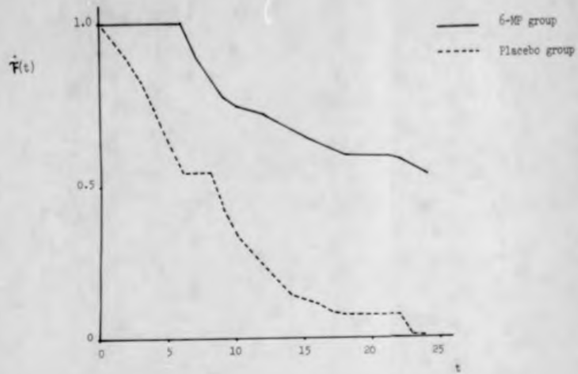
$$\hat{\lambda}_i = \begin{cases} d_{i-1}\,(w_i(t_{(i)}-t_{(i-1)}))^{-1} & i=1,2,\ldots,k \\ \infty & i=k+1. \end{cases}$$

Breslow's method disregards information regarding the exact censoring times and censorings occurring prior to the first death are ignored. This loss of information could be severe if the sample is heavily censored, particularly in situations where several large censoring times exceed the largest uncensored observation. The resulting estimate $\hat{\mathcal{F}}_2(t)$ of $\mathcal{F}(t)$ is zero for $t > t_{(k)}$.

Example

The Kalbfleisch and Prentice approach, with interval width's of 3 units for the 6-MP group and 2 units for the Placebo group, has been used to produce estimated survivor functions in figure 1.1 for the data of example I. It is clear that 6-MP is the superior treatment.

Fig 1.1. Estimated survivor functions for 2 groups in example I (Estimating $\lambda(t)$ as a step function).



Chapter 2

THE TWO-GROUP PROBLEM

[illegible]. Introduction

Single indicator variable

The statistical problem considered in this chapter is one in which independent variation is simply a single binary indicator variable which divides the sample into two groups. Individuals within each group are assumed homogeneous in the sense that group membership is the only factor thought to affect survival. Several approaches to the analysis of the two group situation will be considered and procedures for extension to more than two groups will be indicated where appropriate. Example I will be used throughout the chapter to illustrate how the techniques may be applied.

Notation

Although this chapter mainly considers the two group case it will be convenient to present the notation to be used in the more general K (≥ 2) group situation. For $j=1,\dots,K$ let T_j be a random variable representing survival time with distribution function $F_{T_j}(.)$, observations t_{ji}, $i=1,\dots,n_j$ and corresponding indicators δ_{ji}, $i=1,\dots,n_j$ where

$$\delta_{ji} = \begin{cases} 1 & \text{if } t_{ji} \text{ is a death} \\ 0 & \text{if } t_{ji} \text{ is a censoring.} \end{cases}$$

In addition, let $n = \sum_{j=1}^{K} n_j$ and denote the distinct ordered uncensored observations in the combined sample by $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ with corresponding observed censoring pattern $(d_i, [illegible]; i=0,1,\dots,k)$. For $i=0,1,\dots,k$ let

$$d_{ji} = \begin{cases} 0 & i=0 \\ \text{number of group } j \text{ uncensored observations equal to } t_{(i)} & i=1,\ldots,k. \end{cases}$$

ℓ_{ji} = number of group j censored observations in $[t_{(i)}, t_{(i+1)})$

$$m_{ji} = \sum_{p=i}^{k} (\ell_{jp} + d_{jp}), \quad m_i = \sum_{j=1}^{K} m_{ji} = \sum_{j=i}^{k} (d_j + \ell_j)$$

Comparison of tests

The relative merit of the tests of this chapter will be assessed in two ways. Firstly, in large samples, the criterion of asymptotic relative efficiency (A.R.E) will be used (Kendall and Stuart (1973), p.276). A.R.E. measures the limiting ratio of sample sizes required by two tests to produce the same power for a sequence of parameter values which approach the null value being tested. Secondly, small sample power will be investigated by the following Monte Carlo procedures proposed by Gehan and Thomas (1969) and used subsequently by Lee, Desu and Gehan(1975). The clinical trial situation is simulated by these authors by assuming that individuals in each group enter the trial at a constant rate in the interval $(0, T^*)$. (Note that this corresponds to the random censorship model with

$$H_{Y_1}(y) = H_{Y_2}(y) = \frac{y}{T^*} \quad y \in (0,T^*)) .$$

For i=1,2, group i individuals fail according to the Weibull distribution (see §2.2) with p.d.f.

$$f_i(t) = \lambda_i \, \alpha \, t^{\alpha-1} \exp(-\lambda_i t^{\alpha}), \qquad t > 0.$$

In the two group case, the null hypothesis of interest is H_0: $\lambda_1 = \lambda_2$ against the one-sided alternative H_1: $\lambda_2 > \lambda_1$.

Samples will be generated according to the following plan:

α	λ_1	λ_2	T^*
1.25	0.91498	0.91498	2.17283
		1.14918	1.98527
		1.39338	1.84197
		1.64650	1.72807
		1.90766	1.63490
		2.17620	1.55695
1.00	1.00	1.00	2.00
		1.20	1.79341
		1.40	1.64301
		1.60	1.52816
		1.80	1.43714
		2.00	1.36319
0.75	1.13982	1.13982	1.63007
		1.30684	1.48806
		1.46700	1.37775
		1.62153	1.28887
		1.77129	1.21529
		1.91694	1.15308

If X has a Weibull distribution with parameters α and λ it follows that (see appendix A for details),

$$E(X) = \frac{1}{\lambda^{1/\alpha}} \Gamma\left(\frac{1}{\alpha} + 1\right), \text{ where } \Gamma(x) \text{ denotes the gamma}$$

function. Thus in sampling from exponential distributions (α=1), the mean time to death for group 1 is 1 and for group 2 is successively λ, 1/1.2, 1/1.4, 1/1.6, 1/1.8, 1/2.0. The values of λ_1 and λ_2 when sampling from Weibull distributions with α= 1.25 and α=0.75, are chosen such that the mean time to failure is 1 for group 1 and successively 1, 1/1.2, 1/1.4, 1/1.6, 1/1.8, 1/2.0 for group 2. In addition, when $\alpha=1, \lambda_1 = \lambda_2 = 1$ and T=2.0 the expected proportion of censored observations in the combined sample is $\frac{1}{2}(1 - e^{-2})$ so that in all remaining samples, T^* is chosen to ensure the same expected proportion of censorings. Thus in each case, T^* is the solution of

$$\frac{1}{T^*}\left[\int_0^{T^*}(\exp(-\lambda_1 y^{\alpha}) + \exp(-\lambda_2 y^{\alpha}))dy\right] = 1 - e^{-2}.$$

The simulations will be conducted with equal sample sizes ($n_1 = n_2 = n$) in the two groups.

§2.2. A parametric model

The Weibull distribution

A natural choice for the distribution of survival time T is the Weibull distribution, which may be defined through its hazard function as follows

$$\lambda_T(t) = \lambda\alpha t^{\alpha-1} \qquad \alpha, \lambda > 0, t > 0.$$

The distribution function and p.d.f. of T are given by

$$F_T(t) = 1 - \exp(-\lambda t^{\alpha})$$

$$f_T(t) = \lambda\alpha t^{\alpha-1}\exp(-\lambda t^{\alpha}) \qquad t > 0.$$

The exponential distribution is an important special case when α=1.

For j=1,2, let $F_{T_j}(t) = 1 - \exp(-\lambda_j t^{\alpha})$ 2.1

where $\lambda_1 = \lambda$ and $\lambda_2 = \phi\lambda$. Note that

$$(1 - F_{T_2}(t)) = (1 - F_{T_1}(t))^{\phi} .$$

i.e. the survivor functions considered form a Lehmann family. If the 'shape' parameter α is allowed to take different values in the two groups this property no longer holds. Under these assumptions, the log likelihood function is given by

$$\ell(\phi, \lambda, \alpha) = (\log \lambda + \log \alpha)\sum_{j=1}^{2}\sum_{i=1}^{n_j} \delta_{ji} + (\log\phi)\sum_{i=1}^{n_2} \delta_{2i}$$

$$+ (\alpha-1)\sum_{j=1}^{2}\sum_{i=1}^{n_j} \delta_{ji} \log t_{ji} - \lambda\left\{\sum_{i=1}^{n_1} t_{1i}^{\alpha} + \phi\sum_{i=1}^{n_2} t_{2i}^{\alpha}\right\} \qquad \text{[illegible]}$$

The maximum likelihood estimates $\hat{\phi}, \hat{\lambda}, \hat{\alpha}$ of ϕ, λ, α are then the solutions of the equations

$$\frac{1}{\hat{\lambda}}\sum_{j=1}^{2}\sum_{i=1}^{n_j} \delta_{ji} - \left[\sum_{i=1}^{n_1} t_{1i}^{\hat{\alpha}} + \hat{\phi}\sum_{i=1}^{n_2} t_{2i}^{\hat{\alpha}}\right] = 0$$

$$\frac{1}{\hat{\phi}}\sum_{i=1}^{n_2} \delta_{2i} - \hat{\lambda}\sum_{i=1}^{n_2} t_{2i}^{\hat{\alpha}} = 0$$

$$\frac{1}{\hat{\alpha}}\sum_{j=1}^{2}\sum_{i=1}^{n_j} \delta_{ji} + \sum_{j=1}^{2}\sum_{i=1}^{n_j} \delta_{ji} \log t_{ji} - \hat{\lambda}\left(\sum_{i=1}^{n_1} t_{1i}^{\hat{\alpha}} \log t_{1i} + \hat{\phi}\sum_{i=1}^{n_2} t_{2i}^{\hat{\alpha}} \log t_{2i}\right) = 0.$$

The second partial derivatives of $\ell(\phi, \lambda, \alpha)$ are given by

$$\frac{\partial^2 \ell(\phi,\lambda,\alpha)}{\partial\phi^2} = -\frac{1}{\phi^2}\sum_{i=1}^{n_2} \delta_{2i}$$

$$\frac{\partial^2 \ell(\phi,\lambda,\alpha)}{\partial\phi\,\partial\lambda} = -\sum_{i=1}^{n_2} t_{2i}^{\alpha}$$

$$\frac{\partial^2 \ell(\phi,\lambda,\alpha)}{\partial \lambda^2} = -\frac{1}{\lambda^2} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \delta_{ji}$$

$$\frac{\partial^2 \ell(\phi,\lambda,\alpha)}{\partial \alpha \partial \phi} = -\lambda \sum_{i=1}^{n_2} t_{2i}^{\alpha} \log t_{2i}$$

$$\frac{\partial^2 \ell(\phi,\lambda,\alpha)}{\partial \alpha \partial \lambda} = -\left\{ \sum_{i=1}^{n_1} t_{1i}^{\alpha} (\log t_{1i}) + \phi \sum_{i=1}^{n_2} t_{2i}^{\alpha} (\log t_{2i}) \right\}$$

$$\frac{\partial^2 \ell(\phi,\lambda,\alpha)}{\partial \alpha^2} = -\frac{1}{\alpha^2} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \delta_{ji} - \lambda \left\{ \sum_{i=1}^{n_1} t_{1i}^{\alpha} (\log t_{1i})^2 + \phi \sum_{i=1}^{n_2} t_{2i}^{\alpha} (\log t_{2i})^2 \right\}$$

The expected values of the second derivatives cannot be evaluated unless assumptions concerning the censoring mechanism are made. However, the asymptotic covariance matrix of (ϕ,λ,α) may be estimated consistently by $\underline{V} = [V_{ij}]$, the inverse of the negative of the matrix of second partial derivatives evaluated at the maximum likelihood estimates, and asymptotically

$$(\hat{\phi}, \hat{\lambda}, \hat{\alpha})' \sim N((\phi,\lambda,\alpha)', \underline{V}).$$

An asymptotically efficient test of $H_0: \phi = 1$ against one or two-sided alternatives may then be performed using the relation $\hat{\phi} \sim N(\phi, V_{22})$. Alternatively the standard likelihood ratio test statistic

$$L = -2\,(\ell(\hat{\phi}, \hat{\lambda}, \hat{\alpha}) - \ell(\phi=1, \tilde{\lambda}, \tilde{\alpha})) \quad , \text{ where}$$

$\ell(\phi=1, \tilde{\lambda}, \tilde{\alpha})$ is the maximum value of the log likelihood under the restriction $\phi=1$, may be used. Under H_0, L is distributed asymptotically as χ^2_1. These two tests will be termed the ML and LR tests respectively.

The Exponential Distribution

The analysis in the exponential case is simpler. Replacing α by 1 in 2.2., it follows that $\hat{\phi}$ the maximum likelihood estimate of ϕ is given directly by

$$\hat{\phi} = \left(\frac{\sum_{i=1}^{n_1} t_{1i}}{\sum_{i=1}^{n_1} \delta_{1i}}\right)\left(\frac{\sum_{i=1}^{n_2} t_{2i}}{\sum_{i=1}^{n_2} \delta_{2i}}\right)^{-1} \qquad 2.5.$$

The asymptotic covariance matrix of $(\hat{\phi}, \hat{\lambda})$ may be estimated as before by evaluating second partial derivatives at the maximum likelihood estimates. These second derivatives are given by

$$\frac{\partial^2 \ell(\phi,\lambda)}{\partial\phi^2} = -\frac{1}{\phi^2}\sum_{i=1}^{n_2} \delta_{2i},$$

$$\frac{\partial^2 \ell(\phi,\lambda)}{\partial\phi\partial\lambda} = -\sum_{i=1}^{n_2} t_{2i},$$

$$\frac{\partial^2 \ell(\phi,\lambda)}{\partial\lambda^2} = -\frac{1}{\lambda^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j} \delta_{ji}$$

Tests of hypotheses concerning the parameter ϕ can be performed as in the Weibull case with the obvious simplification.

Although the ML and LR tests are equivalent asymptotically, it is of interest to compare their performance in small samples. Using an extension of the Monte Carlo procedure discussed in §2.1 the small sample powers of these tests are compared at the end of this section in the special case of exponential survival times.

Assumptions concerning censoring mechanism

If assumptions concerning the censoring mechanism are made, the expected values of the second partial derivatives of the log likelihood given above may be evaluated. These results will be presented in the exponential case.

Firstly, under the fixed observation time model if Y_{ji} represents the maximum observable time for individual i in group j, then

$$E\left\{-\frac{\partial^2 \ell(\phi,\lambda)}{\partial\phi^2}\right\} = \frac{1}{\phi^2}\sum_{i=1}^{n_2}(1 - e^{-\lambda\phi Y_{2i}}) = I_{11}(\phi,\lambda)$$

$$E\left\{-\frac{\partial^2 \ell(\phi,\lambda)}{\partial\phi\partial\lambda}\right\} = \frac{1}{\lambda\phi}\sum_{i=1}^{n_2}(1 - e^{-\lambda\phi Y_{2i}}) = I_{12}(\phi,\lambda) = I_{21}(\phi,\lambda)$$

$$E\left\{-\frac{\partial^2 \ell(\phi,\lambda)}{\partial\lambda^2}\right\} = \frac{1}{\lambda^2}\left\{(n_1 + n_2) - \left(\sum_{i=1}^{n_1} e^{-\lambda Y_{1i}} + \sum_{i=1}^{n_2} e^{-\lambda\phi Y_{2i}}\right)\right\}$$

$$= I_{22}(\phi,\lambda).$$

The corresponding quantities in the Weibull case may similarly be obtained, although they involve integrals which need to be evaluated numerically.

Secondly, under the random censorship model with $H_j(.)$ representing the censoring distribution for group j members

$$I_{11}(\phi,\lambda) = \frac{\lambda n_2}{\phi}\int_0^\infty e^{-\lambda\phi t}\,(1 - H_{Y_2}(t))\,dt$$

$$I_{12}(\phi,\lambda) = I_{21}(\phi,\lambda) = n_2\int_0^\infty t\,e^{-\lambda\phi t}\,[\lambda\phi(1 - H_{Y_2}(t)) + H'_{Y_2}(t)]\,dt$$

$$I_{22}(\phi,\lambda) = \frac{n_1}{\lambda}\int_0^\infty e^{-\lambda t}\,(1 - H_{Y_1}(t))\,dt + \frac{n_2\phi}{\lambda}\int_0^\infty e^{-\lambda\phi t}\,(1 - H_{Y_2}(t))\,dt \qquad \text{e.h.}$$

Extension to the Weibull case is straightforward.

In the above situations the asymptotic covariance matrix of $(\hat{\phi},\hat{\lambda})$ is then given by $\underline{I}^{-1}(\phi,\lambda)$ where $\underline{I}(\phi,\lambda) = [I_{ij}(\phi,\lambda)]$. To use the expressions 2.4., knowledge of $H_{Y_1}(.)$ and $H_{Y_2}(.)$ is needed, although the simplifying assumption $H_{Y_1}(t) \equiv H_{Y_2}(t)$ may in certain situations be reasonable.

The F-test

If $T_1, \ldots, T_n$ are independent and identically distributed random variables, with parameter λ then $Y = 2\lambda \sum_{i=1}^{n} T_i$ has a χ^2_{2n} distribution and it follows that the maximum likelihood estimator $\hat{\phi}$ of ϕ at 2.3 has, in the uncensored case, an F-distribution on $(2n_1, 2n_2)$ d.f. under $H_0: \phi = 1$. If, for $j=1,2$, the death process in group j is observed until a fixed number $\sum_{i=1}^{n_j} \delta_{ji}$ of deaths have occurred, the result with d.f. $(2 \sum_{i=1}^{n_1} \delta_{1i}, \; 2 \sum_{i=1}^{n_2} \delta_{2i})$ is exact and is a good approximation (Cox (1953)) if the total observation time is fixed and the number of deaths in each group random. The test procedure based on the above, known as the F test, is asymptotically efficient. Its small sample power is compared at the end of this section with that of the ML and LR tests discussed earlier.

Example

Table 2.1 shows the results of fitting Weibull and Exponential models to the data of example I.

Table 2.1. : Fitting Weibull and Exponential models to data of example I.

Model	M.l.e's	Value of log likelihood at m.l.e.	Estimated variance of estimator
I: Weibull	$\hat{\phi}=2.376$	$\ell(\hat{\phi},\hat{\lambda},\hat{\alpha})$	$var(\hat{\phi}) = 1.166$
	$\hat{\lambda}=0.020$	$=-106.599$	
	$\hat{\alpha}=1.366$		
II: Weibull	$\hat{\lambda}=0.037$	$\ell(\phi=1,\hat{\lambda},\hat{\alpha})$	
$(\phi=1)$	$\hat{\alpha}=1.141$	$=-116.405$	
III: Exptl	$\hat{\phi}=4.602$	$\ell(\hat{\phi},\hat{\lambda})$	
	$\hat{\lambda}=0.025$	$=-108.524$	

$$L_1 = -2(\ell(\phi=1,\hat{\lambda},\hat{\alpha}) - \ell(\hat{\phi},\hat{\lambda},\hat{\alpha})) = 19.652$$

$$L_2 = -2(\ell(\hat{\phi},\hat{\lambda}) - \ell(\hat{\phi},\hat{\lambda},\hat{\alpha})) = 3.890$$

$$Z = \frac{\hat{\phi} - 1}{\sqrt{var(\hat{\phi})}} = \text{[illegible]}$$

Comparison of models I and II yields the test statistics L_1 and Z. Under $H_0: \phi = 1$

i) L_1 is an observation on χ_1^2 and is significant at 0.1%pt.,

ii) Z is an observation on N(0,1) which is not significant at 10% pt.

Comparison of models I and III yields the test statistic L_2 .

Under $H_0: \alpha = 1$, L_2 is an observation on χ_1^2 and is significant at the 5% pt.

The above tests have been considered in the context of two sided alternatives.

Small sample power of the ML, LR and F tests

The small sample power of the ML, LR and F tests are compared in this section using the Monte Carlo procedure discussed in §2.1. The distribution of survival time in each of the two groups is exponential. As the LR test is two-sided the appropriate alternative hypothesis is $H_1 : \phi \neq 1$ so that two-sided significance levels have been used and further samples with $\lambda_2 = 0.2, 0.4, 0.6$, and 0.8 generated according to the following plan:

a	λ_1	λ_2	T^*
1	1	0.2	6.63051
		0.4	3.76444
		0.6	2.79772
		0.8	2.30344

Again T^* is chosen to ensure that the proportion of censored observations is ½ $(1 - e^{-2})$. Samples of size 50, 100, 200, 500 and 1000 with equal numbers in each of the 2 groups have been used. The results are given in table 2.2 which records the proportion of times $H_0 : \phi=1$ was rejected at the 5% level for each value of $\lambda_2=\phi\lambda_1$ considered. Entries for samples of size 50, 100 and 200 are each calculated from 1000 simulations, those for sample size 500 from 500 simulations and 100 simulations were used for sample size 1000. These results indicate that the F and LR tests are to be preferred, particularly in small samples (50 or 100) where the performance of the ML test is very poor. The distributional assumptions concerning ϕ, on which the ML test is based are clearly unsatisfactory. The power functions of the F and LR tests are almost identical for all sample sizes.

Table 2.2 Small sample power of ML, LR and F tests when survival times for groups 1 and 2 are exponential with parameters $\lambda_1=1$ and $\lambda_2=\phi$ respectively. Each entry is propn. of times H_0: $\phi=1$ rejected against two-sided alternative $H_1:\phi\neq1$ at 5% level.

Sample Size	50			100			200			500			1000		
λ_2	ML	LR	F	ML	LR	F	ML	LR	F	ML	LR	F	ML	LR	F
0.2	0.998	0.993	0.993	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.4	0.854	0.734	0.734	0.979	0.950	0.950	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.6	0.449	0.287	0.289	0.653	0.526	0.527	0.868	0.796	0.797	0.994	0.992	0.992	1.000	1.000	1.000
0.8	0.194	0.095	0.097	0.235	0.139	0.141	0.332	0.211	0.212	0.564	0.474	0.474	0.800	0.760	0.760
1.0	0.092	0.040	0.040	0.075	0.050	0.051	0.055	0.046	0.045	0.044	0.042	0.042	0.090	0.090	0.090
1.2	0.029	0.059	0.061	0.020	0.102	0.103	0.066	0.163	0.164	0.248	0.354	0.354	0.540	0.680	0.620
1.4	0.008	0.112	0.117	0.058	0.238	0.232	0.231	0.423	0.424	0.726	0.808	0.808	0.960	1.000	1.000
1.6	0.004	0.205	0.208	0.135	0.423	0.424	0.509	0.709	0.709	0.942	0.980	0.980	1.000	1.000	1.000
1.8	0.005	0.303	0.304	0.246	0.581	0.585	0.750	0.891	0.890	0.999	0.999	0.999	1.000	1.000	1.000
2.0	0.007	0.402	0.405	0.384	0.730	0.730	0.890	0.958	0.958	0.999	0.999	0.999	1.000	1.000	1.000
No. of simul-ations	1000			1000			1000			500			100		

§2.3. Contingency tables

Model

Suppose the time scale is split into k intervals $[\tau_0,\tau_1)$, $[\tau_1,\tau_2)$, ..., $[\tau_{k-1},\tau_k)$ where $\tau_0=0$ and $\tau_k=\infty$, $[\tau_{i-1},\tau_i)$ being termed the i'th interval. All censored observations which occur in the i'th interval will be treated as having occurred at the end of that interval, that is, just prior to τ_i. For j=1,2 let r_{ji} denote the number of group j uncensored observations occurring in the i'th interval and let s_{ji} denote the number of group j uncensored and censored observations $\ge \tau_{i-1}$. The data may then be summarized by k, 2x2 tables, where the ith table is of the form

	survivors	deaths	Total
Group 1	$s_{1i}-r_{1i}$	r_{1i}	s_{1i}
Group 2	$s_{2i}-r_{2i}$	r_{2i}	s_{2i}
Total	$s_i - r_i$	r_i	s_i

If $p_{ji} = p(T_j \ge \tau_i / T_j \ge \tau_{i-1})$ denotes the conditional probability of surviving the i'tn interval, for j=1,2, then λ_{ji}, the logistic transform of p_{ji} is defined by

$$\lambda_{ji} = \log \left\{ \frac{p_{ji}}{1 - p_{ji}} \right\}.$$

The model to be considered is

$$\lambda_{1i} = \mu + \alpha_i \qquad \lambda_{2i} = \mu + \alpha_i + \gamma_i \qquad i=1,\ldots,k \qquad 2.5.$$

and without loss of generality it may be assumed that $\sum_{i=1}^{k} \alpha_i = 0$.
The hypothesis of interest is H_0: $\gamma_i = 0$, $i=1,\ldots,k$ against the general alternative H_1: $\gamma_i \neq 0$ for at least one i.

Statistical analysis

Let R_{ji}, S_{ji}, $j=1,2$ and R_i be random variables corresponding to the observations r_{ji}, s_{ji} $j=1,2$ and r_i in the i'th table and put

$$\underline{R}_j = (R_{j1},\ldots, R_{jk})', \quad \underline{S}_j = (S_{j1},\ldots, S_{jk})' \text{ and } \underline{R} = (R_1,\ldots, R_k)'.$$

Using a straightforward modification of the methods given by Zelen (1971) it follows that inferences concerning $\underline{\gamma} = (\gamma_1,\ldots, \gamma_k)'$ may be based on the distribution of $\underline{R}_2$ conditional on the observed values of $\underline{R}$, $\underline{S}_1$ and $\underline{S}_2$, given by

$$p(\underline{R}_2 = \underline{x} \mid \underline{R} = \underline{r}, \underline{S}_1 = \underline{s}_1, \underline{S}_2 = \underline{s}_2) = \frac{C(\underline{r}, \underline{r}_2)\exp(\underline{\gamma}'\underline{r}_2)}{\sum_{\mathcal{L}} C(\underline{r}, \underline{x})\exp(\underline{\gamma}'\underline{x})} \qquad 2.6.$$

where $\underline{s}_j = (s_{j1},\ldots, s_{jk})$ for $j=1,2$ and

$$C(\underline{r}, \underline{x}) = \prod_{i=1}^{k} \binom{s_{1i}}{r_i - x_i} \binom{s_{2i}}{x_i}$$

and the summation in the denominator is over the set

$$\mathcal{L} = \{\underline{x} : x_i = 0,1,\ldots, \text{[illegible]}, \ i=1,\ldots,k\}.$$

Under H_o, 2.6. reduces to

$$p_o(\underline{R}_2 = \underline{r}_2 | \underline{R}=\underline{r},\ \underline{S}_1 = \underline{s}_1\ ,\ \underline{S}_2 = \underline{s}_2) = \prod_{i=1}^{k} \frac{\binom{s_{1i}}{r_i - r_{2i}}\binom{s_{2i}}{r_{2i}}}{\binom{s_{1i} + s_{2i}}{r_i}}$$

The tail probability associated with a test of H_o against H_1 is then

$$P = \sum_{\underline{w} \varepsilon W} p_o(\underline{R}_2 = \underline{w} |\ \underline{R} = \underline{r})$$

where $W = (\underline{w} = (w_1, \ldots, w_k)' \ : p_o(\underline{R}_2 = \underline{w}|\underline{R} = \underline{r}) \leq p_o(\underline{R}_2 = \underline{r}_2|\underline{R}=\underline{r}))$.
For large S_i, the distribution of R_{2i} conditional on $R_i = r_i$ is approximately $N(\mu_i, \sigma_i^2)$ where

$$\mu_i = \frac{r_i\ s_{2i}}{s_i} \quad \text{and} \quad \sigma_o^2 = \frac{r_i\ s_{1i}\ s_{2i}(s_i - r_i)}{s_i^2(s_i - 1)} \qquad i=1,\ldots,k.$$

Thus, under H_o , $\sum_{i=1}^{k} (R_{2i} - \mu_i)^2/\sigma_i^2$ is distributed asymptotically as χ^2_k. These results were given by Zelen (1973) for the case of extreme censoring. He then modified them to deal with the general censoring situation using an approximation due to Feldstein (1973).

Mantel's statistic

Mantel (1966) also considered the two-group problem and proposed the test statistic

$$M = \left(\sum_{i=1}^{k} r_{2i} - \sum_{i=1}^{k} \mu_i\right)^2 \Big/ \sum_{i=1}^{k} \sigma_i^2 \qquad 2.7.$$

where v_i and σ_i^2 are as before, for testing the difference between the two groups. As Zelen (1973) points out, this would be the appropriate test statistic if the model at 2.5 were

$$\lambda_{1i} = \mu + \beta_i \ , \quad \lambda_{2i} = \mu + \beta_i + \alpha \qquad (2.8)$$

and one wished to test the hypothesis H_0: $\alpha=0$ against H_1: $\alpha\neq0$. Under H_0, M is asymptotically an observation on a random variable having a χ^2_1 distribution. Zelen claims that the assumption implicit at 2.8, that the difference between the two survival probabilities on a logistic scale is constant for each table, is valid only if the observations within each group are from an exponential distribution and the intervals chosen are of equal widths. In view of the connection between subsequent methods and Mantel's statistic, this claim is difficult to support.

A special case

Although the choice of intervals in a contingency table analysis is somewhat arbitrary, it will be convenient for comparison purposes later to consider the following particular situation. Using the notation of the introduction, partition the time axis at points where deathsoccur and construct the intervals

$$[t_{(0)}, t_{(1)}), \ [t_{(1)}, t_{(2)}), \ \ldots, \ [t_{(k-1)}, t_{(k)}), \ [t_{(k)}, t_{(k+1)}),$$

where $t_{(0)} = 0$ and $t_{(k+1)} = \infty$. For $i=0,1,\ldots,k$, the 2x2 table corresponding to the interval $[t_{(i)}, t_{(i+1)})$ is then

	survivors	deaths	Total
Group 1	$m_{1i}-d_{1i}$	d_{1i}	m_{1i}
Group 2	$m_{2i}-d_{2i}$	d_{2i}	m_{2i}
Total	$m_i - d_i$	d_i	m_i

In this case the asymptotic test statistic

$\sum_{i=1}^{k} (R_{2i}-\mu_i)^2/\sigma_i^2$ reduces to

$$S_d = \sum_{i=1}^{k} \left(\frac{d_{2i} - \mu_i}{\sigma_i}\right)^2 \qquad 2.9.$$

where $\mu_i = \dfrac{d_i\, m_{2i}}{m_i}$, $\sigma_i^2 = \dfrac{d_i\, m_{1i}\, m_{2i}\, (m_i - d_i)}{m_i^2 (m_i - 1)}$.

It is assumed that σ_i^2 and μ_i are non zero. Tables in which $m_{ji}=0$ j=1 or 2 contribute no information and are ignored. Note that the first table is ignored i.e. censorings prior to the first death contribute no information.

Correspondingly, Mantel's statistic M is given by

$$M_d = \left(\sum_{i=1}^{k} d_{2i} - \sum_{i=1}^{k} \mu_i\right)^2 \Bigg/ \sum_{i=1}^{k} \sigma_i^2 \qquad 2.10.$$

with μ_i , σ_i^2 as above.

For the data of example I, S_d = 37.09, M_d = 16.73.

S_d is significant at the 0.5% point of χ^2_{17} and M_d is significant at the 0.1% point of χ^2_1.

2.4. The Generalised Wilcoxon test

Gehan's test

Wilcoxon (1945) proposed a statistic for comparing two groups of observations. An extension of the Wilcoxon procedure, applicable when observations are subject to censoring has been considered by Gehan (1965a). Gehan's statistic W is defined as follows:

For $i=1,\ldots,n_1$; $j=1,\ldots,n_2$ let

$$U_{ij} = \begin{cases} +1 & \text{if } \delta_{2j} = 1 \text{ with } t_{1i} > t_{2j} \text{ and } \delta_{1i} = 1 \\ & \qquad\qquad\quad \text{or } t_{1i} \geq t_{2j} \text{ and } \delta_{1i} = 0 \\ -1 & \text{if } \delta_{1i} = 1 \text{ with } t_{1i} < t_{2j} \text{ and } \delta_{2j} = 1 \\ & \qquad\qquad\quad \text{or } t_{1i} \leq t_{2j} \text{ and } \delta_{2j} = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

and put $W = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U_{ij}$. A test of H_0: $F_{T_1}(t) = F_{T_2}(t)$ against one or two-sided alternatives is then constructed by computing the permutation distribution of W over the $(n_1+n_2)!/n_1!n_2!$ possible samples leading to the same combined observed censoring pattern. This test will be referred to as the W test.

Under H_0, Gehan obtains expressions for the mean and variance of W conditional on the observed censoring pattern and, by establishing the asymptotic normality of this quantity, constructs a large sample test of H_0 in the usual way. Mantel (1967), by considering a different representation of the Gehan statistic, has simplified the calculation of the permutation distribution of W and its variance under H_0.

Modifications

Gehan (1965b) has generalised the above techniques to deal with interval censoring.

Breslow (1970) points out that Gehan's permutation test and his results concerning the moments of W are applicable only under the random censorship model, with common censoring distributions, and then modifies the Gehan procedure to deal with situations in which this assumption is not valid. Breslow discusses these techniques in the context of comparing $K(\geq 2)$ groups. Efron (1967) has proposed several modifications of the above procedures which increase their power against certain parametric alternatives. Peto and Peto (1972) have also suggested improvements of the W test.

Example

For the data of example I, the value of W and its moments under H_0 are

$$W = 271, \quad E(W) = 0, \quad \text{var}(W) = 5644.39.$$

The observed value of the asymptotic test statistic $z = W/\sqrt{\text{var}(W)}$ is then 3.59, which, in a test of H_0 against the two-sided general alternative, is significant at the 0.1% point of $N(0,1)$.

Asymptotic Efficiency

Gehan (1965a) has compared the W test and the F test of §2.2 using the criterion of A.R.E. The survival distribution in groups 1 and 2 are thus assumed to be exponential with parameters λ and $\theta\lambda$ respectively. Gehan's calculations are conducted with equal sample sizes in the two groups and in the following two situations:

i) Extreme censoring, observation stops at t^*.

ii) Under the random censorship model with censoring distributions $H_{Y_j}(y)$, $j=1,2$ defined by

$$H_{Y_1}(y) = H_{Y_2}(y) = y/t^* \qquad y \in (0, t^*].$$

In each of these cases, Gehan's results indicate that the W test compares favourably with the F test as $n \to \infty$, although losses in efficiency increase as λt^* increases.

§2.5 Peto and Peto's Logrank test

The K group case: General approach

Peto and Peto (1972) propose a method of comparing $K(\geq 2)$ groups of observations subject to censorship and it will be convenient to formulate their results in this more general context.

Assuming that in general the survivor function $1 - F_{T_j}(.)$ associated with the j^{th} group is of the form

$$(1 - F_{T_j}(t)) = (1 - F(t))^{\theta_j} \qquad j = 1, \ldots, K$$

(Lehmann family of survivor functions), the hypothesis of interest would be H_0: $\theta_j = \theta_0$ against the general alternative H_1: $\theta_j \neq \theta_0$ for at least one j. Let $\lambda(t)$ denote the hazard function corresponding to $F(t)$. The log likelihood function is then

$$\ell(\underline{\theta}) = \sum_{j=1}^{K} \sum_{i=1}^{n_j} \left[\log (1 - F(t_{ji}))^{\theta_j} + \delta_{ji} (\log \theta_j + \log \lambda(t_{ji}))\right] \qquad 2.12.$$

which under H_0 reduces to

$$\ell(\theta_o) = \theta_o \sum_{j=1}^{K} \sum_{i=1}^{n_j} \log(1 - F(t_{ji})) + (\log\theta_o) \sum_{j=1}^{K} \sum_{i=1}^{n_j} \delta_{ji} + \sum_{j=1}^{K} \sum_{i=1}^{n_j} \delta_{ji} \log f(t_{ji}).$$

It follows that $\hat{\theta}_o$ the maximum likelihood estimate of θ_o is given by

$$\hat{\theta}_o = \left(\sum_{j=1}^{K} \sum_{i=1}^{n_j} \delta_{ji} \right) \left[\sum_{j=1}^{K} \sum_{i=1}^{n_j} \log(1 - F(t_{ji})) \right]^{-1} \qquad 2.13.$$

Differentiating 2.12. produces

$$\frac{\partial \ell(\underline{\theta})}{\partial \theta_j} = \sum_{i=1}^{n_j} \log(1 - F(t_{ji})) + \frac{1}{\theta_j} \sum_{i=1}^{n_j} \delta_{ji}$$

$$\frac{\partial^2 \ell(\underline{\theta})}{\partial \theta_j^2} = -\frac{1}{\theta_j^2} \sum_{i=1}^{n_j} \delta_{ji} \qquad j=1,\ldots,K$$

Putting $O_j = \sum_{i=1}^{n_j} \delta_{ji}$, defining

$$E_j = -\theta_o \sum_{i=1}^{n_j} \log(1 - F(t_{ji}))$$

and using the properties that asymptotically

$$E\left\{ \left. \frac{\partial \ell(\underline{\theta})}{\partial \theta_j} \right|_{\theta_j=\theta_o} \right\} = 0 \quad \text{and}$$

$$\operatorname{var}\left\{ \left. \frac{\partial \ell(\underline{\theta})}{\partial \theta_j} \right|_{\theta_j=\theta_o} \right\} = E\left\{ - \left. \frac{\partial^2 \ell(\underline{\theta})}{\partial \theta_j^2} \right|_{\theta_j=\theta_o} \right\} \qquad j=1,\ldots,K$$

it then follows that, under H_o,

$E(O_j - E_j) = 0$ and var $(O_j - E_j) = E(O_j) = E(E_j)$
so that asymptotically

$$\sum_{j=1}^{K} (O_j - E_j)^2/E_j \sim \chi^2_K \, .$$

θ_a at 2.13. may then be used to provide an estimate $\hat{E}_j$ of E_j
and since $\sum_{j=1}^{K} O_j = \sum_{j=1}^{K} \hat{E}_j$.

$$\sum_{j=1}^{K} (O_j - \hat{E}_j)^2/\hat{E}_j \text{ is asymptotically } \chi^2_{K-1}$$

The calculation of $\hat{E}_j$ requires knowledge of $F(t)$. However Peto and Peto point out that using Altshuler's estimate of the common survivor function, $\hat{E}_j$ can be replaced by $\tilde{E}_j = -\sum_{i=1}^{n_j} e(t_{ji})$ where
$e(t) = - \sum_{t(i) \leq t} d_i/m_i$.

The test statistic is then of the form

$$\sum_{j=1}^{K} (O_j - \tilde{E}_j)^2/\tilde{E}_j \qquad 2.14.$$

having null distribution χ^2_{K-1}.

Crowley (1973) has investigated this distributional result and under the random censorship model, but not assuming equal censoring distributions between groups, has shown it to be asymptotically valid.

Note that

$$\sum_{j=1}^{K} \hat{E}_j = \sum_{j=1}^{K} \sum_{i=1}^{n_j} \int_{t} I_{\tau t_{ji}} d\hat{\Lambda}_{t} = \sum_{j=1}^{K} \sum_{i=1}^{k} \frac{d_i}{n_i} n_{ji} = \sum_{j=1}^{K} O_j$$

so that when K=2, 2.14 reduces to

$$(O_1 - \hat{E}_1)^2/\hat{E}_1 + (O_2 - \hat{E}_2)^2/\hat{E}_2 = (O_1 - \hat{E}_1)^2 \left\{ \frac{1}{\hat{E}_1} + \frac{1}{(O_1+O_2-\hat{E}_1)} \right\} .$$

<u>The two group case: Logrank test</u>

In the two group case Peto (1972) discusses a test which is a modification of the above. This test, called the Logrank test, is based on the permutation distribution of n_2 scores from a finite population of $n_1 + n_2$ scores, one for each member of the sample, and may be formulated as follows. Without loss of generality it may be assumed that $\theta_1 = 1$, $\theta_2 = \tau$ so that the hypothesis of interest is H_0: $\tau = 1$ against the general alternative H_1: $\tau \neq 1$. The log likelihood from 2.12 is then

$$\ell(\tau) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \delta_{ji} \log\lambda(t_{ji}) + \sum_{i=1}^{n_1} \log(1 - F(t_{1i}))$$

$$+ \tau \sum_{i=1}^{n_2} \log(1 - F(t_{2i})) + (\log\tau) \sum_{i=1}^{n_2} \delta_{2i} .$$

and it follows that

$$\left. \frac{\partial \ell(\tau)}{\partial \tau} \right|_{\tau=1} = \sum_{i=1}^{n_2} \log(1 - F(t_{2i})) + \sum_{i=1}^{n_2} \delta_{2i}$$

The logrank statistic, L, is obtained on replacing $\log(1 - F(t))$ in this expression by Altshuler's estimate of the common log survivor function, so that

$$L = \sum_{i=1}^{n_2}\left[-\sum_{t_{(\ell)}<t_{2i}} d_\ell/m_\ell\right] + \sum_{i=1}^{n_2}\delta_{2i}$$

$$= \sum_{i=1}^{k}\left[d_{2\ell} - \frac{d_\ell\, m_{2\ell}}{m_\ell}\right] . \qquad 2.15$$

By defining $L_{ji} = \delta_{ji} - \sum_{t_{(\ell)}\leq t_{2i}} d_\ell/m_\ell \qquad j=1,2; i=1,\ldots,n_j$

as a sequence of $n_1 + n_2$ scores,

$$L = \sum_{i=1}^{n_2} L_{2i} .$$

Peto (1972) and Peto and Peto (1972) suggest that an exact test of H_0 against H_1 may be performed by treating L as the sum of n_2 scores randomly selected from the finite population of $n_1 + n_2$ scores. The resulting permutation test, however, will only be valid under the random censorship model with common censoring distributions in the two groups.

Comparison of equations 2.15 and 2.10 shows that the logrank statistic L is identical to the numerator of Mantel's statistic M_d.

The Modified Logrank test

In the discussion of Peto and Peto (1972), Curnow (1972) and Gehan (1972) question the use of Altshuler's estimate of the common survivor function in preference to the PL estimate. Use of the PL estimate $\hat{P}(t)$ of $1 - P(t)$ produces an alternative statistic, L^*, termed the modified Logrank statistic, given by

$$L^* = \sum_{i=1}^{n_2}\delta_{2i} + \sum_{i=1}^{n_2}\sum_{t_{(\ell)}\leq t_{2i}} \log\,(1 - d_\ell/m_\ell).$$

Thomas (1971) has established the asymptotic equivalence of L and L^*

and assuming the random censorship model in which censoring distributions may differ between groups has shown that L is asymptotically normally distributed.

Peto and Peto (1972) indicate procedures for the extension of their methods to incorporate situations in which independent variation between individuals is more extensive.

Example

Table 2.3. illustrates, using the data of example I, the calculation of the statistics L and L*. The numerical difference between their observed values is seen to be small. The test statistic $\sum_{j=1}^{k} (O_j-\tilde{E}_j)^2/\tilde{E}_j$ is also computed for this data.

Power considerations

Peto (1972) claims that the Logrank test has optimal power locally although Crowley (1974) disputes this claim and suggests an alternative justification of the Logrank statistic. For the special case of extreme censoring the locally most powerful property has been established by Johnson and Mehrotra (1972).

Thomas (1971) obtains expressions for the mean and variance of the Logrank statistic under the random censorship model with censoring distributions $H_{Y_j}(.)$, $j=1,2$. In addition, assuming that T_1 and T_2 are distributed exponentially with parameters λ and $\phi\lambda$ respectively, Thomas compares the test based on the asymptotic normality of L, under H_0: $\phi=1$, with the asymptotic test based on the marginal distribution of ϕ (equation 2.3) and shows that when $H_{Y_1}(y)\equiv H_{Y_2}(y)$, the test based on L has A.R.E. equal to 1. For the case $n_1=n_2$ with

Table 2.3: Calculation of Logrank and Modified Logrank statistics for data of Example I.

Observations: 6-MP	Observations: Placebo	i	$t_{(i)}$	d_i	n_i	Logrank score	Mod.Logrank score
		0	0	0	42		
	1,1	1	1	2	42	0.9524	0.9512
	2,2	2	2	2	40	0.9024	0.8999
	3	3	3	1	38	0.8761	0.8732
	4,4	4	4	2	37	0.8221	0.8177
	5,5	5	5	2	35	0.7690	0.7589
6,6,6		6	6	3	33	0.6741	0.6636
6*						−0.3259	−0.3364
7		7	7	1	29	0.6397	0.6286
	8,8,8,8	8	8	4	28	0.4969	0.4745
9*						−0.5031	−0.5255
10		9	10	1	23	0.4535	0.4301
10*					–	−0.5465	−0.5699
	11,11	10	11	2	21	0.3583	0.3301
11*						−0.6417	−0.6699
	12,12	11	12	2	18	0.2472	0.2123
13		12	13	1	16	0.1847	0.1477
	15	13	15	1	15	0.1181	0.0787
16		14	16	1	14	0.0462	0.0043
	17	15	17	1	13	−0.0307	−0.0757
17*,19*,20*						−1.0307	−1.0757
22	22	16	22	2	9	−0.2529	−0.3270
23	23	17	23	2	7	−0.5385	[illegible]
25*,32*,32*, 34*,35*,						}−1.5385	}−1.6635

* indicates censored observation.

$L = 10.2545$ $L^* = 9.7239$

$O_1 = 9$, $O_2 = 21$, $\hat{E}_1 = 19.2468$ and

$$(O_1 - \hat{E}_1)^2 \left[\frac{1}{\hat{E}_1} + \frac{1}{(O_1 + O_2 - \hat{E}_1)} \right] = 15.2183$$

(15.2183 is significant at the 0.1% point of χ^2_1)

$$H_{Y_1}(y) = 1 - e^{-y}, \quad H_{Y_2}(y) = 1 - e^{-ay} \qquad y>0, \quad a>0$$

he evaluates the A.R.E. for various values of a and λ. The A.R.E. is close to unity for most of the parameter values considered.

§2.6 Discussion

Lee, Desu and Gehan (1975), using the Monte Carlo procedure discussed in §2.1, compare the small sample power against one sided alternatives, in the two group case, of some of the tests in this section, namely

i) F test

ii) F_1 test : If T has a Weibull distribution with shape parameter α, then T^{α} has an exponential distribution. The F_1 test is performed by transforming each observation t to $y=t^{\alpha}$ and using the F test on the transformed observations.

iii) M test : The asymptotic form (2.10) of Mantel's test.

iv) W test : The asymptotic form of the Generalised Wilcoxon test.

v) L test : An approximation to the Logrank test, treating L, under H_0, as normally distributed with zero mean and permutational variance

$$n_1 n_2 \sum_{i=1}^{n_1+n_2} L_{2i}^2 / (n_1+n_2)(n_1+n_2-1)$$

vi) ML test : The modified Logrank test with normal approximation as in v).

In sampling from the exponential distribution, the F_1 test does not apply and of the remaining five tests, the F test is most powerful, followed closely by the M, L and ML tests. The W test is less powerful. When samples are from the Weibull distribution with $a = 0.75$ and $a = 1.25$, the F test is not valid and of the remaining tests the F_1 test is most powerful followed by the ML test. The M, L and W tests then follow in order of decreasing power. These simulations were based on sample sizes $n_1 = n_2 = 50$ and censoring rate 43%. Further samples were generated with differing sample sizes ($n_1 = n_2 = 20$) and censoring rates (0%, 10%, 25%, 75%, 90%) and regardless of the test considered it was found that power increased with increasing sample size and decreasing censoring rate. The above authors also generated samples ($n_1 = n_2 = 50$) from the Weibull distribution with different shape parameters ($a=1$ in group 1, $a=0.85$ in group 2) and found that in this case the W test is most powerful followed by the ML, M and L tests. Since the ML and L tests are formulated under the assumption that the survivor functions in the 2 groups derive from a Lehmann family this result is not surprising. For further small sample comparisons of the W, L and F tests see also Thomas (1971) and Efron (1967).

Chapter 3

REGRESSION MODELS FOR SURVIVAL DATA

§3.1 Introduction

Regression Models

In the last chapter, methods of analysis were discussed in a particular situation where individuals were classified according to a finite number of groups (perhaps corresponding to different treatments) and survival experience compared between groups. Usually other factors such as age of individual, white blood count at time of treatment, severity of disease and so on will affect survival time. The present chapter considers models which allow for the investigation of, and adjustment for, such concomitant variation.

If no censoring were present, normal theory least squares methods might be applied using some suitable transformation on survival time. However, these methods are not easily adapted to the censored case, although in particular situations some work in this area has been carried out (Sampford (1954), Sampford and Taylor (1959), Nelson and Hahn (1972), (1973), Hartley and Hocking (1971)).

Rather, recent work has proceeded on the use of regression models for distributions which are thought to approximate closely the true distribution of survival time, such as the exponential or Weibull distributions. §3.2 investigates a logistic exponential regression model appropriate in the survival data context. §3.3. looks closely at models in which the independent variables are assumed to have a multiplicative effect on the hazard function.

Notation

The notation to be used in this and subsequent chapters is an extension of that introduced in §2.1.

For $i = 1, \ldots, n$, let T_i be a continuous random variable representing survival time for the i'th individual with distribution function $F_{T_i}(t)$ and corresponding independent variables $\underline{x}_i' = (x_{i1}, x_{i2}, \ldots, x_{ip})$. Denote the observation on T_i by t_i with indicator

$$\delta_i = \begin{cases} 1 & \text{if } t_i \text{ is a death} \\ 0 & \text{if } t_i \text{ is a censoring.} \end{cases}$$

Additional notation to be used will be introduced at the beginning of the appropriate section.

§3.2. Logistic-Exponential Model

Model and Analysis

For $i = 1, \ldots, n$ assume that T_i is exponentially distributed with parameter λ_i. Split the time axis into unit intervals and for deaths let τ_i represent the interval in which individual i dies. For censorings let τ_i denote the last complete interval in which individual i was observed to have not yet died i.e. approximate censorings to have occurred at the beginning of the corresponding unit interval. Then if $Q_i = e^{-\lambda_i}$, the probability of individual i surviving a unit interval conditional on entering it, the log likelihood function of $\underline{\lambda}' = (\lambda_1, \ldots, \lambda_n)$ is given by $\ell(\underline{\lambda}) = \sum_{i=1}^{n} \left\{\tau_i \log Q_i + \delta_i \log \left[\frac{1-Q_i}{Q_i}\right]\right\}$.

Myers et. al. (1973) propose a model in which $\log\left(\frac{1-Q_i}{Q_i}\right)$ is linear in the independent variables, i.e.

$$\log\left(\frac{1-Q_i}{Q_i}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji} \qquad i = 1, \ldots, n. \qquad 3.1$$

It follows that

$$\log Q_i = -\log \left\{1 + \exp \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}\right)\right\} = -\lambda_i .$$

Substitution of these expressions in $\ell(\lambda)$ yield the log likelihood function $\ell(\beta_0, \underline{\beta})$. The above authors indicate procedures for the maximum likelihood estimation of β_0, $\underline{\beta}' = (\beta_1, \ldots, \beta_p)$.

Myers et. al. however noted that the definition of what constitutes a time interval has a direct effect upon the resulting parameter estimates. To achieve time scale invariance they postulated the existence of a parameter W and modified their model such that

$$\log\left[\frac{1 - Q_i^W}{Q_i^W}\right] = \beta_o + \sum_{j=1}^{p} \beta_j x_{ji} \quad , \qquad 3.2$$

where $Q_i^W = e^{-\lambda_i W}$ is the conditional probability of individual i surviving an interval of length W. The resulting log likelihood $\ell(\beta_o, \underline{\beta}, W)$ is as above with

$$\log Q_i = -\frac{1}{W} \log \left\{1 + \exp\left(\beta_o + \sum_{j=1}^{p} \beta_j x_{ji}\right)\right\} \quad \text{and}$$

$$\text{Log}\left[\frac{1 - Q_i}{Q_i}\right] = \log \left[\left\{1 + \exp\left(\beta_o + \sum_{j=1}^{p} \beta_j x_{ji}\right)\right\}^{1/W} - 1\right] .$$

The above authors encountered difficulties in estimating $\underline{\beta}$ and W in this modified model due to a lack of dependence of the fit on W and suggested that W might be prespecified to overcome this problem.

Time dependent exponential parameter

Mantel and Hankey (1975) considered this model further and suggested an alternative generalisation of the model at 3.1. They questioned the validity of assuming λ_i to be independent of time and included in their model a continuous time function g(t) such that

$$\lambda_i(t) = \log \left[1 + \exp\left\{\beta_o + \sum_{j=1}^{p} \beta_j x_{ji} + g(t)\right\}\right] .$$

This leads to an alternative specification of the model at 3.1 as

$$\log\left[\frac{1 - Q_{ik}}{Q_{ik}}\right] = \beta_o + \sum_{j=1}^{p} \beta_j x_{ji} + g(k) \quad , \qquad 3.3$$

where Q_{ik} is the conditional probability of individual i surviving the kth unit interval. Assuming that g(k) may be approximated by $\sum_{\ell=1}^{s} \gamma_\ell k^\ell$, a polynomial of degree s, the log likelihood function $\ell(\beta_o, \underline{\beta}, \underline{\gamma})$ is given by

$$\ell(\beta_o, \underline{\beta}, \underline{\gamma}) = \sum_{i=1}^{n} \left\{ \sum_{k=1}^{\tau_i} \log Q_{ik} + \delta_i \log \left[\frac{1 - Q_{i\tau_i}}{Q_{i\tau_i}}\right] \right\}$$

where $\log Q_{ik} = -\log\left\{1 + \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ji} + \sum_{k=1}^{m} \gamma_k k^{a}\right)\right\}$

Mantel and Hankey develop procedures for estimating the parameters in the model and discuss an application using the data given in Myers et. al.

Estimation of covariance matrix

Myers et. al. in calculating the covariance matrix of β_0, $\underline{\beta}$ (and W) in the model at 3.1 (and 3.2) evaluate the expected values of the second partial derivatives of the log likelihood under the fixed observation time model. However, Mantel and Hankey question the use of this procedure and suggest that significance tests concerning the parameters for all three models considered be carried out using the large sample likelihood ratio procedure.

Alternatively, the expected values, may be evaluated by noting that under the random censorship model, with censoring distribution $H_Y(y)$,

$$E(\delta_i) = \int_0^{\infty} \lambda_i e^{-\lambda_i t} (1 - H_Y(t)) \, dt$$

$$E(\tau_i) = \sum_{u=1}^{m} u \int_{u-1}^{u} e^{-\lambda_i t} \left[\lambda_i (1 - H_Y(t)) + e^{-\lambda_i} H'_Y(t+1)\right] dt.$$

Assuming a specific form for $H_Y(y)$ the expected values of τ_i and δ_i may be substituted in place of τ_i and δ_i in the second partial derivatives of the log likelihood to yield the covariance matrix of parameter estimators. (The second derivatives are simple linear functions of τ_i and δ_i). The introduction of the function g makes the use of this technique for the model at 3.3 computationally complex.

§1.3. Proportional Hazard Models

The Cox Model

Cox (1972) proposes a model in which, for $i = 1, \ldots, n$, the hazard function $\lambda_i(t)$ for the i'th individual is given by

MODEL I : $\lambda_i(t) = \lambda_0(t) \exp(\underline{\beta}' \underline{x}_i)$

where $\underline{\beta}' = (\beta_1, \ldots, \beta_p)$ is a vector of unknown parameters and $\lambda_0(t)$ is an unknown function of time. Note that for any two individuals $i, j, i \neq j$,

$$\lambda_i(t) = \lambda_j(t) \exp\{ \underline{\beta}' (\underline{x}_i - \underline{x}_j) \}$$

so that the model is of the proportional hazards type. An attractive feature of model I is that the function $\lambda_0(t)$, termed the underlying hazard function, is left arbitrary. In the next section models in which $\lambda_0(t)$ takes a specific form will be considered.

The Exponential and Weibull models

Prentice (1973) has considered two models in which the function $\lambda_0(t)$ takes a specific form.

MODEL II : If $\lambda_0(t) = \lambda\alpha t^{\alpha-1}$ in model I then

$$\lambda_i(t) = \lambda\alpha t^{\alpha-1} \exp(\underline{\beta}' \underline{x}_i). \qquad \alpha, \lambda > 0$$

MODEL III : If $\lambda_0(t) = \lambda$ in model I then

$$\lambda_i(t) = \lambda \exp(\underline{\beta}' \underline{x}_i).$$

Note that if $\alpha = 1$, model II reduces to III. The random variable T_i, representing survival time for the i'th individual is exponential under model III and Weibull under model II.

In the two group case

$$\underline{x}_i = x_i = \begin{cases} 0 & \text{group 1 members,} \\ 1 & \text{group 2 members} \end{cases}$$

model II reduces to

$$\lambda(t) = \begin{cases} \lambda\alpha t^{\alpha-1} & \text{group 1,} \\ \lambda\alpha e^{\beta} t^{\alpha-1} & \text{group 2} \end{cases}$$

a simple reparameterization of the model discussed in §2.2 with $\alpha = \gamma$ and $\beta = \log \sigma$.

Model III was first considered by Glasser (1967) for a single independent variable.

Myers et. al. (1973) point out an interesting connection between model IV and the form of the logisitic exponential model at 3.2. From 3.2 and for small W,

$$\log\left[\frac{1 - Q_i^W}{Q_i^W}\right] = \log\left[e^{\lambda_i W} - 1\right] = \log(\lambda_i W)$$

so that

$$\lambda_i = \frac{1}{W} e^{\lambda_0} \exp(\underline{\beta}' \underline{x}_i),$$

which is equivalent to the expression for $\lambda_i(t)$ in model III. Byar and Mantel (1975) have considered this connection in greater detail.

<u>Inclusion of strata</u>

In applications the assumption that particular independent variable(s) act multiplicatively on the hazard function may not be true. To incorporate such a situation (or alternatively to eliminate those independent variables not of primary interest), Kalbfleisch (1974b) suggests an extension of model I.

Suppose that the individuals may be split into s strata according to the values of the independent variable(s) violating the assumptions of model I (or not of primary concern) and for i=1, ..., n_j, j=1, ..., s let T_{ji} be a random variable representing survival time of the ith individual in the jth stratum. The above author proposes a model in which the hazard function of T_{ji} is given by

MODEL IV : $\lambda_{ji}(t) = \lambda_{oj}(t) \exp(\underline{\beta}' \underline{x}_{ji})$,

where $\underline{x}_{ji}$ is the vector of independent variables to be included in the description of the model for the ith member of the jth stratum. Similarly, in this situation, models II and III may be adjusted respectively as

MODEL V : $\lambda_{ji}(t) = \lambda_j \alpha_j t^{\alpha_j - 1} \exp(\underline{\beta}' \underline{x}_{ji})$.

MODEL VI : $\lambda_{ji}(t) = \lambda_j \exp(\underline{\beta}' \underline{x}_{ji})$.

Note that the proportional hazards assumption is retained within strata for each of the above models. Further generality may be obtained on allowing independent variables included in the models to have different effects between strata.

Holt and Prentice (1974) have considered models IV and VI, and model V (with $n_j = n$ for all j) in the matched pair situation i.e. $n_j = 2$, $j = 1, \ldots, n_2 = n$. These models are essentially of a different type to the general within strata models given above in that introduction of new observations introduces new parameters (or new functions in the case of model IV). The resulting methods of inference will not be considered here and the reader is referred to Holt and Prentice for details.

Related models

Although models I to VI will provide the main subject for study in the remainder of this work, several other models of the proportional hazards type have received attention. Two that will be briefly considered here, of the exponential type, closely relate to model III and could easily be extended to the stratified situation in an obvious way.

Firstly, a model proposed by Fiegl and Zelen (1965), assumes that the hazard function for individual i is given by

$$\lambda_i(t) = (a + \underline{\beta}' \underline{x}_i)^{-1} .$$

Zippin and Armitage (1966) extend the method of analysis given by Fiegl and Zelen to incorporate censored data. Several authors who have subsequently used this model have encountered computational difficulties, due to the restriction that, for $i=1, \ldots, n$, $a + \underline{\beta}' \underline{x}_i > 0$. The methods outlined by Mantel and Myers (1971) have gone some way towards solving these problems although analysis remains cumbersome.

Secondly, Greenberg et. al. (1974) propose a model in which

$$\lambda_i(t) = a + \underline{\beta}' \underline{x}_i \ .$$

Again the restriction $a + \underline{\beta}' \underline{x}_i > 0$, $i=1, \ldots, n$ is necessary and causes similar difficulties.

Thus if ease of application is considered as a criterion for choosing between different models (as it probably is in the analysis of data) either of the above should only be used if their approximation to the true situation is thought (or discovered) to be better than any of the models I to VI.

The form of exp $(\underline{\beta}' \underline{x})$

The exponent in each of the models I to VI is quite flexible in that, as Cox (1972) points out, $\underline{\beta}' \underline{x}$ may be replaced by any general function $h(\underline{\beta}, \underline{x})$ of the independent variables. A relatively simple transformation which may assist in the physical interpretation of the models I, II and III is obtained by using, for $i = 1, \ldots, n$

$$z_{ij} = x_{ij} - \bar{x}_j \ , \quad \bar{x}_j = \sum_{i=1}^{n} x_{ij} / n \ ,$$

for some or all of $j = 1, \ldots, p$. Similar transformations within strata will allow similar interpretations in models IV, V and VI.

In addition, Cox (1972) suggested that time dependent independent variables might be included in the specification of model I. For example, in the two group situation, a suitable form of model I might be

$$\lambda_i(t) = \begin{cases} \lambda_o(t) & \text{group 1 members} \\ \lambda_o(t) \exp(\beta_1 + \beta_2 t) & \text{group 2 members} \end{cases}$$

Such variables may similarly be used in models II to VI. Note however that time dependent variables destroy the proportional hazards assumption. (The comments of Kalbfleisch and Prentice (1972a) on how such variables would affect model I are misleading). Discussion of the validity of including time dependent variables of this type is given in §4.6 and §4.7 and their use regarding goodness of fit is considered in §6.1 and §6.2. Otherwise it will be assumed that independent variables are prespecified and not functions of time.

§3.4. Summary

The remainder of this work will be primarily concerned with models I to VI. Chapter 4 investigates the estimation of, and significance tests concerning, relevant parameters in the models. The efficiency of these inferential procedures will be discussed in chapter 5 and techniques for assessing 'goodness of fit' to the models in chapter 6. Chapter 7 illustrates the use of these models in an example.

Chapter 4

THE ANALYSIS OF PROPORTIONAL HAZARD MODELS

§4.1. Introduction

Summary

Inferential procedures arising from models I to VI will be discussed in this chapter. Parameter estimation will be achieved by the methods of maximum likelihood (§4.2 and §4.3) although other approaches, marginal likelihood (§4.6), partial likelihood (§4.7) and Bayesian (§4.8) will be considered. §4.4 investigates methods of estimating covariance matrices of relevant parameter estimators and §4.5 indicates tests of significance concerning their values. For the parametric models, results will usually be given for models II and V only. Corresponding expressions for models III and VI may be deduced as special cases.

Tied Data

It will be assumed throughout that random variables representing survival time are continuous. Frequently, however, data will be recorded in a form involving ties. If these are small in number, a random breaking of the ties will usually be adequate. To cover the possibility of a large number of ties, Cox (1972) discusses a logistic model closely related to model I. Kalbfleisch and Prentice (1973) extend their marginal likelihood approach to incorporate tied data, retaining the form of model I. See also Breslow (1974). Each of these methods may be employed within strata, under model IV. Analysis using any of the parametric models II, III, V and VI is unaffected by the possibility of tied data.

Notation

Let $t_1^* < \ldots < t_n^*$ denote the ordered censored and uncensored dependent observations with corresponding indicators

$$\delta_i^* = \begin{cases} 1 & \text{if } t_i^* \text{ is a death} \\ 0 & \text{if } t_i^* \text{ is a censoring,} \end{cases}$$

and independent variables $\mathbf{z}_i^*$, and denote by $t_{(1)} < t_{(2)} \ldots < t_{(k)}$ the ordered uncensored survival times.

When dealing with models IV, V and VI the notation may be extended in an obvious way. In the j'th stratum, let $t_{j1}^* < \ldots < t_{jn_j}^*$ be the ordered observations, with indicators δ_{ji}^* and independent variables $\mathbf{z}_{ji}^*$. Corresponding unordered quantities will be denoted by t_{ji}, δ_{ji} and $\mathbf{z}_{ji}$.

[illegible] Formation of likelihood functions

Model I

Cox (1972), in computing the likelihood function under model I, considers only time points t_i^* at which deaths occur $(\delta_i^* = 1)$. Given the set of individuals who have observations on survival time $\geq t_i^*$, the probability that the death is on the individual as observed is

$$\frac{\lambda_0(t_i^*) \exp(\boldsymbol{\beta}'\mathbf{z}_i^*)}{\sum_{j=i}^{n} \lambda_0(t_i^*) \exp(\boldsymbol{\beta}'\mathbf{z}_j^*)} .$$

The required likelihood is then obtained as the product, over deaths, of such terms and

$$L(\underline{\beta}) = \prod_{i=1}^{n} \left(\exp(\underline{\beta}'\underline{z}_i^*) / \sum_{j=i}^{n} \exp(\underline{\beta}'\underline{z}_j^*)\right)^{\delta_i^*} \qquad 4.1.$$

Further discussion on the formation of this likelihood is given in §4.6 and §4.7.

Models II and III

Under model II the likelihood function is given by

$$L(\underline{\beta}, \lambda, \alpha) = \prod_{i=1}^{n} \left(\lambda\, \alpha\, t_i^{\alpha-1}\, e^{\underline{\beta}'\underline{z}_i}\right)^{\delta_i} \exp\left(-\lambda t_i^{\alpha}\, e^{\underline{\beta}'\underline{z}_i}\right) \qquad 4.2.$$

Models IV, V and VI

The approach employed for model I suggests that the likelihood function $L(\underline{\beta})$ under model IV may be constructed as the product over strata, of terms

$$\prod_{i=1}^{n_j} \left(\exp(\underline{\beta}'\underline{z}^*_{ji}) / \sum_{k=i}^{n_j} \exp(\underline{\beta}'\underline{z}^*_{jk})\right)^{\delta_{ji}^*}$$

and thus

$$L(\underline{\beta}) = \prod_{j=1}^{m} \prod_{i=1}^{n_j} \left(\exp(\underline{\beta}'\underline{z}^*_{ji}) / \sum_{k=i}^{n_j} \exp(\underline{\beta}'\underline{z}_{jk})\right)^{\delta_{ji}^*} \qquad 4.3.$$

The likelihood function $L(\underline{\beta}, \underline{\lambda}, \underline{\alpha})$ under model V is formed in a similar way using 4.2 and

$$L(\underline{\beta}, \underline{\lambda}, \underline{\alpha}) = \prod_{j=1}^{m} \prod_{i=1}^{n_j} \left(\lambda_j\, \alpha_j\, t_{ji}^{\alpha_j-1}\, e^{\underline{\beta}'\underline{z}_{ji}}\right)^{\delta_{ji}} \exp\left(-\lambda_j\, t_{ji}^{\alpha_j}\, e^{\underline{\beta}'\underline{z}_{ji}}\right) \qquad 4.4.$$

where $\underline{\lambda}' = (\lambda_1, \ldots, \lambda_m)$ and $\underline{\alpha}' = (\alpha_1, \ldots, \alpha_m)$.

4.3. Parameter and function estimation

Model I

From 4.1. the log likelihood function for β is given by

$$\ell(\beta) = \sum_{i=1}^{n} \delta_i^* \left[\beta' x_i^* - \log\left\{\sum_{j=i}^{n} \exp(\beta' x_j^*)\right\}\right]. \qquad 4.5.$$

Differentiating

$$\frac{\partial \ell(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} \delta_i^* \left\{ x_{ik}^* - \frac{\sum_{j=i}^{n} x^*_{jk} \exp(\beta' x_j^*)}{\sum_{j=i}^{n} \exp(\beta' x_j^*)} \right\} \quad k=1,\ldots,p \qquad 4.6.$$

and

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_\ell \, \partial \beta_k} = - \sum_{i=1}^{n} \delta_i^* \left\{ \sum_{j=i}^{n} x^*_{j\ell} x^*_{jk} \exp(\beta' x_j^*) \sum_{j=i}^{n} \exp(\beta' x_j^*) \right.$$

$$\left. - \sum_{j=i}^{n} x^*_{j\ell} \exp(\beta' x_j^*) \sum_{j=i}^{n} x^*_{jk} \exp(\beta' x_j^*) \right\} \Big/ \left\{ \sum_{j=i}^{n} \exp(\beta' x_j^*) \right\}^2$$

$$= U_{\ell k}(\beta) \qquad \ell,k=1,\ldots,p \qquad 4.7.$$

The maximum likelihood estimate $\hat{\beta}$ of β is the solution of $\frac{\partial \ell(\beta)}{\partial \beta_k} = 0$, $k=1,\ldots,p$. With few independent variables the likelihood may be tabulated directly to obtain parameter estimates. Alternatively a Newton-Raphson iterative procedure using 4.6 and 4.7 will yield $\hat{\beta}$. Computation of the second partial derivatives at 4.7., however, may prove tedious and a search method using 4.5 may be preferable with large data sets. More will be said about thecomputational aspects of model fitting in chapter 7.

The problem of estimating $\lambda_0(t)$ has been considered by several authors. Kalbfleisch and Prentice (1973) begin by approximating this quantity as a step function as in §1.3. Estimation of the steps λ_i is a direct extension of the methods of that section and given $\hat{\beta}$ the above authors show that the maximum likelihood estimators $\hat{\lambda}_i$ of λ_i $i=1,2,\dots,r$ are given by

$$\hat{\lambda}_i = \begin{cases} d_i\left\{\sum_{s=1}^{p_i+q_i}(t_{is}-b_{i-1})\exp(\hat{\beta}' z_{is}) + (b_i-b_{i-1})\sum_{h=i+1}^{r}\sum_{s=1}^{p_h+q_h}\exp(\hat{\beta}' z_{hs})\right\}^{-1} & i=1,\dots,r-1 \\ d_r\left\{\sum_{s=1}^{p_r+q_r}(t_{rs}-b_{r-1})\exp(\hat{\beta}' z_{rs})\right\}^{-1} & i=r \end{cases} \quad (4.8)$$

where t_{is}, $s=1,\dots,p_i+q_i$ represent observations on survival time in I_i with corresponding independent variables z_{is}. The resulting estimate of the survivor function

$$\bar{F}(t) = \exp\left\{-e^{\beta' z}\int_0^t \lambda_0(u)\,du\right\}$$

for an individual with independent variables z is then

$$\hat{\bar{F}}_i(t) = \begin{cases} \exp\left\{-t\,\hat{\lambda}_1 e^{\hat{\beta}' z}\right\} & t \in I_1 \\ \exp\left[-\left\{(t-b_{i-1})\hat{\lambda}_i + \sum_{k=1}^{i-1}\hat{\lambda}_k(b_k-b_{k-1})\right\}e^{\hat{\beta}' z}\right] & t \in I_i \end{cases}$$

$i=2,\dots,r$

Alternatively, the Breslow approach of §1.3 yields estimates

$$\hat{\lambda}_i = \begin{cases} \left\{(t_{(i)}-t_{(i-1)})\sum_{j=i}^{D}\exp(\hat{\beta}' z^*_j)\right\}^{-1} & i=1,\dots,k \\ \infty & i=k+1 \end{cases}$$

which may again be used to obtain estimated survivor functions as above. To achieve a form which generalises the PL estimate, Breslow (1974) shows that the probability of surviving interval I_i conditional on entering it may be estimated by

$$\hat{\pi}_i = \begin{cases} 1 - \exp\left(-\hat{\lambda}_i \left(t_{(i)} - t_{(i-1)}\right) e^{\hat{\underline{\beta}}' \underline{x}_i}\right) & i=1,\dots,k-1 \\ 0 & i=k. \end{cases}$$

and the corresponding estimate of the survivor function is

$$\hat{F}_2(t) = \begin{cases} 1 & t \in I_1 \\ \prod_{\ell=1}^{i-1} \left\{ 1 - e^{\hat{\underline{\beta}}' \underline{x}} \left(\sum_{j=\ell}^{n} e^{\hat{\underline{\beta}}' \underline{x}^*_j} \right)^{-1} \right\} & t \in I_i \end{cases} \qquad i=2,\dots,k$$

Oakes (1972) and Cox (1972) have also considered the estimation of $\lambda_o(t)$. The approach of Oakes is similar to that of Breslow. Cox, assuming that $\lambda_o(t)$ is zero except at points where deaths occur, performs a separate maximum likelihood estimation procedure for $\lambda_o(t)$ at each of these points. The resulting estimate of the underlying survivor function is a further generalization of the PL estimate.

Models II and III

From 4.2 the log likelihood under model II is

$$\ell(\underline{\beta}, \lambda, \alpha) = (\log\lambda + \log\alpha) \sum_{i=1}^{n} \delta_i + (\alpha-1) \sum_{i=1}^{n} \delta_i \log t_i + \sum_{i=1}^{n} \delta_i \underline{\beta}' \underline{x}_i - \lambda \sum_{i=1}^{n} t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i).$$

Differentiating,

$$\frac{\partial \ell(\underline{\beta}, \lambda, \alpha)}{\partial \beta_j} = \sum_{i=1}^{n} \delta_i x_{ij} - \lambda \sum_{i=1}^{n} x_{ij} t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i) \qquad j=1,\dots,p$$

$$\frac{\partial \ell(\underline{\beta}, \lambda, \alpha)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i)$$

$$\frac{\partial \ell(\underline{\beta}, \lambda, \alpha)}{\partial \alpha} = \frac{1}{\alpha} \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \delta_i \log t_i - \lambda \left\{ \sum_{i=1}^{n} (\log t_i) t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i) \right\}. \qquad 4.9$$

The second partial derivatives of $\ell(\underline{\beta},\lambda,\alpha)$ are

$$\frac{\partial^2 \ell(\underline{\beta},\lambda,\alpha)}{\partial\beta_h \partial\beta_j} = -\lambda \sum_{i=1}^{n} x_{ih} x_{ij} t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i) \qquad h,j=1,\ldots,p$$

$$\frac{\partial^2 \ell(\underline{\beta},\lambda,\alpha)}{\partial\lambda \partial\beta_j} = -\sum_{i=1}^{n} x_{ij} t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i) \qquad j=1,\ldots,p$$

$$\frac{\partial^2 \ell(\underline{\beta},\lambda,\alpha)}{\partial\alpha \partial\beta_j} = -\lambda \sum_{i=1}^{n} x_{ij} (\log t_i)\, t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i) \qquad j=1,\ldots,p$$

$$\frac{\partial^2 \ell(\underline{\beta},\lambda,\alpha)}{\partial\lambda^2} = -\frac{1}{\lambda^2} \sum_{i=1}^{n} \delta_i$$

$$\frac{\partial^2 \ell(\underline{\beta},\lambda,\alpha)}{\partial\alpha \partial\lambda} = -\sum_{i=1}^{n} (\log t_i)\, t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i)$$

$$\frac{\partial^2 \ell(\underline{\beta},\lambda,\alpha)}{\partial\alpha^2} = -\frac{1}{\alpha^2} \sum_{i=1}^{n} \delta_i - \lambda \sum_{i=1}^{n} (\log t_i)^2 t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i) \qquad 4.10.$$

Maximum likelihood estimates $\hat{\underline{\beta}}, \hat{\lambda}, \hat{\alpha}$ of $\underline{\beta}, \lambda, \alpha$ may be obtained as with model I.

<u>Models IV, V and VI</u>

From 4.3. the log likelihood under model IV is

$$\ell(\underline{\beta}) = \sum_{j=1}^{s} \sum_{i=1}^{n_j} \delta_{ji}^{*} \left[\underline{\beta}'\underline{x}_{ji} - \log\left\{\sum_{k=i}^{n_j} \exp(\underline{\beta}'\underline{x}^{*}_{ki})\right\}\right]$$

and first and second derivatives of $\ell(\underline{\beta})$ are simply sums over strata of terms like 4.6 and 4.7. The functions $\lambda_{01}(.),\ldots,\lambda_{0s}(.)$ may be estimated by performing separate estimation procedures, as for model I, within each stratum. Under model V, the log likelihood $\ell(\underline{\beta},\underline{\lambda},\underline{\alpha})$ and its derivatives may be computed directly from 4.9 and 4.10.

§4.4. Evaluation of covariance matrices

Estimation using maximum likelihood estimates

The covariance matrix of relevant parameter estimators in each of the models may be estimated as the inverse of the negative of the matrix of second partial derivatives evaluated at the maximum likelihood estimates. The usual large sample distributional results for maximum likelihood estimates are valid, although these properties for estimators resulting from models I and IV require further justification (see §4.7). Asymptotically in the within strata models is meant in the sense that, the quantities $q_j = n_j/n$, $j=1,\ldots,s$ remain constant, while the total sample size $n = \sum_{j=1}^{s} n_j \to \infty$.

Uncensored case

Unless assumptions concerning the censoring mechanism are made, expected values of second partial derivatives in each of the models cannot be evaluated. Relatively simple results however may be obtained in the uncensored case (all individuals observed to death). Under model II, it may be shown that (see Appendix A for details),

$$E(T_i^{\alpha}) = \frac{e^{-\underline{\beta}'\underline{x}_i}}{\lambda} L(1,0)$$

$$E\{(\log T_i)T_i^{\alpha}\} = \frac{e^{-\underline{\beta}'\underline{x}_i}}{\alpha\lambda}\{L(1,1) - (\log \lambda + \underline{\beta}'\underline{x}_i)\, L(1,0)\}$$

$$E\{(\log T_i)^2 T_i^{\alpha}\} = \frac{e^{-\underline{\beta}'\underline{x}_i}}{\alpha^2\lambda}\{L(1,2) - 2(\log \lambda + \underline{\beta}'\underline{x}_i)L(1,1) + (\log\lambda + \underline{\beta}'\underline{x}_i)^2 L(1,0)\},$$

where $L(1,0) = 1$, $L(1,1) = 1-\omega$, $L(1,2) = (\omega-1)^2 + \frac{\pi^2}{6} - 1$,

and $\omega = 0.5772 \ldots$ is Euler's constant.

The asymptotic covariance matrix of the maximum likelihood estimators $\hat{\underline{\beta}}$, $\hat{\lambda}$ and $\hat{\alpha}$ is then $I^{II}(\underline{\beta},\lambda,\alpha)^{-1}$ where $I^{II}(\underline{\beta},\lambda,\alpha) = [I_{ij}^{II}(\underline{\beta},\lambda,\alpha)]$ is a symmetric matrix with elements

$$I^{II}_{jk}(\underline{\beta},\lambda,\alpha) = \sum_{i=1}^{n} x_{ij}\, x_{ik}$$

$$I^{II}_{j\,p+1}(\underline{\beta},\lambda,\alpha) = \frac{1}{\lambda}\sum_{i=1}^{n} x_{ij}$$

$$I^{II}_{j\,p+2}(\underline{\beta},\lambda,\alpha) = \frac{1}{\alpha}\sum_{i=1}^{n} x_{ij}\,(L(1,1) - (\log\lambda + \underline{\beta}'\underline{z}_i)) \qquad j,k=1,\ldots,p$$

$$I^{II}_{p+1\,p+1}(\underline{\beta},\lambda,\alpha) = \frac{n}{\lambda^2}$$

$$I^{II}_{p+1\,p+2}(\underline{\beta},\lambda,\alpha) = \frac{1}{\alpha\lambda}\sum_{i=1}^{n}(L(1,1) - (\log\lambda + \underline{\beta}'\underline{z}_i))$$

$$I^{II}_{p+2\,p+2}(\underline{\beta},\lambda,\alpha) = \frac{1}{\alpha^2}\sum_{i=1}^{n}(1 + L(1,2) - 2(\log\lambda + \underline{\beta}'\underline{z}_i)L(1,1) + (\log\lambda + \underline{\beta}'\underline{z}_i)^2).$$

Similar results for model V are obtained using

$$E(T_{ji}^{\alpha_j}) = \frac{e^{-\underline{\beta}'\underline{z}_{ji}}}{\lambda_j}\, L(1,0)$$

$$E((\log T_{ji})\, T_{ji}^{\alpha_j}) = \frac{e^{-\underline{\beta}'\underline{z}_{ji}}}{\alpha_j\,\lambda_j}\,(L(1,1) - (\log\lambda_j + \underline{\beta}'\underline{z}_{ji}))$$

$$E((\log T_{ji})^2 T_{ji}^{\alpha_j}) = \frac{e^{-\underline{\beta}'\underline{z}_{ji}}}{\alpha_j^2\lambda_j}\,(L(1,2) - 2(\log\lambda_j + \underline{\beta}'\underline{z}_{ji})L(1,1) + (\log\lambda_j + \underline{\beta}'\underline{z}_{ji})^2)$$

<u>Assumptions concerning the censoring mechanism</u>

Under the fixed observation time model, if, for $i=1,\ldots,n$ Y_i represents the maximum observable time for individual i, then in model II

$$E(\delta_i) = 1 - \exp(-\,Y_i^{\alpha} e^{\underline{\beta}'\underline{z}_i})$$

$$E(T_i^{\alpha}) = \frac{e^{-\underline{\beta}'\underline{z}_i}}{\lambda}\,(1 - \exp(-\lambda Y_i^{\alpha} e^{\underline{\beta}'\underline{z}_i}))$$

$$E((\log T_i)\, T_i^{\alpha}) = \int_0^{Y_i}(\log u)\, u^{2\alpha-1}\,\lambda\alpha e^{\underline{\beta}'\underline{z}_i}\exp(-\lambda u^{\alpha} e^{\underline{\beta}'\underline{z}_i})\,du + (\log Y_i)Y_i^{\alpha}\exp(-\lambda Y_i^{\alpha} e^{\underline{\beta}'\underline{z}_i})$$

$$E((\log T_i)^2 T_i^{\alpha}) = \int_0^{Y_i} (\log u)^2 u^{2\alpha-1} \lambda\alpha e^{\beta' x_i} \exp(-\lambda u^{\alpha} e^{\beta' x_i})\,du$$

$$+ (\log Y_i)^2 Y_i^{\alpha} \exp(-\lambda Y_i^{\alpha} e^{\beta' x_i}).$$

when $\alpha \neq 1$, evaluation of the last two terms cannot be achieved analytically. However in the exponential case, these terms are not needed and the required quantities are

$$E(\delta_i) = 1 - \exp(-\lambda Y_i e^{\beta' x_i})$$

$$E(T_i) = \frac{e^{-\beta' x_i}}{\lambda}\{1 - \exp(-\lambda Y_i e^{\beta' x_i})\}.$$

Under the random censorship model

$$E(\delta_i) = \int_0^{\infty} \lambda e^{\beta' x_i} \alpha t^{\alpha-1} \exp(-\lambda e^{\beta' x_i} t^{\alpha}) \{1 - H_{Y_i}(t)\}\,dt$$

$$E(T_i^{\alpha}) = \int_0^{\infty} t^{\alpha} C_i(t, \beta, \lambda, \alpha)\,dt$$

$$E((\log T_i) T_i^{\alpha}) = \int_0^{\infty} (\log t)\, t^{\alpha} C_i(t, \beta, \lambda, \alpha)\,dt$$

$$E((\log T_i)^2 T_i^{\alpha}) = \int_0^{\infty} (\log t)^2 t^{\alpha} C_i(t, \beta, \lambda, \alpha)\,dt,$$

where $$C_i(t, \beta, \lambda, \alpha) = \lambda e^{\beta' x_i} \alpha t^{\alpha-1} \exp(-\lambda e^{\beta' x_i} t^{\alpha})(1 - H_{Y_i}(t))$$

$$+ \exp(-\lambda e^{\beta' x_i} t^{\alpha})\, H'_{Y_i}(t),$$

and for $i=1,\ldots,n$, Y_i is a random variable representing period of observation for individual i with distribution function $H_{Y_i}(.)$.

Similar results are obtained in an obvious way for models V and VI.

4.5 Tests of significance concerning parameters

Stepwise procedures

The purpose of an analysis using any of the above models will usually be to select those independent variables having a significant effect upon survival. This may be achieved by a forward stepwise procedure similar to that used in standard multiple regression, the effect of each new independent variable introduced into the model being assessed using the large sample likelihood ratio test procedure. A backward stepwise procedure, fitting a model with all independent variables included and eliminating each one in turn, is an alternative approach used in a related context by Greenberg and Bayard (1974). The former method will prove more useful for applications in which the number of independent variables is large. In addition the hypotheses H_o: $\alpha=1$ and $H_o:\alpha_1=\alpha_2=\ldots=\alpha_s=1$ are appropriate for distinguishing between models II and III and between models V and VI respectively and may be assessed using the large sample test mentioned above.

Model I - connection with logrank test

In discussing model I, Cox (1972) indicates that the global null hypothesis $H_o:\underline{\beta} = \underline{0}$ may be tested by noting that the statistic $\left[\frac{\partial L(\underline{0})}{\partial \beta_1}, \ldots, \frac{\partial L(\underline{0})}{\partial \beta_p}\right]'$ is asymptotically normally distributed, under H_o, with zero mean vector and covariance matrix $\underline{U}(\underline{0})^{-1}$ where $U(\underline{\beta}) = [-H_{ij}(\underline{\beta})]$. In the two group case where

$$z_i^* = x_i^* = \begin{cases} 0 & \text{group 1 members} \\ 1 & \text{group 2 members} \end{cases}$$

this statistic reduces to

$$\frac{\partial \ell(0)}{\partial \beta} = \sum_{i=1}^{n} \delta_i^* \left\{ x_i^* - \frac{\sum_{j=i}^{n} x_j^*}{n-i+1} \right\} \quad \text{and}$$

$$U(0) = \sum_{i=1}^{n} \delta_i^* \left\{ \frac{(n-i+1) \sum_{j=i}^{n} x_j^* - \left(\sum_{j=i}^{n} x_j^* \right)^2}{(n-i+1)^2} \right\}.$$

In the notation of §2.1, $\sum_{i=1}^{n} \delta_i^* x_i^* = \sum_{i=1}^{k} m_{1i}$

and

$$\frac{\delta_i^*}{n-i+1} \sum_{j=i}^{n} x_j^* = \begin{cases} 0 & \text{if } \delta_i^* = 0 \\ \dfrac{m_{2i}}{m_i} & \text{if } \delta_i^* = 1 \end{cases},$$

so that $\dfrac{\partial \ell(0)}{\partial \beta} = \sum_{i=1}^{k} m_{1i} - \sum_{i=1}^{k} \dfrac{m_{2i}}{m_i}$ and

$$U(0) = \sum_{i=1}^{k} \left[\frac{m_i m_{2i} - m_{2i}^2}{m_i^2} \right] = \sum_{i=1}^{k} \frac{m_{1i} m_{2i}}{m_i^2}.$$

Comparison with 2.10 shows that the above test is equivalent to Mantel's test based on the statistic M_d. The connection with the Logrank test at 2.15 automatically follows.

§4.6. Marginal likelihood approach

Introduction

The techniques of marginal likelihood to be used in this section have been developed along similar lines by Fraser (1968) and by Kalbfleisch and Sprott (1970), for the purpose of eliminating nuisance parameters.

Model I

Several contributors to the discussion of Cox (1972) were unhappy about the formation of the likelihood function at 4.1 and Kalbfleisch and Prentice (1973) have justified its form within the framework of marginal likelihood. These authors argue that in the uncensored case, the rank vector is sufficient for $\underline{\beta}$ in the absence of knowledge of $\lambda_0(.)$ (i.e. marginally sufficient for $\underline{\beta}$, Barnard (1963)). The marginal likelihood of $\underline{\beta}$, $L(\underline{\beta})$ is then proportional to the distribution of the rank vector. In the censored case, the full rank vector is not observed and Kalbfleisch and Prentice suggest that the marginal likelihood is sensibly based on the probability that the rank vector is one of those possible under the observed sample. The resulting expression is identical to the form 4.1. It is important to note that this extension to the censored case cannot be justified formally within the context of marginal likelihood. In addition, the marginal likelihood approach assumes that $\lambda_0(.)$ is not identically zero over an open interval of the positive real line, and that independent variables are not functions of time. The marginal sufficiency arguments break down if time dependent covariates are included in the model.

Models II and III

For model II, in the uncensored case, $\underline{A} = (A_2, \ldots, A_n)'$, where $A_i = T_i/T_1$, $i=2,\ldots,n$, is marginally sufficient for $\underline{\beta}, \alpha$ and the marginal likelihood $L(\underline{\beta}, \alpha)$ of $\underline{\beta}, \alpha$ is proportional to the p.d.f. $f_a(\underline{a})$ of $\underline{A}$. The p.d.f. of $\underline{T} = (T_1, \ldots, T_n)'$ is

$$f_{\underset{\sim}{\Delta}}(\underset{\sim}{t}) = \lambda^n \alpha^n \left(\prod_{i=1}^{n} t_i^{\alpha-1}\right) \exp\left(\sum_{i=1}^{n} \underset{\sim}{\beta}'\underset{\sim}{x}_i - \lambda \sum_{i=1}^{n} t_i^{\alpha} e^{\underset{\sim}{\beta}'\underset{\sim}{x}_i}\right).$$

Applying the multivariate transformation

$$T_1 = T_1, \quad T_i = A_i T_1 \qquad i=2,\ldots,n$$

and integrating T_1 from the resulting expression it follows that

$$f_{\underset{\sim}{\Delta}}(\underset{\sim}{a}) = (n-1)!\,\alpha^{n-1} \exp\left(\sum_{i=1}^{n} \underset{\sim}{\beta}'\underset{\sim}{x}_i\right) \prod_{i=1}^{n} a_i^{\alpha-1} \left(\sum_{i=1}^{n} a_i^{\alpha} e^{\underset{\sim}{\beta}'\underset{\sim}{x}_i}\right)^{-n} \quad \text{where } a_1=1$$

$$= L(\underset{\sim}{\beta}, \alpha).$$

In the censored case the marginal likelihood of $\underset{\sim}{\beta}, \alpha$ is proportional to the probability that $\underset{\sim}{\Delta}$ is one of those possible given the sample. (As in the case of model I this extension to the censored case cannot be justified formally). Without loss of generality it may be assumed that

$$\delta_1 = \delta_2 = \ldots = \delta_r = 1 \qquad r \geq 2$$

$$\delta_{r+1} = \delta_{r+2} = \ldots = \delta_n = 0,$$

and the event of interest is

$A_2 = a_2, \ldots, A_r = a_r, \; A_{r+1} > a_{r+1}, \ldots, A_n > a_n$ having probability

$$\int_0^{\infty} p(T_1 = t_1, A_2 = a_2, \ldots, A_r = a_r, A_{r+1} > a_{r+1}, \ldots, A_n > a_n)\, dt_1$$

$$= \int_0^{\infty} p(T_{r+1} > t_1 a_{r+1}, \ldots, T_n > t_1 a_n / T_1 = t_1)\, p(T_1 = t_1, A_2 = a_2, \ldots, A_r = a_r)\, dt_1$$

$$= \lambda^r \alpha^r \prod_{i=1}^{r} a_i^{\alpha-1} \exp\left(\sum_{i=1}^{r} \underline{\beta}' \underline{x}_i\right) \int_0^\infty t_1^{n\alpha-1} \exp\left(-\lambda t_1^\alpha \sum_{i=1}^{n} a_i^\alpha e^{\underline{\beta}'\underline{x}_i}\right) dt_1$$

$$= (r-1)!\, \alpha^{r-1} \prod_{i=1}^{n} a_i^{\delta_i(\alpha-1)} \exp\left(\sum_{i=1}^{n} \delta_i \underline{\beta}'\underline{x}_i\right) \left(\sum_{i=1}^{n} a_i^\alpha e^{\underline{\beta}'\underline{x}_i}\right)^{-r}$$

$$= L(\underline{\beta}, \alpha) \qquad 4.11.$$

<u>Models IV, V and VI</u>

Corresponding arguments in model IV indicate that the marginal likelihood of $\underline{\beta}$ arises out of the joint distribution of the set of rank vectors (with the usual extension to the censored case), one for each stratum, and the resulting expression is identical to 4.3. Similar considerations also extend results to incorporate model V.

<u>Inferential procedures</u>

Methods of inference based on the marginal likelihoods in models II and III have been discussed by Prentice (1973). He suggests that tests of significance concerning $\underline{\beta}$, α in model II and $\underline{\beta}$ in model III be conducted by comparing the null values to be tested with the corresponding distribution of the marginally sufficient statistic. Similar methods may be used with models V and VI. In each of these models suitable estimators of the parameters are provided by the mode of the marginal likelihood. Differentiating the log marginal likelihood of $\underline{\beta}, \alpha$ in model II shows that $\hat{\underline{\beta}}, \hat{\alpha}$ are the solutions of

$$\frac{1}{\hat{\alpha}} \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \delta_i \log t_i - r \log\left(\sum_{i=1}^{n} t_i^{\hat{\alpha}} e^{\hat{\underline{\beta}}'\underline{x}_i}\right) = 0$$

$$\sum_{i=1}^{n} \delta_i x_{ij} - \left(\sum_{i=1}^{n} \delta_i\right) \sum_{i=1}^{n} t_i^{\hat{\alpha}} x_{ij} e^{\hat{\underline{\beta}}'\underline{x}_i} \left(\sum_{i=1}^{n} t_i^{\hat{\alpha}} e^{\hat{\underline{\beta}}'\underline{x}_i}\right)^{-1} = 0, \quad j=1,\dots,p.$$

Comparison with 4.9 shows that the resulting estimates are identical to those obtained using the standard maximum likelihood procedure. Estimates in model V, obtained as above, also coincide with those obtained by maximum likelihood.

The large sample properties of maximum likelihood, however, have yet to be established for methods based on marginal likelihood. The implications of Kalbfleisch and Prentice (1973), that these should necessarily hold in general, must be treated with caution, although for the special cases of the marginal likelihoods based on models I and IV, such large sample results are valid, as will be seen in the next section.

4.7 Partial likelihood approach

Definition and properties

Cox (1975) has clarified the position concerning his 'conditional' likelihood for model I through the concept of partial likelihood, which in Cox's notation may be defined as follows:

Suppose $\underline{Y}$ is a random variable with p.d.f. $f_{\underline{Y}}(\underline{y} ; \underline{\theta})$ which may be transformed to a sequence of (possibly vector-valued) random variables $(X_j, S_j ; j=1,\ldots,m)$ the transformation not depending on the unknown parameter $\underline{\theta}$. Then

$$f_{X_1,X_2,\ldots,X_j,S_1,S_2,\ldots,S_j} = \prod_{j=1} f_{X_j,S_j/X_1,\ldots,X_{j-1},S_1,\ldots,S_{j-1}}$$

$$= \prod_{j=1}^{m} f_{X_j/X_1,\ldots,X_{j-1},S_1,\ldots,S_{j-1}} \prod_{j=1}^{m} f_{S_j/X_1,\ldots,X_j,S_1,\ldots,S_{j-1}}$$

and the second term of this expression is defined as the partial likelihood based on $\underline{S}$ in the sequence $(X_j, S_j \; ; \; j=1,\ldots,m)$. (N.B. $f_{X_1,X_2,\ldots,X_j,S_1,S_2,\ldots,S_j}$ is an abbreviated form of $f_{X_1,X_2,\ldots,X_j,S_1,S_2,\ldots,S_j}(x_1, x_2,\ldots, x_j, s_1, s_2,\ldots, s_j)$ and so on). If a suitable transformation is available such that the partial likelihood depends only on the parameters of interest then inferences concerning these parameters may be based on this likelihood. Cox discusses several points associated with the uniqueness and formation of partial likelihoods. In addition he shows that the standard large sample properties of maximum likelihood, that is

i) asymptotic normality of parameter estimators,

ii) consistency of the matrix of 2nd partial derivatives, evaluated at either the parameter estimates or the true parameter values, in the estimation of the covariance matrix,

iii) large sample χ^2 test procedure based on the likelihood ratio,

are all valid when dealing with partial likelihoods.

Models I and IV

Using the notation of §4.1 for $j=1,\ldots,k$ let S_j represent the event, individual with independent variable $\underline{z}_j^{*}$ dies at $t_j^{*}(=t_{(i)})$, and let Y_j reperesent the event that a death occurs at t_j^{*} and the individuals censored in $[t_{(i-1)}, t_{(i)})$ are as observed. Then, Cox argues that the resulting partial likelihood for model I is 4.1. Crowley (1974) makes these points mathematically explicit for the two sample problem. A further point of importance is that this approach allows the inclusion of time dependent covariates.

It follows directly that the likelihood under model IV, at 4.3, may be interpreted as a partial likelihood and the inclusion of time dependent covariates is also permitted in this model.

§4.8. Bayesian approach

Models II and III

A Bayesian approach to the analysis of models II and III might sensibly choose the non-informative priors $\pi(\beta, \lambda) = \frac{1}{\lambda}$ for model III and $\pi(\beta,\lambda,\alpha) = \frac{1}{\alpha\lambda}$ for model II. The posterior density under model II is then

$$\pi(\beta,\lambda,\alpha) = \frac{(\alpha\lambda)^{\sum_1^n \delta_i - 1} \prod_{i=1}^{n} t_i^{\delta_i(\alpha-1)} \, e^{\sum_1^n \delta_i \beta' z_i}}{\exp(\lambda \sum_{i=1}^{n} t_i^{\alpha} \exp(\beta' z_i))} ,$$

and using the result

$$\int_0^\infty x^k e^{-bx} = \frac{k!}{b^{k+1}}$$, where k is a non negative integer, it follows

that the marginal posterior density of α, β is

$$\pi(\beta,\alpha) = \int_0^\infty \pi(\beta,\lambda,\alpha)\, d\lambda$$

$$= \frac{\left(\sum_{i=1}^{n} \delta_i - 1\right)! \, \alpha^{\sum_1^n \delta_i - 1} \exp\left[\sum_{i=1}^{n} \delta_i \beta' z_i\right] \prod_{i=1}^{n} t_i^{\delta_i(\alpha-1)}}{\left\{\sum_{i=1}^{n} t_i^{\alpha} \exp(\beta' z_i)\right\}^{\sum_1^n \delta_i}} \qquad 4.12$$

which is identical to the expression 4.11 for $L(\beta,\alpha)$ (except for a constant of proportionality).

Models V and VI

In model V with prior density

$$\pi(\beta,\lambda,\sigma) = \frac{1}{\prod_{j=1}^{n} \lambda_j \sigma_j} \quad .$$

the marginal posterior density of β, σ is the product over strata of terms 4.12., identical to the marginal likelihood under model V.

Chapter 5

EFFICIENCY COMPARISONS

§5.1. Introduction

Measuring efficiency

This chapter is concerned mainly with the relative efficiency of methods of estimation based on model I when the true model is either II or III. Similar comparisons between the within strata models are also considered.

A suitable measure of asymptotic efficiency of methods of estimation is given by Kendall and Stuart (1972, p.19) for estimators which are asymptotically normally distributed. If θ_1 is an asymptotically efficient estimator of the particular $\underline{\beta}$ component θ of interest and θ_2 is another estimator with $\mathrm{var}(\theta_1) = \frac{a_1}{n^s}$, $\mathrm{var}(\theta_2) = \frac{a_2}{n^s}$ as $n \to \infty$, then the asymptotic relative efficiency of θ_2 (compared to θ_1) is given by

$$R_{2,1} = \left(\frac{a_1}{a_2}\right)^{\frac{1}{s}} .$$

The quantity $R_{2,1}$ may be interpreted in large samples as the inverse ratio of sample sizes needed to give the estimators equal variances. In all applications to be considered here, $s = 1$. Results concerning the asymptotic efficiency of testing procedures may be obtained by exploiting the connection between A.R.E. and estimating efficiency outlined by Kendall and Stuart (1972, p 284/5).

Monte Carlo methods, similar to those of §2.1 will be used to verify these asymptotic results and to assess the effect of censoring on them. Efficiency comparisons in small samples will also be achieved by simulation.

The within strata models are investigated in §5.5. The large and small sample results for model I are considered for the 2 and K group cases in §5.2., while §5.3 and§5.4 investigate the one and two independent variable situations respectively.

<u>Asymptotic covariance matrix of $\hat{\beta}$ in parametric models</u>

It is convenient at this stage to consider again the results of §4.4 concerning the asymptotic covariance matrix of $\hat{\beta}$ for the parametric models in the uncensored case. Under model II, the information matrix may be partitioned in a natural way as

$$\underset{\sim}{I}^{II}(\underset{\sim}{\beta},\lambda,\alpha) = \begin{bmatrix} \underset{\sim}{A} & \vdots & \underset{\sim}{C} \\ \cdots & \cdots & \cdots \\ \underset{\sim}{C}' & \vdots & \underset{\sim}{B} \end{bmatrix}$$

where $\underset{\sim}{A}$, a $p \times p$ symmetric matrix, has elements

$$a_{kj} = I_{kj}^{II}(\underset{\sim}{\beta},\lambda,\alpha) \qquad k,j=1,\ldots,p,$$

$\underset{\sim}{C}$, a $p \times 2$ matrix has elements

$$C_{kj} = I_{k\,p+j}^{II}(\underset{\sim}{\beta},\lambda,\alpha) \qquad k=1,\ldots,p\ ;\ j=1,2,$$

and $\underset{\sim}{B}$, a 2×2 symmetric matrix, has elements

$$B_{kj} = I_{p+k\ p+j}^{II}(\underset{\sim}{\beta},\lambda,\alpha) \qquad k,\ j=1,2.$$

The asymptotic covariance matrix $\left[\underset{\sim}{I}^{II}(\underset{\sim}{\beta},\lambda,\alpha)\right]^{-1}$ may then be conveniently written as

$$\begin{bmatrix} \underset{\sim}{M} & \vdots & -\underset{\sim}{M}\,\underset{\sim}{C}\,\underset{\sim}{B}^{-1} \\ \cdots & \cdots & \cdots \\ -\underset{\sim}{B}^{-1}\,\underset{\sim}{C}'\underset{\sim}{M} & \vdots & \underset{\sim}{B}^{-1} + \underset{\sim}{B}^{-1}\,\underset{\sim}{C}'\,\underset{\sim}{M}\,\underset{\sim}{C}\,\underset{\sim}{B}^{-1} \end{bmatrix}$$

where $\underset{\sim}{M} = (\underset{\sim}{A} - \underset{\sim}{C}\,\underset{\sim}{B}^{-1}\,\underset{\sim}{C}')^{-1}$ and the marginal distribution of $\hat{\beta}$ is asymptotically $N(\underset{\sim}{\beta},\ \underset{\sim}{M})$.

Similarly, under model V, the information matrix $\underset{\sim}{I}^{V}(\underset{\sim}{\beta},\underset{\sim}{\lambda},\underset{\sim}{\alpha})$ may be partitioned as above and $\underset{\sim}{A}$, a $p \times p$ symmetric matrix, has elements

$$A_{kl} = \sum_{j=1}^{s}\sum_{i=1}^{n_j} x_{jik}\, x_{jil} \qquad k,\ l=1,\ldots,p,$$

$\underset{\sim}{C} = \left[\underset{\sim}{C}_1 \ \vdots\ \underset{\sim}{C}_2\right]$ where $\underset{\sim}{C}_1$, $\underset{\sim}{C}_2$ are $p \times s$ matrices with elements

$$C_{\lambda\, kj} = \frac{1}{\lambda_j} \sum_{i=1}^{n_j} x_{jik} ,$$

$$C_{\alpha\, kj} = \frac{1}{\alpha_j} \sum_{i=1}^{n_j} x_{jik} \left(L(1,1) - (\log \lambda_j + \underline{\beta}'\underline{x}_{ji})\right),$$

$k = 1,\ldots,p\ ;\ j=1,\ldots,\ m.$

The matrix $\underline{B}$ may be written in the form

$$\begin{bmatrix} \underline{B}_{\lambda} & \vdots & \underline{B}_{\lambda\alpha} \\ \cdots & \cdots & \cdots \\ \underline{B}'_{\lambda\alpha} & \vdots & \underline{B}_{\alpha} \end{bmatrix}$$

where $\underline{B}_{\lambda} = \text{diag}_m \,(n_j/\lambda_j^2)$

$$\underline{B}_{\alpha} = \text{diag}_m \left[\frac{\lambda_2}{\alpha_j} \sum_{i=1}^{n_j} \left(1 + L(1,2) - 2(\log \lambda + \underline{\beta}'\underline{x}_{ji})\, L(1,1) + (\log\lambda + \underline{\beta}'\underline{x}_{ji})^2\right)\right]$$

and

$$\underline{B}_{\lambda\alpha} = \text{diag}_m \left[\frac{1}{\alpha_j\lambda_j} \sum_{i=1}^{n_j} \left(L(1,1) - (\log \lambda_j + \underline{\beta}'\underline{x}_{ji})\right)\right] , \text{ and}$$

$\text{diag}_m\,(a_j)$ denotes the $m \times m$ diagonal matrix with $(j,j)^{th}$ element a_j. Asymptotically $\hat{\underline{\beta}} \sim N(\underline{\beta},\ \underline{M})$ where $\underline{M} = (\underline{A} - \underline{C}'\ \underline{B}^{-1}\ \underline{C})^{-1}$ is as before. Results for models III and VI are obvious special cases of the above.

§5.2 The two and K group cases

Large sample efficiency

As pointed out in §4.5, the model I statistic $\frac{\partial \ell(0)}{\partial \beta}$ in the two group case is identical to the logrank statistic at 2.15. The results at the end of §2.5 indicate that the test based on this statistic is asymptotically fully efficient under random censorship, when the censoring distributions in the two groups are equal and the true distribution of survival time is exponential. Further asymptotic results have also been discussed in that section.

In the K-group case, where

$\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \ldots, x_{i\,K-1})$ $\quad x_{ij} = \begin{cases} 1 & \text{if } i \text{ in group } j \\ 0 & \text{otherwise,} \end{cases}$

the null hypothesis $H_0: \beta_1 = \beta_2 = \ldots = \beta_{K-1} = 0$ can be tested using the statistic given in §4.5 and under H_0

$$\left[\frac{\partial \ell(\underset{\sim}{0})}{\partial \underset{\sim}{\beta}}\right]^1 \underset{\sim}{U}(\underset{\sim}{0})^{-1} \left[\frac{\partial \ell(\underset{\sim}{0})}{\partial \underset{\sim}{\beta}}\right] \overset{\text{asymp}}{\sim} \chi^2_{K-1} \qquad 5.1.$$

Crowley (1973) has extended the above 2 group results under identical conditions showing that the test based on 5.1. is asymptotically fully efficient. Again losses in efficiency occur for unequal censoring distributions.

Small sample power

In the two group case, Lee, Desu and Gehan (1975) using the Monte Carlo procedure of §2.1 have evaluated the small sample power of the single tailed test based on the statistic $\frac{\partial \ell(0)}{\partial \beta}$, treating $\frac{\partial \ell(0)}{\partial \beta} \Big/ \sqrt{U(0)}$ as $N(0,1)$ under $H_0: \beta = 0$. These authors incorporate this test in the comparisons summarized in §2.6. and in both the censored and uncensored situations show that this test compares favourably with a) the F-test when the true distribution of survival time is exponential and b) the F_1-test when the distribution is Weibull. The small sample efficiency of maximum likelihood estimation based on model I to that under model III, in the 2 group case is considered in §5.3 as a special case of the single independent variable situation.

§5.3 A single independent variable

The results of Kalbfleisch

The relative efficiency of the method of estimation based on model I compared to that based on model III has been considered in the single independent variable uncensored case by Kalbfleisch (1974a).

Using the marginal likelihood $L(\beta)$ of β in model III, he evaluates the information $I^{III}(\beta)$ about β contained in the statistic $\underline{A}$, on which this marginal likelihood is based as

$$I^{III}(\beta) = E\left\{-\frac{\partial^2}{\partial\beta^2}\log L(\beta)\right\}.$$

In model I, the information $I^{I}(\beta)$ contained in the rank vector is evaluated at $\beta=0$. The efficiency $\epsilon_n(0)$ of an estimation procedure based on the distribution of the rank vector (model I) compared to one based on the distribution of $\underline{A}$ (marginal likelihood approach to model III) at $\beta=0$ and in a sample of size n is then the ratio of the informations contained in these statistics and

$$\epsilon_n(0) = I^{I}(0)/I^{III}(0) \qquad 5.2.$$

$\epsilon_n(0)$ is tabulated for various values of n and

$$\epsilon(0) = \lim_{n\to\infty} \epsilon_n(0) = 1.$$

For non-zero values of β, $I^{I}(\beta)$ cannot be evaluated analytically and Kalbfleisch obtains an approximation to $I^{I}(\beta)$, valid in the neighbourhood of $\beta=0$, by expanding $\log I^{I}(\beta)$ as a Taylor series about $\beta=0$. The resulting efficiency measure $\epsilon_n(\beta)$ is evaluated for large n to yield an asymptotic measure $\epsilon(\beta)$.

The interpretation of these asymptotic results in terms of two particular estimators is unclear as the methods given by Kalbfleisch in model III are based on the concepts of marginal likelihood (see comments in §4.6 on inferential procedures). These difficulties are overcome in this section by replacing the marginal likelihood $L(\beta)$ by the standard likelihood $L(\beta,\lambda)$. The detailed derivation given, runs parallel to that of Kalbfleisch.

Large sample efficiency in uncensored case

It will be convenient to transform the independent variable x by putting $z_i = x_i - \bar{x}$, for $i=1,\ldots,n$ where $\bar{x} = \sum_{i=1}^{n} x_i/n$.

The results of §5.1 with $p=1$ indicate that the model III information matrix is given by

$$I^{III}(\beta,\lambda) = \begin{bmatrix} n\mu_2 & 0 \\ 0 & n/\lambda^2 \end{bmatrix}$$

where, for $j=1,2,\ldots$, $\mu_j = \frac{1}{n}\sum_{i=1}^{n} z_i^{\,j}$ is the j'th central moment of the finite population $(z_1, z_2, \ldots, z_n)$. It then follows that $\hat\beta_{III}$, the maximum likelihood estimator of β in model III, is distributed asymptotically as $N(\beta, 1/_{n\mu_2})$.

Under Model I, the likelihood function can be written

$$L(\beta) = \prod_{i=1}^{n}\left(\sum_{j=i}^{n} e^{\beta z_j^*}\right)^{-1} \quad \text{where } z_j^* = z_j - \bar z, \; j=1,\ldots,n,$$

and using 4.7., the 2nd derivative of the log likelihood is

$$\frac{\partial^2 \ell(\beta)}{\partial\beta^2} = -\sum_{i=1}^{n}\left\{\left(\sum_{j=i}^{n} z_j^{*2} e^{\beta z_j^*}\right)\left(\sum_{j=i}^{n} e^{\beta z_j^*}\right) - \left(\sum_{j=i}^{n} z_j^* \, e^{\beta z_j^*}\right)^2\right\}\left(\sum_{j=i}^{n} e^{\beta z_j^*}\right)^{-2}$$

$$= -g_2(\beta)$$

where $g_j(\beta) = -\frac{\partial^j \ell(\beta)}{\partial\beta^j}$ $j=1,2$.

Putting $I^I(\beta) = E(g_2(\beta)) = \sum_{\underline{r}} g_2(\beta)\, L(\beta)$, where the summation is over all $n!$ possible rank vectors $\underline{r}$, it follows that asymptotically $\hat\beta_I$, the maximum likelihood estimator of β under model I, has distribution $N(\beta, 1/_{I}I_{(\beta)})$. The asymptotic variance of $\hat\beta_I$ cannot be evaluated explicitly for non-zero β. However at $\beta=0$ relatively simple results may be obtained, where

$$I^I(0) = E_p\left\{\sum_{i=1}^{n}\frac{1}{n-i+1}\sum_{j=i}^{n} z_j^{*2} - \sum_{i=1}^{n}\frac{1}{(n-i+1)^2}\left(\sum_{j=i}^{n} z_j^*\right)^2\right\}$$

and E_p denotes expectation over the permutation distribution of the finite population $(z_1, z_2, \ldots, z_n)$. Since

$$E_P\left(\sum_{j=i}^{n} z_j^{*2}\right) = (n - i + 1)\,\mu_2 \quad \text{and}$$

$$E_P\left\{\left(\sum_{j=i}^{n} z_j^{*}\right)^2\right\} = (n - i + 1)\,\mu_2\left(1 - \frac{n-i}{n+1}\right) .$$

$$I^I(0) = \frac{n\mu_2}{n-1}\sum_{i=1}^{n}\frac{n-i}{n-i+1} = \frac{n\mu_2}{n-1}\, s_n = n\mu_2 + o(n)$$

where

$$s_n = \sum_{i=1}^{n}\frac{n-i}{n-i+1} \quad \text{and } f(n) = o(n) \Longleftrightarrow \lim_{n\to\infty}\frac{f(n)}{n} = 0.$$

At $\beta = 0$, the asymptotic variances of $\hat{\beta}_I$ and $\hat{\beta}_{III}$ are simple inverse functions of the sample size, so that $R_{I,III}(0)$, the asymptotic relative efficiency of $\hat{\beta}_I$ (compared to $\hat{\beta}_{III}$) is 1. Thus the method of estimation, $\hat{\beta}_I$, using model I is asymptotically fully efficient when $\beta=0$ and the connection between estimating efficiency and A.R.E. mentioned earlier implies that a test of H_0: $\beta=0$ based on the distribution of $\hat{\beta}_I$ has A.R.E. equal to 1 when compared to the asymptotically efficient test based on the marginal distribution of $\hat{\beta}_{III}$.

For non-zero β, $I^I(\beta)$ cannot be evaluated analytically and some approximations are needed. Expanding $\log I^I(\beta)$ as a Taylor series about the value $\beta=0$

$$\log I^I(\beta) = \log I^I(0) + \frac{\beta}{I^I(0)}\,\frac{\partial I^I(0)}{\partial\beta}$$

$$+ \frac{\beta^2}{2(I^I(0))^2}\left[I^I(0)\,\frac{\partial^2 I^I(0)}{\partial\beta^2} - \left\{\frac{\partial I^I(0)}{\partial\beta}\right\}^2\right] + \cdots \qquad 5.3$$

Note that the log transformation ensures that $I^I(\beta)$ remains positive for all values of β.

Evaluating term by term,

$$I^I(0) = n\mu_2 + o(n)$$

$$\frac{\partial I^I(0)}{\partial \beta} = E_p\{g_3(0) - g_2(0)g_1(0)\} = o(n)$$

$$\frac{\partial^2 I^I(0)}{\partial \beta^2} = E_p\left[g_4(0) - 2g_3(0)g_1(0) - (g_2(0))^2 + g_2(0)(g_1(0))^2\right]$$

$$= -2n\mu_2^2 + o(n).$$

For details of the calculations leading to these results see Kalbfleisch (1974a). From 5.3 it follows that, in the neighbourhood of $\beta = 0$, $I^I(\beta) = n\mu_2 e^{-\mu_2\beta^2}$ and the asymptotic relative efficiency of $\hat{\beta}_I$ at β is given by

$$E_{I,II}(\beta) = e^{-\mu_2\beta^2}. \qquad 5.4.$$

If the true distribution is Weibull and model II is appropriate

$$\underline{I}^{II}(\beta,\lambda,\alpha) = \left[\begin{array}{c:cc} n\mu_2 & 0 & -\frac{\beta n\mu_2}{\alpha} \\ \hdashline 0 & & \\ -\frac{\beta n\mu_2}{\alpha} & & \underline{B} \end{array}\right] = \left[\begin{array}{c:c} A & \underline{C} \\ \hdashline \underline{C}' & \underline{B} \end{array}\right]$$

where $\underline{B} = [B_{ij}]$ with

$B_{11} = \frac{n}{\lambda^2}$, $B_{12} = \frac{n}{\alpha\lambda}(1 - \omega - \log\lambda) = B_{21}$ and

$B_{22} = \frac{n}{\alpha^2}\left((1 - \omega - \log\lambda)^2 + \frac{\pi^2}{6} + \beta^2\mu^2\right)$ and using the results of §5.1.

$$\underline{C}\,\underline{B}^{-1}\underline{C}' = C_{12}^2 B_{11}(B_{11}B_{22} - B_{12}^2)^{-1} = n\beta^2\mu^2\left(\frac{\pi^2}{6} + \beta^2\mu^2\right)^{-1}$$

so that asymptotically

$\mathrm{var}(\hat{\beta}_{II}) = (A - \underline{C}\,\underline{B}^{-1}\underline{C}')^{-1} = \frac{1}{n\mu_2} + \frac{6\beta^2}{n\pi^2}$, and in the neighbourhood

of $\beta = 0$,

$$R_{I,II}(\beta) = e^{-\mu^2\beta^2}\left(1 + \frac{6\beta^2\mu^2}{\pi^2}\right). \qquad 5.5.$$

Note that $R_{I,II}(0) = 1$ and that, since

$$e^{\mu^2\beta^2} \geq 1 + \frac{6\beta^2\mu_2}{\pi^2} \geq 1,$$

asymptotically and in the neighbourhood of $\beta = 0$

$\text{var}(\hat{\beta}_I) \geq \text{var}(\hat{\beta}_{II}) \geq \text{var}(\hat{\beta}_{III})$ so that

$R_{I,II}(\beta) \geq R_{I,III}(\beta)$.

Note also from the above analysis that

$$R_{II,III}(\beta) = \left(1 + \frac{6\beta^2\mu_2}{\pi^2}\right)^{-1}.$$

These asymptotic efficiencies have been evaluated in table 5.1. for various values of $|\beta|$ with $\mu_2 = 1$. Note that $R_{I,III}(\beta)$ exceeds 0.75 for $|\beta| < 0.5364$. In the 2 group case with $z_i = \begin{cases}-1 & \text{group 1}\\ +1 & \text{group 2}\end{cases}$ and equal numbers in each group this corresponds to a ratio of approximately 3:1 for the failure rates in the 2 groups.

Kalbfleisch, using computer simulation, indicates that 5.4. is a good approximation when the distribution of the finite population $z_1, z_2, \ldots, z_n$ is symmetric but discrepancies occur if the distribution is skew. To investigate these claims and to assess the effects of censoring on the efficiency results, these and other simulations are given here. 1000 observations were randomly generated (500 in each group) for model III with $\lambda=1$, $p=1$ and $z_i = \begin{cases}-1 & \text{group 1}\\ +1 & \text{group 2}\end{cases}$ for values of $\beta = -0.5(0.1)\,0.5$.
On each occasion, estimates $\hat{\beta}_I$ and $\hat{\beta}_{III}$ of β were obtained using a Newton-Raphson method. This procedure was repeated 20 times and for each β value, sample variances of $\hat{\beta}_I$ and $\hat{\beta}_{III}$ calculated. The results are given in table 5.2A). Two further situations were also considered. Firstly, the finite population $(z_1, z_2, \ldots, z_{1000})$ consituted a random sample from a standard normal distribution and secondly from a unit exponential distribution. In each case the finite population was standardised

<u>Table 5.1</u>.

A.R.E.'s $R_{I,III}(\beta)$, $R_{I,II}(\beta)$ and $R_{II,III}(\beta)$ when $\mu_2 = 1$ for $|\beta| = 0(0.05)0.5$.

$\|\beta\|$	$R_{I,III}(\beta)$	$R_{I,II}(\beta)$	$R_{II,III}(\beta)$
0	1	1	1
0.05	0.998	0.999	0.998
0.10	0.990	0.996	0.994
0.15	0.978	0.991	0.987
0.20	0.961	0.984	0.976
0.25	0.939	0.975	0.963
0.30	0.914	0.964	0.948
0.35	0.885	0.951	0.931
0.40	0.852	0.935	0.911
0.45	0.817	0.917	0.890
0.50	0.779	0.897	0.868
0.5364	0.750	0.881	0.851

<u>Table 5.2</u>. Sample means and variances of $\bar{\delta}_I$ and $\bar{\delta}_{III}$ and estimated A.R.E. of $\bar{\delta}_I$ compared to $\bar{\delta}_{III}$ for $\delta = -0.5(0.1)0.5$

A) Two group case. (20 simulations)

δ	$R_{I,III}(\delta)$ using 5.4	$\bar{\delta}_{III}$ average	$\bar{\delta}_{III}$ sample variance($\times 10^2$)	$\bar{\delta}_I$ average	$\bar{\delta}_I$ sample variance($\times 10^2$)	estimated $R_{I,III}(\delta)$
−0.5	0.779	−0.4829	0.1594	−0.4848	0.1758	0.906
−0.4	0.852	−0.3829	0.1594	−0.3848	0.1694	0.941
−0.3	0.914	−0.2829	0.1594	−0.2851	0.1591	1.001
−0.2	0.961	−0.1829	0.1594	−0.1842	0.1588	1.004
−0.1	0.990	−0.0829	0.1594	−0.0836	0.1603	0.994
0.0	1.000	0.0171	0.1594	0.0170	0.1611	0.989
0.1	0.990	0.1171	0.1594	0.1182	0.1590	1.002
0.2	0.961	0.2171	0.1594	0.2196	0.1661	0.959
0.3	0.914	0.3171	0.1594	0.3211	0.1757	0.907
0.4	0.852	0.4171	0.1594	0.4230	0.1807	0.882
0.5	0.779	0.5171	0.1594	0.5251	0.1897	0.840

B) Independent variables observations from N(0,1). (10 simulations)

s	$R_{I,III}(s)$ using 5.5.	s_{III} average	s_{III} sample variance($\times 10^2$)	s_I average	s_I sample variance($\times 10^2$)	estimated $R_{I,III}(s)$
-0.5	0.779	-0.4987	0.1127	-0.5034	0.2012	0.560
-0.4	0.852	-0.3987	0.1127	-0.4014	0.1800	0.626
-0.3	0.914	-0.2987	0.1127	-0.3011	0.1537	0.733
-0.2	0.961	-0.1987	0.1127	-0.2009	0.1313	0.858
-0.1	0.990	-0.0987	0.1127	-0.1001	0.1110	1.015
0.0	1.000	0.0013	0.1127	0.0015	0.1061	1.062
0.1	0.990	0.1013	0.1127	0.1020	0.1060	1.063
0.2	0.961	0.2013	0.1127	0.2026	0.1073	1.050
0.3	0.914	0.3013	0.1127	0.3044	0.1143	0.986
0.4	0.852	0.4013	0.1127	0.4069	0.1188	0.948
0.5	0.779	0.5013	0.1127	0.5084	0.1265	0.891

C) Independent variables observations from unit exponential distribution.

(10 simulations)

θ	$R_{I,III}(\theta)$ using 5.4.	$\bar{\theta}_{III}$ average	θ_{III} sample variance($\times 10^2$)	$\bar{\theta}_I$ average	θ_I sample variance($\times 10^2$)	estimated $R_{I,III}(\theta)$
-0.5	0.779	-0.5088	0.1422	-0.5056	0.2300	0.618
-0.4	0.852	-0.4088	0.1422	-0.4093	0.2392	0.594
-0.3	0.914	-0.3088	0.1422	-0.3140	0.2069	0.687
-0.2	0.961	-0.2088	0.1422	-0.2122	0.1582	0.899
-0.1	0.990	-0.1088	0.1422	-0.1106	0.1619	0.878
0.0	1.000	-0.0088	0.1422	-0.0092	0.1428	0.996
0.1	0.990	0.0912	0.1422	0.0919	0.1468	0.969
0.2	0.961	0.1912	0.1422	0.1932	0.1515	0.939
0.3	0.914	0.2912	0.1422	0.2954	0.1611	0.883
0.4	0.852	0.3912	0.1422	0.3972	0.1720	0.827
0.5	0.779	0.4912	0.1422	0.4978	0.1721	0.827

by subtraction of the sample mean and division by the sample standard deviation to ensure that $\mu_1 = 0$ and $\mu_2 = 1$. Results from 10 simulations are given in tables 5.2.B) and 5.2.C) respectively.

In the three cases considered, the expression 5.4. is in reasonably close agreement with the A.R.E. obtained from computer simulation, although the results are generally less stable than those of Kalbfleisch. Particularly in the two latter cases, the estimated values of $R_{I,III}(\beta)$ lack symmetry about $\beta=0$. The reason for this is not clear.

Effects of censoring

To assess the effect of censoring on the efficiency results of the previous section the simulation study was extended. At each simulation, having generated a random sample of 1000 observations from the appropriate form of model III, the clinical trial situation was simulated (as in §2.1) by assuming that individuals enter the trial at a constant rate in the interval $(0, T^*)$ and termination of the trial at T^* gave a set of censored and uncensored observations. T^* was chosen so that the expected proportion of censorings was successively 0.3 and 0.6. Thus for each of the 3 finite populations $(z_1, z_2, \ldots, z_{1000})$ and for each β value considered, T^* was the solution of

$$\frac{1}{n} \sum_{i=1}^{1000} e^{-\beta z_i}(1 - \exp(-T^* e^{\beta z_i})) = p$$

for $p = 0.3$ and 0.6. Tables 5.3 and 5.4 present the results for $p=0.3$ and $p = 0.6$ respectively.

<u>Table 5.3</u>. Sample means and variances of $\hat{\beta}_I$ and $\hat{\beta}_{III}$ and estimated A.R.E. of $\hat{\beta}_I$ compared to $\hat{\beta}_{III}$ for β = -0.5(0.1)0.5 and 30% censoring.

A) Two group case (20 simulations)

β	$\hat{\beta}_{III}$ average	$\hat{\beta}_{III}$ sample variance($\times 10^2$)	$\hat{\beta}_I$ average	$\hat{\beta}_I$ sample variance($\times 10^2$)	estimated $R_{I,III}(\beta)$
-0.5	-0.4812	0.2136	-0.4804	0.2541	0.840
-0.4	-0.3808	0.2164	-0.3815	0.2472	0.876
-0.3	-0.2804	0.1828	-0.2802	0.2034	0.899
-0.2	-0.1828	0.1184	-0.1827	0.1241	0.954
-0.1	-0.0817	0.1535	-0.0815	0.1570	0.978
0.0	0.0275	0.1570	0.0277	0.1565	1.003
0.1	0.1196	0.1860	0.1197	0.1872	0.994
0.2	0.2255	0.1538	0.2259	0.1543	0.997
0.3	0.3220	0.1777	0.3218	0.1783	0.997
0.4	0.4232	0.1573	0.4242	0.1609	0.978
0.5	0.5233	0.1987	0.5247	0.2004	0.991

b) Independent variables observations from N(0,1) distribution. (10 simulations)

β	$\hat{\beta}_{III}$ average	$\hat{\beta}_{III}$ sample variance($\times 10^2$)	$\hat{\beta}_{I}$ average	$\hat{\beta}_{I}$ sample variance($\times 10^2$)	estimated $e_{I,III}(\beta)$
-0.5	-0.5057	0.3112	-0.5066	0.3115	0.999
-0.4	-0.4136	0.2936	-0.4151	0.3200	0.918
-0.3	-0.3046	0.3119	-0.3048	0.3299	0.946
-0.2	-0.2058	0.2644	-0.2058	0.2731	0.968
-0.1	-0.1059	0.1881	-0.1075	0.1931	0.974
0.0	0.0070	0.2748	0.0058	0.2724	1.009
0.1	0.1071	0.2202	0.1077	0.2077	1.061
0.2	0.2174	0.2359	0.2168	0.2037	1.158
0.3	0.3063	0.1860	0.3065	0.1932	0.963
0.4	0.4108	0.1744	0.4110	0.1442	1.210
0.5	0.5076	0.2063	0.5098	0.2091	0.987

C) Independent variables observations from unit exponential distribution (10 simulations)

β	$\bar{\beta}_{III}$ average	$\bar{\beta}_{III}$ sample variance(×10²)	$\bar{\beta}_I$ average	$\bar{\beta}_I$ sample variance(×10²)	estimated $R_{I,III}(\beta)$
-0.5	-0.4939	0.4067	-0.4986	0.4100	0.992
-0.4	-0.3998	0.3135	-0.4003	0.3168	0.990
-0.3	-0.2943	0.1782	-0.2958	0.1792	0.994
-0.2	-0.2020	0.1606	-0.2030	0.1485	1.082
-0.1	-0.1005	0.2044	-0.0996	0.1969	1.038
0.0	-0.0069	0.3000	-0.0074	0.3076	0.975
0.1	0.0903	0.1863	0.0904	0.1976	0.943
0.2	0.1866	0.2163	0.1872	0.2289	0.945
0.3	0.2902	0.1580	0.2921	0.2087	0.757
0.4	0.3964	0.1078	0.3993	0.1289	0.837
0.5	0.4972	0.1236	0.5009	0.1557	0.794

Table 5.4. Sample means and variances of $\hat{\beta}_I$ and $\hat{\beta}_{III}$ and estimated A.R.E. of $\hat{\beta}_I$ compared to $\hat{\beta}_{III}$ for $\beta = -0.5(0.1)\ 0.5$ and 60% censoring.

A) Two group case (20 simulations)

β	$\hat{\beta}_{III}$ average	$\hat{\beta}_{III}$ sample variance($\times 10^2$)	$\hat{\beta}_I$ average	$\hat{\beta}_I$ sample variance($\times 10^2$)	estimated $R_{I,III}(\beta)$
-0.5	-0.4626	0.3476	-0.4627	0.3634	0.956
-0.4	-0.3715	0.3015	-0.3705	0.3089	0.976
-0.3	-0.2640	0.3148	-0.2631	0.3172	0.992
-0.2	-0.1751	0.1764	-0.1749	0.1719	1.027
-0.1	-0.0678	0.4003	-0.0676	0.3952	1.013
0.0	0.0282	0.2773	0.0276	0.2778	0.998
0.1	0.1308	0.3300	0.1302	0.3296	1.001
0.2	0.2290	0.1545	0.2283	0.1466	1.054
0.3	0.3338	0.2323	0.3319	0.2280	1.019
0.4	0.4360	0.3434	0.4344	0.3403	1.009
0.5	0.5272	0.2332	0.5236	0.2292	1.017

B) Independent variable observations from N(0,1) distribution

(10 simulations)

β	$\bar{\beta}_{III}$ average	β_{III} sample variance($\times 10^2$)	$\bar{\beta}_I$ average	β_I sampled variance($\times 10^2$)	estimated $R_{I,III}(\beta)$
-0.5	-0.5036	0.3270	-0.5078	0.3272	0.999
-0.4	-0.3893	0.5869	-0.3911	0.5799	1.012
-0.3	-0.2952	0.3416	-0.2964	0.3302	1.034
-0.2	-0.1737	0.4937	-0.1739	0.4878	1.012
-0.1	-0.0905	0.3774	-0.0918	0.3695	1.021
0.0	0.0180	0.5277	0.0186	0.5254	1.004
0.1	0.0989	0.4130	0.0998	0.4076	1.013
0.2	0.2302	0.3382	0.2313	0.3384	0.999
0.3	0.3268	0.5139	0.3258	0.5061	1.015
0.4	0.4181	0.2264	0.4179	0.2129	1.063
0.5	0.5186	0.2710	0.5180	0.2577	1.052

C) Independent observations from unit exponential distribution

(10 simulations)

δ	$\hat{\delta}_{III}$ average	$\hat{\delta}_{III}$ sample variance($\times 10^2$)	$\hat{\delta}_{I}$ average	$\hat{\delta}_{I}$ sample variance($\times 10^2$)	estimated [illegible]
-0.5	-0.4625	0.7823	-0.4630	0.7927	0.987
-0.4	-0.3745	0.5730	-0.3742	0.5632	1.017
-0.3	-0.2622	0.2836	-0.2617	0.2858	0.992
-0.2	-0.1756	0.6420	-0.1761	0.6484	0.990
-0.1	-0.0879	0.3471	-0.0871	0.3454	1.005
0.0	0.0253	0.3053	0.0245	0.3060	0.998
0.1	0.0981	0.4214	0.0977	0.4111	1.025
0.2	0.1969	0.2952	0.1968	0.2862	1.031
0.3	0.2968	0.2848	0.2973	0.2887	0.986
0.4	0.4153	0.1896	0.4169	0.1917	0.989
0.5	0.4956	0.2457	0.4956	0.2395	1.026

These results clearly indicate that, in general, the large sample efficiency of $\bar{\beta}_I$ compared to $\hat{\beta}_{III}$ increases when the data is subject to censorship. In addition greater severity of censoring improves the efficiency even further. This general trend is particularly marked in the two situations where the independent variables are randomly generated samples from standard normal and unit exponential distributions. In the 2 group case, the estimated value of $R_{I,III}(\beta)$ is greater than 99% for all β values considered with censoring proportion 0.6. Corresponding lower bounds in the standard normal and unit exponential situations are 99% and 98.5% respectively.

Small sample considerations

In a sample of size n, Kalbfleisch suggests that the relative efficiency of $\bar{\beta}_I$ compared to $\hat{\beta}_{III}$ is given by

$$R_{I,III,n}(\beta) = \epsilon_n(0)e^{-\frac{1}{2}\beta^2} \qquad 5.6.$$

where $\epsilon_n(0) = \frac{n+1}{n(n-1)} s_n$ is the value of the information ratio obtained at 5.2. by Kalbfleisch. The validity of this approximation is assessed by the above author using computer simulation although details are not given. Tables 5.5A) and 5.5B) present computer simulated estimates (obtained as before) of this relative efficiency in the 2 group case for various values of n, with no censoring and 30% censoring respectively. The expression 5.6. is also evaluated in tables 5.5A).

<u>Table 5.5.A)</u> Relative efficiency $R_{I\_III\_n}(\theta)$ using 5.6 for $\mu_2=1$ and $\theta= -0.5(0.1)0.5$ with estimated values obtained by simulation in the 2 group case and no censoring.

n / θ	20 using simulated	20 5.6 estimate	30 using simulated	30 5.6 estimate	[illegible] using simulated	[illegible] 5.6 estimate	50 using simulated	50 5.6 estimate	80 using simulated	80 5.6 estimate	110 using simulated	110 5.6 estimate
-0.5	0.706	0.568	0.722	0.593	0.731	0.588	0.738	0.641	0.749	0.716	0.755	0.791
-0.4	0.772	0.537	0.790	0.666	0.800	0.648	0.807	0.697	0.819	0.794	0.826	0.866
-0.3	0.828	0.565	0.847	0.735	0.858	0.710	0.866	0.746	0.879	0.854	0.886	0.916
-0.2	0.871	0.567	0.890	0.756	0.902	0.753	0.910	0.823	0.924	0.843	0.931	0.914
-0.1	0.897	0.597	0.917	0.800	0.929	0.777	0.938	0.803	0.952	0.866	0.960	0.912
0.0	0.906	0.752	0.927	0.753	0.939	0.779	0.947	0.857	0.962	0.878	0.969	0.920
0.1	0.897	0.791	0.917	0.761	0.929	0.754	0.938	0.851	0.952	0.881	0.960	0.892
0.2	0.871	0.738	0.890	0.737	0.902	0.747	0.910	0.813	0.924	0.871	0.931	0.831
0.3	0.828	0.673	0.847	0.664	0.858	0.736	0.866	0.777	0.879	0.867	0.886	0.773
0.4	0.772	0.671	0.790	0.637	0.800	0.660	0.807	0.765	0.819	0.819	0.826	0.725
0.5	0.706	0.569	0.722	0.598	0.731	0.555	0.738	0.747	0.749	0.756	0.755	0.705
no. of simulations	100		100		100		100		70		50	

B) Estimated values of relative efficiency for $\mu_2=1$ and $\beta=-0.5(0.1)0.5$ obtained by simulation. 2 group case with 30% censoring.

β \ n	20	30	40	50	80
-0.5	0.672	0.726	0.671	0.746	0.817
-0.4	0.668	0.771	0.743	0.797	0.829
-0.3	0.641	0.833	0.717	0.766	0.955
-0.2	0.688	0.772	0.888	0.765	0.943
-0.1	0.715	0.755	0.822	0.898	0.932
0.0	0.851	0.756	0.825	0.891	0.921
0.1	0.862	0.704	0.917	0.952	1.037
0.2	0.864	0.888	0.863	0.953	0.843
0.3	0.823	0.713	0.838	0.900	0.814
0.4	0.707	0.774	0.678	0.890	0.898
0.5	0.712	0.688	0.703	0.811	0.830

The simulation results suggest that 5.6. is an overestimate of the relative efficiency in the 2 group case, although the approach to full efficiency is at a comparable rate. The results in table 5.5B) again indicate that the relative efficiency increases when censoring is imposed.

§5.4. Two independent variables

Introduction

The situation of two independent variables considered in this section is one that frequently occurs in medical statistics, where the factor of special interest is perhaps treatment group, while the second independent variable is some other factor, not of primary interest, such as age or sex etc.

Large sample efficiency in uncensored case

The calculations involved in the two variable uncensored case are a natural extension of those in §5.3. For $j=1,2$, let

$z_{ij} = x_{ij} - x_j\ ,\quad i=1,\dots,n,$

where $\underline{x}_i = (x_{i1}, x_{i2})$ and $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$

Under model III

$$\underline{I}^{III}(\beta_1, \beta_2, \lambda) = \left[\begin{array}{cc:c} \multicolumn{2}{c:}{\underline{A}} & 0 \\ \hdashline 0 & 0 & n/\lambda^2 \end{array}\right] \quad \text{where } \underline{A} = n\begin{bmatrix} \mu_{2,0} & \mu_{1,1} \\ \mu_{1,1} & \mu_{0,2} \end{bmatrix}$$

and $\mu_{j,k} = \frac{1}{n}\sum_{i=1}^{n} z_{i1}^{\,j}\ z_{i2}^{\,k}$. Note that $\mu_{0,1} = \mu_{1,0} = 0$.

so that asymptotically

$$\hat{\beta}_{III} \sim N(\beta, A^{-1})$$

and $\hat{\beta}_{III,1}$ has asymptotic marginal distribution $N\left(\beta_1, \frac{\mu_{0,2}}{n(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)}\right)$

The likelihood function under model I in the uncensored case is

$$L(\beta_1,\beta_2) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta_1 z_{i1}^* + \beta_2 z_{i2}^*}}{\sum_{j=i}^{n} e^{\beta_1 z_{j1}^* + \beta_2 z_{j2}^*}} \right\}$$

where $z_{ij}^* = x_{ij}^* - \bar{x}_j$ $\quad i=1,\ldots,n;\ j=1,2.$

The 2nd partial derivatives of the log likelihood are, using 4.7.,

$$\frac{\partial^2 l(\beta_1,\beta_2)}{\partial\beta_\ell\,\partial\beta_k} = -\sum_{i=1}^{n}\left\{\left[\left(\sum_{j=i}^{n} z_{j\ell}^* z_{jk}^* e^{\beta_1 z_{j1}^* + \beta_2 z_{j2}^*}\right)\left(\sum_{j=i}^{n} e^{\beta_1 z_{j1}^* + \beta_2 z_{j2}^*}\right)\right]\right.$$

$$\left. -\left(\sum_{j=i}^{n} z_{j\ell}^* e^{\beta_1 z_{j1}^* + \beta_2 z_{j2}^*}\right)\left(\sum_{j=i}^{n} z_{jk}^* e^{\beta_1 z_{j1}^* + \beta_2 z_{j2}^*}\right)\left[\sum_{j=i}^{n} e^{\beta_1 z_{j1}^* + \beta_2 z_{j2}^*}\right]^{-2}\right\},$$

$$= \begin{cases} -g_{2,0}(\beta_1,\beta_2) & \ell=k=1, \\ -g_{0,2}(\beta_1,\beta_2) & \ell=k=2, \\ -g_{1,1}(\beta_1,\beta_2) & \ell=1, k=2 \text{ or } \ell=2, k=1. \end{cases}$$

where $g_{i,j}(\beta_1,\beta_2) = \dfrac{-\partial^{i+j} l(\beta_1,\beta_2)}{\partial\beta_1^{\,i}\,\partial\beta_2^{\,j}}$ $\quad i,j=1,2,\ldots$

$$g_{i,0}(\beta_1,\beta_2) = -\frac{\partial^i l(\beta_1,\beta_2)}{\partial\beta_1^{\,i}}, \quad g_{0,i}(\beta_1,\beta_2) = -\frac{\partial^i l(\beta_1,\beta_2)}{\partial\beta_2^{\,i}} \quad i=1,2,\ldots$$

Putting

$$I^I_{11}\ (\beta_1,\beta_2) = E\ \{g_{2,0}(\beta_1,\beta_2)\} = \sum_{\underline{x}} g_{2,0}(\beta_1,\beta_2)\ L\ (\beta_1,\beta_2),$$

$$I^I_{12}\ (\beta_1,\beta_2) = E\ \{g_{1,1}(\beta_1,\beta_2)\} = \sum_{\underline{x}} g_{1,1}(\beta_1,\beta_2)\ L\ (\beta_1,\beta_2)$$

$$= I^I_{2,1}(\beta_1,\beta_2),$$

$$I^I_{2,2}(\beta_1,\beta_2) = E(g_{0,2}(\beta_1,\beta_2)) = \sum_{\underline{x}} g_{0,2}(\beta_1,\beta_2)\ L(\beta_1,\beta_2),$$

the maximum likelihood estimator $\hat{\underline{\beta}}_I$ of $\underline{\beta} = (\beta_1,\beta_2)'$ has asymptotic distribution $N(\underline{\beta},\ \underline{I}^I(\beta_1,\beta_2)^{-1})$ where $\underline{I}^I(\beta_1,\beta_2) = [I_{ij}{}^I(\beta_1,\beta_2)]$ and the asymptotic marginal distribution of $\hat\beta_{I1}$ is $N(\beta_1,\ V_I(\beta_1,\beta_2))$, where

$$V_I(\beta_1,\beta_2) = I^I{}_{12}(\beta_1,\beta_2)\ \{I^I_{11}(\beta_1,\beta_2)\ I^I_{22}(\beta_1,\beta_2) - I^I_{12}(\beta_1,\beta_2)^2\}^{-1}.$$

$V_I(\beta_1,\beta_2)$ cannot, in general, be evaluated analytically but again relatively simple results may be obtained at $\beta_1 = \beta_2 = 0$.

$$I^I_{11}(0,0) = E_p\ \{g_{2,0}(0,0)\} = \frac{n}{n-1}\ s_n\ \mu_{2,0}$$

$$I^I_{12}(0,0) = E_p\ \{g_{1,1}(0,0)\} = \frac{n}{n-1}\ s_n \mu_{1,1}$$

$$I^I_{22}(0,0) = E_p\ \{g_{0,2}(0,0)\} = \frac{n}{n-1}\ s_n\ \mu_{0,2}$$

and

$$V_I(0,0) = \frac{n-1}{n s_n}\ \frac{\mu_{0,2}}{(\mu_{2,0}\ \mu_{0,2} - \mu^2_{1,1})} = \frac{\mu_{0,2}}{n(\mu_{2,0}\ \mu_{0,2} - \mu^2_{1,1}) + o(n)}$$

so that at $\beta_1 = \beta_2 = 0$, the asymptotic relative efficiency $R_{I,III}(0,0)$ of $\hat\beta_{I1}$ compared to $\hat\beta_{IIII}$ is equal to 1. This implies that at $\beta_2 = 0$, a test of H_0: $\beta_1 = 0$ based on the marginal distribution of $\hat\beta_{I1}$ is asymptotically fully efficient.

As in the single independent variable case an approximation to the asymptotic variance of $\hat\beta_{II}$, valid in the neighbourhood of $(\beta_1,\beta_2) = (0,0)$, may be obtained for non-zero β_1,β_2 by expanding $\log V_I(\beta_1,\beta_2)$ as a Taylor series about $(0,0)$. The log transformation again ensures that $V_I(\beta_1,\beta_2)$ is strictly positive for all β_1 and β_2.

$$\log V_I(\beta_1,\beta_2) = \log V_I(0,0) + \frac{\beta_1}{V_I(0,0)}\frac{\partial V_I(0,0)}{\partial\beta_1} + \frac{\beta_2}{V_I(0,0)}\frac{\partial V_I(0,0)}{\partial\beta_2}$$

$$+ \frac{\beta_1^2}{2[V_I(0,0)]^2}\left\{V_I(0,0)\frac{\partial^2 V_I(0,0)}{\partial\beta_1^2} - \left[\frac{\partial V_I(0,0)}{\partial\beta_1}\right]^2\right\}$$

$$+ \frac{\beta_1\beta_2}{2[V_I(0,0)]^2}\left\{V_I(0,0)\frac{\partial^2 V_I(0,0)}{\partial\beta_1\partial\beta_2} - \left[\frac{\partial V_I(0,0)}{\partial\beta_1}\right]\left[\frac{\partial V_I(0,0)}{\partial\beta_2}\right]\right\}$$

$$+ \frac{\beta_2^2}{2[V_I(0,0)]^2}\left\{V_I(0,0)\frac{\partial^2 V_I(0,0)}{\partial\beta_2^2} - \left[\frac{\partial V_I(0,0)}{\partial\beta_2}\right]^2\right\} + \ldots$$

It may be shown (the details are given in Appendix B) that in the neighbourhood of $(\beta_1,\beta_2) = (0,0)$ and for large n

$$V_I(\beta_1,\beta_2) \approx \frac{\mu_{0,2}}{n(\mu_{2,0}\,\mu_{0,2} - \mu_{1,1}^2)}\exp\left\{\beta_1^2\frac{(\mu_{2,0}\,\mu_{0,2}-\mu_{1,1}^2)}{\mu_{0,2}}\right\}$$

and hence

$$R_{I,III}(\beta_1,\beta_2) \approx \exp\left\{-\frac{\beta_1^2}{\mu_{0,2}}(\mu_{2,0}\,\mu_{0,2}-\mu_{1,1}^2)\right\} \qquad 5.7.$$

Note that $R_{I,III}(0,\beta_2) = 1$, so that the estimator $\hat\beta_{II}$ is asymptotically fully efficient and a test of H_0: $\beta_1 = 0$ based on the marginal distribution of $\hat\beta_{I1}$ has asymptotic relative efficiency equal to 1 compared to the efficient test based on $\hat\beta_{III}$. These results are independent of the value of β_2, although it should be noted that use of the Taylor expansion implies that 5.7 is valid only for (β_1,β_2) in the neighbourhood of (0,0).

Putting $\mu_{1,1}=0$ in 5.7,

$R_{I,III}(\beta_1,\beta_2) = \exp\left(-\beta_1^2\mu_{2,0}\right)$

which is identical to the asymptotic relative efficiency at 5.4. for the single variable case. This result supports the claim, made by Kalbfleisch (1974), that 5.4. is a good guide in multiparameter problems provided that the independent variables are nearly uncorrelated.

Under model II the elements of the information matrix $\underset{\sim}{I}^{II}(\beta_1,\beta_2,\lambda,\sigma)$ are readily available. $\underset{\sim}{A}$ is as in model III, $\underset{\sim}{B} = [b_{ij}]$, $\underset{\sim}{C} = [c_{ij}]$ where

$$B_{11} = \frac{n}{\sigma^2}, \quad B_{12} = -\frac{n}{\sigma^2}(1-\gamma-\log\lambda) = B_{21}$$

$$B_{22} = \frac{n}{\sigma^2}\left\{(1-\gamma-\log\lambda)^2 + \frac{\pi^2}{6} + \beta_1^2\mu_{2,0} + \beta_2^2\mu_{2,0} + 2\beta_1\beta_2\mu_{1,1}\right\},$$

$$C_{11} = 0 = C_{21}, \quad C_{12} = -\frac{n}{\sigma}(\beta_1\mu_{2,0}+\beta_2\mu_{1,1}), \quad C_{22} = -\frac{n}{\sigma}(\beta_1\mu_{1,1}+\beta_2\mu_{0,2})$$

so that $\underset{\sim}{M} = (\underset{\sim}{A} - \underset{\sim}{C}\,\underset{\sim}{B}^{-1}\underset{\sim}{C}')^{-1}$

$$= \frac{1}{n}\left(\frac{\pi^2}{6}+\beta_1^2\mu_{2,0}+\beta_2^2\mu_{0,2}+2\beta_1\beta_2\mu_{1,1}\right)\begin{bmatrix}\frac{\pi^2}{6}\mu_{2,0}+\beta_2^2\mu & \frac{\pi^2}{6}\mu_{1,1}-\beta_1\beta_2\mu \\ \frac{\pi^2}{6}\mu_{1,1}-\beta_1\beta_2\mu & \frac{\pi^2}{6}\mu_{0,2}+\beta_1^2\mu\end{bmatrix}^{-1}$$

where $\mu = (\mu_{2,0}\,\mu_{0,2} - \mu_{1,1}^2)$. It follows that asymptotically

$$\operatorname{var}(\hat{\beta}_{II,1}) = \frac{1}{n}\left[\frac{\mu_{0,2}}{\mu} + \frac{6\beta_1^2}{\pi^2}\right] \text{ and thus}$$

$$R_{I,II}(\beta_1,\beta_2) = \left[1 + \frac{6\beta_1^2\mu}{\pi^2\mu_{0,2}}\right]\exp\left[-\frac{\beta_1^2\mu}{\mu_{0,2}}\right]$$

in the neighbourhood of $(\beta_1,\beta_2) = (0,0)$.

Note also that

$$R_{II,III}(\theta_1,\theta_2) = \left[1 + \frac{\theta_1\theta_2}{\sigma^2\,\mu_{2,0}}\right].$$

$R_{I,II}(\theta_1,\theta_2)$ and $R_{II,III}(\theta_1,\theta_2)$ coincide with the corresponding expressions for the single independent variable case when $\mu_{1,1}=0$ and the comments associated with $R_{I,III}(\theta_1,\theta_2)$ concerning this reduction apply here. Table 5.6. evaluates these asymptotic efficiencies for various values of $\mu_{1,1}$ and θ_1 for the case $\mu_{2,0} = \mu_{0,2} = 1$. Results for $\mu_{1,1} = 0$ are as in table 5.1.

§5.5. Within strata models

Introduction

The large sample efficiency results of §5.3. and §5.4. in the uncensored case are extended in this section to the within strata models IV, V and VI.

Single independent variable case

It will again be convenient to transform linearly the independent variable by defining, for $j=1,\ldots,s$

$$z_{ji} = x_{ji} - \bar{x}_j \quad i=1,\ldots,n_j \quad \text{where } \bar{x}_j = \sum_{i=1}^{n_j} x_{ji} \Big/ n_j .$$

Let $\mu_{(j)r} = \sum_{i=1}^{n_j} z_{ji}^r \Big/ n_j$ $j=1,\ldots,s$; $r = 1,2,\ldots$. Denote the r^{th} central moment of the finite population $\{x_{j1},\ldots,x_{jn_j}\}$ within the j^{th} strata. Note that

$$\mu_{(1)1} = \mu_{(2)1} = \ldots = \mu_{(s)1} = 0.$$

Table 5.6. A.R.E's $R_{I,III}(\beta_1,\beta_2)$, $R_{I,II}(\beta_1,\beta_2)$ and $R_{II,III}(\beta_1,\beta_2)$ when $\mu_{2,0} = \mu_{0,2} = 1$ for $|\mu_{1,1}| = 0.2(0.2)0.8$ and $|\beta_1| = 0(0.1)0.5$.

$\|\mu_{1,1}\|$	$\|\beta_1\|$	$R_{I,III}(\beta_1,\beta_2)$	$R_{I,II}(\beta_1,\beta_2)$	$R_{II,III}(\beta_1,\beta_2)$
0.2	0	1	1	1
	0.1	0.990	0.996	0.994
	0.2	0.962	0.985	0.977
	0.3	0.917	0.965	0.950
	0.4	0.858	0.938	0.915
	0.5	0.787	0.901	0.873
0.4	0	1	1	1
	0.1	0.992	0.997	0.995
	0.2	0.967	0.987	0.980
	0.3	0.927	0.970	0.956
	0.4	0.874	0.946	0.924
	0.5	0.811	0.914	0.887
0.6	0	1	1	1
	0.1	0.994	0.997	0.996
	0.2	0.975	0.990	0.985
	0.3	0.944	0.977	0.966
	0.4	0.903	0.959	0.941
	0.5	0.852	0.935	0.911
0.8	0	1	1	1
	0.1	0.996	0.999	0.998
	0.2	0.986	0.994	0.991
	0.3	0.968	0.987	0.981
	0.4	0.944	0.977	0.966
	0.5	0.914	0.964	0.948

In addition let $\mu_r = \frac{1}{n} \sum_{j=1}^{s} \sum_{i=1}^{n_j} z_{ji}^r = \sum_{j=1}^{s} q_j \, \mu_{(j)r}$. $r=1,2,\ldots$

where $q_j = n_j/n$ $j=1,\ldots,s$ and $n = \sum_{j=1}^{s} n_j$ is the total sample size.

Under model VI, the information matrix is then given by

$$\underline{I}^{IV}(\beta,\underline{\lambda}) = \begin{bmatrix} n\,\mu_2 & \vdots & \underset{1\times s}{\underline{0}} \\ \cdots & \cdots & \cdots \\ \underset{s\times 1}{\underline{0}} & \vdots & \mathrm{diag}_s(n_j/\lambda_j^2) \end{bmatrix}$$

and it follows that asymptotically,

$$\hat{\beta}_{VI} \sim N(\beta,\ 1/_{n\mu_2}).$$

Similarly, under model V the results of §5.1. indicate that the information matrix may be written as

$$\underline{I}^{IV}(\underline{\beta},\underline{\lambda},\underline{\alpha}) = \begin{bmatrix} A & \vdots & \underset{1\times s}{\underline{0}} & \vdots & \underline{C}_\alpha \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \underset{s\times 1}{\underline{0}} & \vdots & \underline{B}_\lambda & \vdots & \underline{B}_{\lambda\alpha} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \underline{C}'_\alpha & \vdots & \underline{B}'_{\lambda\alpha} & \vdots & \underline{B}_\alpha \end{bmatrix}$$

where $A = n\mu_2$,

$$\underline{C}_\alpha = \left[-\frac{\beta n_1 \mu_{(1)2}}{\alpha_1},\ -\frac{\beta n_2 \mu_{(2)2}}{\alpha_2},\ \ldots,\ -\frac{\beta n_s \mu_{(s)2}}{\alpha_s} \right] ,$$

$$\underline{B}_\lambda = \mathrm{diag}_s\ (n_j/\lambda_j^2).$$

$$B_{\lambda\alpha} = \text{diag}_s \left[\frac{n_j}{\alpha_j \lambda_j} \left\{ 1 - \omega - \log \lambda_j \right\} \right].$$

$$B_{\alpha} = \text{diag}_s \left(\frac{n_j}{\alpha_j^2} \left[\left\{ 1 - \omega - \log \lambda_j \right\}^2 + \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right] \right).$$

and asymptotically $\hat{\theta}_V \sim N(\theta, M)$ where

$M = (A - C\, B^{-1} C')^{-1}$. In this case $CB^{-1}C'$ reduces to

$$C_\alpha (B_\alpha^{-1} + B_\alpha^{-1}\, B_{\lambda\alpha}'\, M^*\, B_{\lambda\alpha}\, B_\alpha^{-1})C_\alpha'.$$

where $M^* = (B_\lambda - B_{\lambda\alpha} B_\alpha^{-1}\, B_{\lambda\alpha}')^{-1}$.

Evaluating term by term

$$B_\alpha^{-1} = \text{diag}_s \left(\frac{\alpha_j^2}{n_j} \left[\left\{ 1 - \omega - \log \lambda_j \right\}^2 + \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right]^{-1} \right).$$

$$B_{\lambda\alpha} B_\alpha^{-1}\, B_{\lambda\alpha}' = \text{diag}_s \left(\frac{n_j}{\lambda_j^2} \left\{ 1 - \omega - \log \lambda_j \right\}^2 \left[\left\{ 1 - \omega - \log \lambda_j \right\}^2 + \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right]^{-1} \right).$$

$$M^* = \text{diag}_s \left(\frac{\lambda_j^2}{n_j} \left[\left\{ 1 - \omega - \log \lambda_j \right\}^2 + \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right] \left\{ \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right\}^{-1} \right)$$

$$B_\alpha^{-1}\, B_{\lambda\alpha}'\, M^*\, B_{\lambda\alpha}\, B_\alpha^{-1}$$

$$= \text{diag}_s \left(\frac{\alpha_j^2}{n_j} \left\{ 1 - \omega - \log \lambda_j \right\}^2 \left[\left[\left\{ 1 - \omega - \log \lambda_j \right\}^2 + \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right] \left\{ \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right\} \right]^{-1} \right)$$

$$B_\alpha^{-1} + B_\alpha^{-1}\, B_{\lambda\alpha}'\, M^*\, B_{\lambda\alpha} B_\alpha^{-1} = \text{diag}_s \left(\frac{\alpha_j^2}{n_j} \left\{ \frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right\}^{-1} \right)$$

so that

$$N = \left\{ n\mu_2 - \sum_{j=1}^{s} [illegible] \left[\frac{\pi^2}{6} + \beta^2 \mu_{(j)2} \right]^{-1} \right\}^{-1}$$ and asymptotically

$$\tilde{\beta}_V \sim N \left[\beta, \left\{ n \sum_{j=1}^{s} [illegible] \left[1 + \frac{6\beta^2}{\pi^2} \mu_{(j)2} \right]^{-1} \right\}^{-1} \right]$$

It then follows that

$$R_{V,VI}(\beta) = [illegible] \left(\sum_{j=1}^{s} [illegible] \mu_{(j)2} \left[1 + [illegible] \mu_{(j)2} \right]^{-1} \right)$$

Note that $R_{V,VI}(0) = 1$.

Under model IV, terms appearing in $I^{IV}(\beta)$, the information about β contained in the set of s rank vectors, may be evaluated as sums over strata of corresponding model I quantities. At $\beta=0$

$$I^{IV}(0) = \sum_{j=1}^{s} \frac{[illegible]}{n_j - 1} [illegible] \, {}^{*}n_j, \quad \text{where} \quad [illegible]_{n_j} = \sum_{i=1}^{n_j} \frac{n_j - 1}{n_j - i + 1}$$

$$= [illegible] + o(n)$$

and asymptotically var $(\hat{\beta}_{IV}) = \frac{1}{[illegible]}$. Comparison with the asymptotic distributions of $\tilde{\beta}_V$ and $\hat{\beta}_{VI}$ shows that $R_{IV,V}(0) = R_{IV,VI}(0) = 1$. For non-zero β, $\log I^{IV}(\beta)$ may be expanded as a Taylor series, about $\beta=0$, as at 5.3 and similar considerations yield

$$\frac{[illegible]}{[illegible]} = \sum_{j=1}^{s} o(n_j) = o(n),$$

$$\frac{\partial^2 I^{IV}(0)}{\partial \beta^2} = -2\sum_{j=1}^{s} n_j \mu^2{}_{(j)2} + \sum_{j=1}^{s} o(n_j) = -2n\sum_{j=1}^{s} q_j \mu^2{}_{(j)2} + o(n),$$

so that for large n and in the neighbourhood of $\beta=0$

$$I^{IV}(\beta) = n\mu_2 \exp\left(-\frac{\beta^2}{\mu_2}\sum_{j=1}^{s} q_j\ \mu^2_{(j)2}\right).$$

It then follows that

$$R_{IV,VI}(\beta) = \exp\left(-\frac{\beta^2}{\mu_2}\ \sum_{j=1}^{s} q_j \mu^2{}_{(j)2}\right) \text{ and}$$

$$R_{IV,\ V}(\beta) = \mu\left(\sum_{j=1}^{s} \frac{q_j \mu_{(j)2}}{1+6\beta^2 \mu_{(j)2}/\sigma^2}\right)^{-1} \exp\left(-\frac{\beta^2}{\mu_2}\sum_{j=1}^{s} q_j \mu^2{}_{(j)2}\right)$$

in the neighbourhood of $\beta=0$.

<u>Two independent variables</u>

For $k = 1,2$, let $z_{jik} = x_{jik} - \bar{x}_{jk}$ $\quad i=1,\dots,n_j;\ j=1,\dots,s$, where $\bar{x}_{jk} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{jik}$. In addition put

$$\mu_{(j)r_1,r_2} = \frac{1}{n_j}\sum_{i=1}^{n_j} z_{ji1}^{r_1}\ z_{ji2}^{r_2} \qquad j=1,\dots,s \text{ with}$$

$$\mu_{r_1,r_2} = \frac{1}{n}\sum_{j=1}^{s}\sum_{i=1}^{n_j} z_{ji1}^{r_1}\, z_{ji2}^{r_2} = \sum_{j=1}^{s} q_j \mu_{(j)r_1,r_2},\ r_1,r_2=1,2,\dots$$

Note that $\mu_{(j)0,1} = \mu_{(j)1,0} = 0 \qquad j=1,\dots,s.$

Under model VI, the information matrix is given by

$$\underline{I}^{VI}(\beta_1,\beta_2,\underline{\lambda}) = \left[\begin{array}{c:c} \underline{A} & \underset{2\times s}{\underline{0}} \\ \hdashline \underset{s\times 2}{\underline{0}} & \mathrm{diag}_s\left(\frac{n_j}{\lambda_j^2}\right) \end{array}\right]$$

where $\underset{\sim}{A} = n \begin{bmatrix} \mu_{2,0} & \mu_{1,1} \\ \mu_{1,1} & \mu_{0,2} \end{bmatrix}$, and asymptotically

$$\hat{\beta}_{VI} \sim N\left(\beta_1, \frac{\mu_{0,2}}{n(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)}\right).$$

Under model V, the information matrix may be written as

$$\underset{\sim}{I}^{V}(\beta_1,\beta_2,\underset{\sim}{\lambda},\underset{\sim}{\alpha}) = \left[\begin{array}{c:c:c} \underset{\sim}{A} & \underset{2\times s}{\underset{\sim}{0}} & \underset{\sim}{C}_{\alpha} \\ \hdashline \underset{s\times 2}{\underset{\sim}{0}} & \underset{\sim}{B}_{\lambda} & \underset{\sim}{B}_{\lambda\alpha} \\ \hdashline \underset{\sim}{C}_{\alpha}' & \underset{\sim}{B}_{\lambda\alpha} & \underset{\sim}{B}_{\alpha} \end{array}\right]$$

where $\underset{\sim}{A}$ is as in model VI,

$$\underset{\sim}{C}_{\alpha} = \begin{bmatrix} -\frac{n_1}{\alpha_1}(\beta_1\mu(1)_{2,0}+\beta_2\mu(1)_{1,1}),\ldots, & -\frac{n_s}{\alpha_s}(\beta_1\mu(s)_{2,0}+\beta_2\mu(s)_{1,1}) \\ -\frac{n_1}{\alpha_1}(\beta_1\mu(1)_{1,1}+\beta_2\mu(1)_{0,2}),\ldots, & -\frac{n_s}{\alpha_s}(\beta_1\mu(s)_{1,1}+\beta_2\mu(s)_{0,2}) \end{bmatrix},$$

$$\underset{\sim}{B}_{\lambda} = \mathrm{diag}_s\left(-\frac{n_j}{\lambda_j^2}\right),$$

$$\underset{\sim}{B}_{\lambda\alpha} = \mathrm{diag}_s\left(\frac{n_j}{\alpha_j\lambda_j}\left\{1-\psi-\log\lambda_j\right\}\right) \quad \text{and}$$

$$\underset{\sim}{B}_{\alpha} = \mathrm{diag}_s\left(\frac{n_j}{\alpha_j^2}\left\{(1-\psi-\log\lambda_j)^2+\frac{\pi^2}{6}+\beta_1^2\mu(j)_{2,0}+2\beta_1\beta_2\mu(j)_{1,1}+\beta_2^2\mu(j)_{0,2}\right\}\right)$$

A direct extension of the algebra used in the single independent variable case yields the result that asymptotically $\underset{\sim}{\hat{\beta}}_V \sim N(\underset{\sim}{\beta}, \underset{\sim}{M})$ where

$$\underset{\sim}{M} = \frac{1}{n}\begin{bmatrix} \sum_{j=1}^{s} \frac{q_j}{D_j}\left[\frac{\pi^2}{6}\mu_{(j)2,0}+\beta_2^2\mu_{(j)}\right] & \sum_{j=1}^{s} \frac{q_j}{D_j}\left(\frac{\pi^2}{6}\mu_{(j)1,1}-\beta_1\beta_2\mu_{(j)}\right) \\ \sum_{j=1}^{s} \frac{q_j}{D_j}\left[\frac{\pi^2}{6}\mu_{(j)1,1}-\beta_1\beta_2\mu_{(j)}\right) & \sum_{j=1}^{s} \frac{q_j}{D_j}\left(\frac{\pi^2}{6}\mu_{(j)0,2}+\beta_1^2\mu_{(j)}\right) \end{bmatrix}^{-1}$$

with $D_j = \frac{\pi^2}{6} + \beta_1^2\mu_{(j)2,0} + 2\beta_1\beta_2\mu_{(j)1,1} + \beta_2^2\,\mu_{(j)0,2}$ and

$\mu_{(j)} = \mu_{(j)2,0}\,\mu_{(j)0,2} - \mu^2_{(j)1,1}$ $\qquad j=1,\ldots,s,$ and the asymptotic variance of $\hat{\beta}_{V\,1}$ may be obtained in the usual way.

Under model IV considerations similar to those for model I yield the asymptotic result

$$\mathrm{var}\,(\hat{\beta}_{IV\,1}) = V_{IV}(\beta_1,\beta_2)$$
$$= \frac{\mu_{0,2}}{n\mu}\exp\Big[\Big\{\mu^2_{0,2}\sum_{j=1}^{s} q_j(\beta_2\mu_{(j)2,0}+\beta_2\mu_{(j)1,1})^2+\mu^2_{1,1}\sum_{j=1}^{s} q_j(\beta_1\mu_{(j)1,1}+\beta_2\mu_{(j)0,2})^2$$
$$-2\mu_{0,2}\mu_{1,1}\sum_{j=1}^{s} q_j(\beta_1\mu_{(j)1,1}+\beta_2\mu_{(j)0,2})(\beta_2\mu_{(j)2,0}+\beta_2\mu_{(j)1,1})\Big\}\Big/\mu_{0,2}\mu\Big]$$

where $\mu = \mu_{2,0}\mu_{0,2}-\mu^2_{1,1}$, in the neighbourhood of $(\beta_1,\beta_2) = (0,0)$. Details of the calculations leading to this result are given in appendix C. Expressions for the asymptotic relative efficiencies $R_{IV,V}(\beta_1,\beta_2)$, $R_{IV,VI}(\beta_1,\beta_2)$ and $R_{V,VI}(\beta_1,\beta_2)$ are obtained as ratios of appropriate asymptotic estimator variances.

Chapter 6

MODEL CHECKING

§6.1 Introduction

Summary

Choosing the appropriate form of a model and subsequent examination of its fit to the data are two important points that will be considered in this chapter. The rest of this section concerns the choice of a model prior to formal fitting. Methods of checking model assumptions after fitting are discussed in §6.2.

Initial investigations

An initial step in any analysis will usually be to 'screen' a large number of independent variables for those likely to be of some interest regarding prognostic prediction. At this stage only some information on the amount of dependence is required and comparing median survival times (obtained as the 50% percentile of the survivor function estimated as in §1.3) between various subgroups of the data defined by the independent variables may well be adequate.

Having chosen a relatively small subset on which to focus attention more direct methods of assessing the way in which independent variables affect survival are available. Under model I, for each independent variable $k = 1,\ldots,p$,

$$\lambda_i(t) = \lambda_o(t)\, e^{\beta_k x_{ik}} \exp(\tilde{\underline{\beta}}'\, \tilde{\underline{x}}_i)$$

where $\tilde{\underline{x}}_i$ and $\tilde{\underline{\beta}}$ are as before with x_{ik} and β_k omitted. For a binary variable x_k, this relationship is equivalently

$$\lambda_i(t) = \begin{cases} \lambda_o(t)\, e^{\beta_k} \exp(\tilde{\underline{\beta}}'\, \tilde{\underline{x}}_i) & x_{ik} = 1 \\ \lambda_o(t) \exp(\tilde{\underline{\beta}}'\, \tilde{\underline{x}}_i) & x_{ik} = 0 \end{cases}$$

The assumption that the variable x_k acts on the hazard function in this way may be tested by fitting a model of type IV with

$$\lambda_i(t) = \begin{cases} \lambda_{o1}(t) \exp(\tilde{\underline{\beta}}' \tilde{\underline{x}}_i) & x_{ik} = 1 \\ \lambda_{o2}(t) \exp(\tilde{\underline{\beta}}' \tilde{\underline{x}}_i) & x_{ik} = 0, \end{cases}$$

estimating $\lambda_{o1}(t)$ and $\lambda_{o2}(t)$ and assessing the connection between these functions. An appropriate means of assessment is provided by plotting log underlying cumulative hazard functions (log $\tilde{\Delta}_{o1}(t)$ and log $\tilde{\Delta}_{o2}(t)$) against t. Constant differences should result. For discrete variables taking more than 2 values this procedure can be extended in an obvious way to provide useful information concerning the way in which the variable acts. Appropriate groupings allow similar techniques for continuous variables.

§6.2 Assessing goodness of fit

Time dependent covariates

The possibility of using time dependent covariates was mentioned briefly in §3.3 and the justification of their inclusion was given in §4.7. In this general situation, models I, II and III are

$$\lambda_i(t) = \lambda_o(t) \exp(\underline{\beta}' \underline{x}_i(t))$$
$$\lambda_i(t) = \lambda \alpha t^{\alpha-1} \exp(\underline{\beta}' \underline{x}_i(t))$$
$$\lambda_i(t) = \lambda \exp(\underline{\beta}' \underline{x}_i(t))$$

with corresponding likelihoods

$$\prod_{i=1}^{n} \left\{ \frac{\exp(\underline{\beta}' \underline{x}_i^*(t_i^*))}{\sum_{j=i} \exp(\underline{\beta}' \underline{x}_j^*(t_i^*))} \right\}^{\delta_i^*}, \qquad \text{Model I}$$

$$\prod_{i=1}^{n} \left\{ \lambda \alpha t_i^{\alpha-1} \exp(\underline{\beta}' \underline{x}_i(t_i)) \right\}^{\delta_i} \exp\left\{ -\lambda t_i^{\alpha} \exp(\underline{\beta}' \underline{x}_i(t_i)) \right\}, \qquad \text{Model II.}$$

An example of the use of such covariates in assessing the appropriateness of the proportional hazards assumption is given by Cox (1972). In the analysis of the data of Ex.I, Cox uses his discrete form of model I (see §4.1) with covariates $x_{i1} = \begin{cases} 0 & \text{6-MP group} \\ 1 & \text{Placebo group} \end{cases}$ $\quad x_{i2} = x_{i1}(t-10)$

(The constant 10 is included to avoid unduly large numbers in the exponent). Using the large sample likelihood ratio test, the coefficient of x_{i2} was found to be not significantly different from zero. General procedures for choosing particular funtional forms for such covariates are difficult to write down explicitly, although the graphical methods of the previous section may yield some information.

Breslow's 'test for parallelism'

In the two group case with additional independent variable x_2 model I takes the form

$$\lambda_i(t) = \lambda_o(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2}), \qquad (i=1,\ldots,n)$$

where $x_{i1} = \begin{cases} 0 & \text{group 1 members} \\ 1 & \text{group 2 members} \end{cases}$

An assumption implicit in this specification of the model is that x_2 is independent of group membership, that is, the effect of x_2 is the same within the two groups. A more general specification of the form

$$\lambda_i(t) = \lambda_o(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2}) \qquad 6.1$$

will allow this assumption to be assessed, the appropriate hypothesis of interest being $H_o: \beta_3=0$ (The usual large sample test may be employed to test H_o). This is the basis of the 'test for parallelism' considered by Breslow (1974) who writes 6.1 equivalently as

$$\lambda_i(t) = \begin{cases} \lambda_o(t) \exp(\beta_{21} x_{i2}) & \text{group 1 members} \\ \lambda_o(t) \exp(\beta_1 + \beta_{22} x_{i2}) & \text{group 2 members} \end{cases}$$

and investigates the hypothesis $\beta_{21} = \beta_{22}$.

In more general covariate situations, product terms may be included in the model from the outset of the analysis and their coefficients assessed in the usual manner to examine the significance of such interactions.

<u>The use of residuals</u>

Model I may be written in the following equivalent way:
For $i = 1, \ldots, n$, let ϵ_i be a random variable having a unit exponential distribution with

$$\epsilon_i = e^{\underset{\sim}{\beta}'\underset{\sim}{x}_i} \int_o^{T_i} \lambda_o(u)du = h_i^{I}(T_i; \underset{\sim}{\beta}, \lambda_o(\cdot)).$$

This expression of the model allows the use of the methods of Cox and Snell (1968) to define a set of 'crude' residuals

$$R_i = h_i^{I}(T_i; \hat{\underset{\sim}{\beta}}, \hat{\lambda}_o(.)) = e^{\hat{\underset{\sim}{\beta}}'\underset{\sim}{x}_i} \int_o^{T_i} \hat{\lambda}_o(u)du$$

where $\hat{\underset{\sim}{\beta}}$ is the maximum likelihood estimate of $\underset{\sim}{\beta}$ and $\hat{\lambda}_o(.)$ is the estimate of $\lambda_o(.)$ (see §4,3). In the uncensored case the R_i's should exhibit approximately the properties of a random sample of size n from a unit exponential distribution. Information concerning possible dependence of the error quantities on the $\underset{\sim}{x}_i$'s may be gained from plots of 'crude' residuals against corresponding independent variable values for each such variable. Plotting ordered residuals against expected order statistics provides a check of the assumed distributional form of the ϵ_i's.

Similarly, models II and III can be expressed respectively through the transformations

$$\epsilon_i = \lambda e^{\underset{\sim}{\beta}'\underset{\sim}{x}_i} T_i^{\alpha} = h_i^{II}(T_i; \underset{\sim}{\beta}, \lambda, \alpha), \qquad i = 1, \ldots, n,$$

$$\epsilon_i = \lambda e^{\underset{\sim}{\beta}'\underset{\sim}{x}_i} T_i = h_i^{III}(T_i; \underset{\sim}{\beta}, \lambda), \qquad i = 1, \ldots, n,$$

where the ϵ_i's are as above. 'Crude' residuals are obtained on replacing parameters by their maximum likelihood estimates. To extend these methods to the censored case, note that for a censored observation t^*, information on the true 'crude' residual, is of the form

$$R_i > r_i^* = h_i(t^* ; \underset{\sim}{\beta}, \cdot)$$

so that proceeding as before produces a set of exact and censored 'crude' residuals. Under any of the above models the error quantities $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ have survivor function (ε) satisfying

$$\log \;(\varepsilon) = -\varepsilon = -\Lambda(\varepsilon)$$

and a plot of log survivor function, estimated from the 'crude' residuals, provides a check of the distributional assumptions. Altshuler's approach (§1.3) to the estimation of a cumulative hazard function is particularly appropriate here.

Expressing models IV, V and VI respectively through the transformations

$$\varepsilon_{ji} = e^{\underset{\sim}{\beta}'\underset{\sim}{x}_{ji}} \int_0^{T_{ji}} \lambda_{oj}(u)du = h_{ji}^{IV}(T_{ji}; \underset{\sim}{\beta}, \lambda_{oj}(\cdot))$$

$$\varepsilon_{ji} = \lambda_j \, e^{\underset{\sim}{\beta}'\underset{\sim}{x}_{ji}} \, T_{ji}^{\alpha_j} = h_{ji}^{V}(T_{ji}; \underset{\sim}{\beta}, \lambda_j, \alpha_j)$$

$$\varepsilon_{ji} = \lambda_j \, e^{\underset{\sim}{\beta}'\underset{\sim}{x}_{ji}} \, T_{ji} = h_{ji}^{VI}(T_{ji}; \underset{\sim}{\beta}, \lambda_j)$$

where in each case, for $j = 1, \ldots, s$ $i = 1, \ldots, n_j$, ε_{ji} is exponentially distributed with unit mean allows corresponding methods to be used having fitted a model of the within stratatype. In large data sets these techniques should provide an adequate check of model assumptions.

In the uncensored case improvements of the above prodecures are possible. Cox and Snell, in a general context, suggest transformation of the'crude' residuals R_i to form 'modified' residuals R_i' which reflect the properties of the ε_i's more closely. These refinements involve the calculation of means, variances and covariances of the R_i's and resulting choice of a suitable transformation. As an example these authors have considered the uncensored case of model III with a single independent variable. For the p independent variables case, extensions of their calculations (details are given in appendix D) show that, to $o(\frac{1}{n})$, for $i, i_1 = 1, \ldots, n$

$$E(R_i) = 1 + \frac{p}{2n} + \sum_{k=1}^{p} b_k z_{ik} - \frac{1}{2} \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{ik} z_{i\ell}$$

$$= 1 + a_i$$

$$E(R_i^2) = 2 - \frac{2}{n} + \frac{2p}{n} + 4\left(\sum_{k=1}^{p} b_k z_{ik} - \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{ik} z_{i\ell} \right)$$

$$= 2 + c_i$$

$$E(R_i R_{i_1}) = 1 + (a_i + a_{i_1}) - \frac{1}{n} - \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{ik} z_{i_1 \ell}$$

$$= 1 + c_{ii_1} \qquad i \neq i_1$$

where $b_k = E(\hat{\beta}_k - \beta_k) = \frac{1}{2} \sum_{r=1}^{p} \sum_{s=1}^{p} \sum_{t=1}^{p} I^{rk} I^{st} \sum_{u=1}^{p} z_{ur} z_{us} z_{ut}$,

and $I^{k\ell}$ is the (k,ℓ)th element of the inverse of the model III information matrix. The standardised form of model III has been used in these calculations, i.e. with $z_{ij} = x_{ij} - \bar{x}_j$, $j=1,\ldots,p$; $i=1,\ldots,n$. The size of the correction terms a_i, c_i and c_{ii_1} provides a check of the validity of the assumption that the properties of the crude residuals are a close approximation to those of the ε_i's. In the case of model III, Cox and Snell suggest as a suitable transformation $R_i' = \left(\frac{R_i}{1-f_i}\right)^{1+k_i}$ $i=1,\ldots,n$, where the k_i's and f_i's are small. Assuming that this transformation provides random variables R_i' each having a unit exponential distribution they show that, for $i=1,\ldots,n$,

$$E(R_i) = (1-\ell_i)\,\Gamma\left(1 + \frac{1}{1+k_i}\right),$$

$$E(R_i^2) = (1-\ell_i)^2\,\Gamma\left(1 + \frac{2}{1+k_i}\right).$$

Equating these expressions to the approximations given earlier, expanding as a Taylor series about $k_i = \ell_i = 0$ and ignoring higher order terms of k_i and ℓ_i it follows that

$$k_i = \tfrac{1}{2}(4a_i - c_i),$$

$$\ell_i = \tfrac{1}{2}(1-\omega)\,c_i - (3-2\omega)a_i.$$

Observed values of the 'modified' residuals in the uncensored case may now be computed and used as before to check model assumptions. In addition, pairwise correlations between residuals may be examined by noting that, to the order considered,

$$\text{cov}(R_i, R_j) = E(R_i R_j) - E(R_i)E(R_j)$$

$$= -\frac{1}{n} - \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{jk} z_{j\ell} = \text{corr}(R_i, R_j)$$

since the R_i's have approximately unit variance.

Under model VI extension of the calculations for model III (see appendix §D3) yield to $o(\frac{1}{n})$, for $j, j_1 = 1, \ldots, s$; $i = 1, \ldots, n_j$, $i_1 = 1, \ldots, n_{j_1}$,

$$E(R_{ji}) = 1 + \frac{p}{2n_j} + \sum_{k=1}^{p} b_k z_{jik} - \frac{1}{2} \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{jik} z_{ji\ell}$$

$$= 1 + a_{ji},$$

$$E(R_{ji}\ ) = 2 - \frac{2}{n_j} + \frac{2p}{n_j} + 4\left(\sum_{k=1}^{p} b_k z_{jik} - \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{jik} z_{ji\ell}\right)$$

$$= 2 + c_{ji}$$

$$E(R_{ji}R_{j_1i_1}) = 1 - \frac{1}{n_j}\delta_{jj_1} + (a_{ji} + a_{j_1i_1}) - \sum_{k=1}^{p}\sum_{\ell=1}^{p} I^{k\ell} x_{ji\ell}x_{j_1i_1k}$$

$$= 1 + c_{ji,j_1i_1}, \qquad ji \neq j_1i_1$$

where for $k=1,\ldots,p$, $b_k = E(\hat\beta_k - \beta_k)$

$$= \frac{1}{2}\sum_{r=1}^{p}\sum_{t=1}^{p}\sum_{u=1}^{p} I^{rk}I^{tu}\sum_{j=1}^{m}\sum_{i=1}^{n_j} x_{jir}x_{jit}x_{jiu}.$$

$I^{k\ell}$ is the (k,ℓ)th element of the model VI information matrix and δ_{jj_1} is the Kronecker delta. Again the standardised form of the model has been used. Defining 'modified' residuals $R'_{ji} = \left(\frac{R_{ji}}{1-\varepsilon_{ji}}\right)^{1+h_{ji}}$ each having a unit exponential distribution yields, to the order considered,

$$a_{ji} = \frac{1}{2}[\ldots_{ji} - \ldots_{ji}], \ \ldots_{ji} = \frac{1}{2}[\ldots \ldots_{ji} - (3-2w)\ldots_{i}].$$

Finally, covariances between residuals are given by

$$\mathrm{cov}(R_{ji}, R_{j_1i_1}) = \frac{-1}{n_j}\delta_{jj_1} - \sum_{k=1}^{p}\sum_{\ell=1}^{p} I^{k\ell} x_{ji\ell}x_{j_1i_1k} + o\left(\frac{1}{n}\right).$$

Similar methods for models II and V may be used, although the necessary algebraic results require a considerably greater amount of algebra. The results of appendix §A3 are relevant here.

<u>Further results</u>

Greenberg et. al. (1974) using a related model (see §3.3) have suggested comparison of observed and expected deaths in checking model assumptions. The approach could be used here in the uncensored case although extension to the general censoring situation will require either information on potential censoring times (fixed observation time model) or assumptions regarding the censoring distribution for each individual (random censorship model).

In addition the above authors suggest the use of half replicates in assessing goodness of fit. In this approach the data is randomly devided into two groups, one of which is used to estimate parameters, allowing prediction of the survival pattern for the second group. The observed and predicted patterns may then be compared. Censored data again makes this technique intractible.

§6.3 Discussion

A starting point in data analysis is likely to be graphical checks of the type mentioned in §6.1 and fitting of an appropriate form of model I, or model IV if particular independent variable(s) appear to violate the proportional hazards assumption. (Alternatively, the inclusion of time-dependent covariates may be considered for such variables (see §6.2)). The results of chapter 5 suggest that, while investigating the effects of independent variables, little is to be gained from an efficiency standpoint by imposing more stringent assumptions required by the parametric models II and III or V and VI.

Selection of those independent variables meriting inclusion in the model may then be carried out using a stepwise procedure of the type discussed in §4.5. Together with checks of the model (§6.2) these methods select relevant independent variables and appropriately model their effects on survival. Estimation of the underlying hazard function (or functions in model IV) allows explicit expression of the hazard function for an individual as well as definition of a 'crude' residual. It is at this stage that investigation of particular parametric forms for $\lambda_o(.)$ (or $\lambda_{o1}(.),...,\lambda_{os}(.)$) seems appropriate

In the next chapter an example using the above type of analysis will be presented.

Chapter 7

EXAMPLE AND CONCLUDING REMARKS

§7.1 Example

The data

The investigation to be reported here arose from a clinical trial on patients with histologically confirmed cirrhosis assessing the effect of Prednisone treatment. The trial, conducted by the 'Copenhagen Study Group for Liver Diseases', began on 1st January 1962 and was terminated on January 1st 1969 for the purposes of data analysis. A detailed account is given by Juhl et.al(1974). Each patient entering the trial was randomly allocated to either prednisone treatment or placebo tablets. It was thought that several additional factors might be of prognostic importance, that is, age of patient at entry into trial, sex, average daily alcohol consumption for a specified period prior to entry, the activity of the cirrhosis (a well defined biochemical factor) and the absence/presence of ascites.

In order to illustrate the methods discussed in this work, attention will be confined to that subgroup of 177 male patients whose alcohol consumption was above the median value amongst all males, and who had information on all the remaining variables mentioned above. 86 of these patients were members of the control group (placebo tablets) while 91 received Prednisone (treatment group). All independent variables are binary except age, which has been transformed by subtraction of the age sample mean within this group, 59.58 years and division by the group standard deviation, 9.51 years. Of the 110 uncensored survival times, 102 were distinct, there being 4 pairs of 2 tied observations. As regards analysis under models I and IV these ties, where necessary, have been broken at random.

Table 7.1 Data from clinical trial conducted by the 'Copenhagen Study Group for Liver Diseases' on 177 Male alcoholics with cirrhosis of the liver.

Censored (0) or Uncensored (1)	Survival time (days)	Independent Variables Age*	Ascites	Activity	Treatment
1	13	-0.27	-1	1	1
1	15	1.64	1	-1	1
1	19	0.54	1	1	1
1	26	-1.97	1	1	-1
1	32	1.64	1	1	1
1	33	1.24	-1	-1	1
1	36	-1.27	-1	-1	1
1	39	1.64	-1	-1	-1
1	40	1.64	1	1	1
1	45	0.74	-1	-1	1
1	48	-0.77	-1	-1	1
1	56	0.94	1	1	1
1	57	0.74	1	1	-1
1	66	0.04	-1	-1	-1
1	82	-1.67	-1	-1	1
1	90	-0.77	1	1	1
1	90	0.64	1	1	1
0	91	0.84	1	-1	1
1	103	0.84	1	-1	1
0	108	-0.87	-1	-1	1
0	111	0.64	-1	-1	1
1	112	1.84	1	1	-1
1	114	1.04	-1	1	1
1	117	0.34	1	1	1
1	117	2.04	-1	1	1
1	118	0.74	1	-1	1
1	122	-0.57	1	1	1
0	126	-0.17	1	-1	1

1	140	1.34	-1	-1	1
1	155	1.24	1	1	1
1	163	0.74	1	-1	1
0	164	-0.17	-1	-1	1
1	168	0.84	-1	-1	1
1	191	0.44	-1	-1	-1
1	193	0.64	-1	-1	-1
0	193	0.64	1	1	1
0	193	1.54	-1	-1	1
0	197	-1.17	1	-1	1
0	198	0.44	1	-1	1
0	209	-0.17	-1	-1	1
1	223	0.54	1	-1	1
1	256	0.54	1	1	1
1	261	-0.67	1	-1	1
0	263	1.34	-1	-1	1
1	266	-0.87	-1	-1	1
1	269	1.44	1	1	1
1	286	0.54	-1	-1	1
1	304	0.14	1	-1	1
1	322	1.34	1	-1	1
1	344	-0.27	1	1	1
1	346	0.84	1	1	1
1	346	0.34	1	-1	1
0	353	0.94	-1	-1	1
1	361	-0.97	1	-1	1
0	361	-0.07	1	-1	1
0	362	0.14	-1	-1	1
0	363	-0.67	-1	-1	1
1	366	0.74	1	1	-1
0	400	0.74	1	-1	1
1	403	-0.27	1	-1	-1
1	405	1.04	1	-1	1
0	411	-2.17	1	-1	-1

1	415	-0.37	1	-1	1
1	422	-0.37	-1	1	-1
1	427	-0.57	1	1	1
1	454	-0.17	1	-1	1
1	459	0.34	1	-1	-1
1	469	0.94	1	1	-1
1	473	0.64	-1	-1	-1
1	479	0.44	-1	-1	1
1	534	0.44	-1	-1	1
1	647	0.54	1	-1	1
1	689	0.24	1	-1	1
0	710	-0.47	1	-1	1
1	723	-0.67	1	1	1
0	750	-0.57	-1	-1	-1
1	752	0.14	1	-1	1
0	754	-0.77	-1	-1	1
1	777	1.84	-1	-1	1
1	825	1.84	-1	-1	1
1	841	-0.17	-1	-1	1
1	851	-1.37	-1	1	1
1	879	0.34	-1	-1	1
1	941	0.54	-1	-1	1
1	975	-0.37	-1	-1	1
1	1057	0.04	1	1	1
1	1057	0.14	1	-1	1
1	1065	1.04	-1	1	1
0	1069	-1.07	1	-1	-1
1	1078	0.14	-1	-1	1
0	1084	-1.87	-1	-1	1
1	1101	-0.27	-1	-1	1
1	1114	-1.27	-1	-1	-1
1	1141	-0.27	1	-1	1
1	1142	0.14	1	-1	-1
1	1182	-0.17	1	1	1

1	1198	0.34	1	-1	1
0	1228	-1.67	1	-1	1
0	1233	0.84	-1	-1	-1
1	1242	1.54	-1	-1	1
1	1252	0.24	1	-1	1
0	1274	-0.27	-1	1	1
1	1316	-1.37	1	-1	1
1	1359	0.54	-1	-1	-1
0	1370	-3.28	-1	-1	1
1	1377	1.64	1	-1	1
0	1378	-1.67	-1	-1	1
0	1477	0.24	1	-1	1
1	1544	-1.97	-1	-1	1
1	1614	0.84	-1	-1	-1
0	1614	-1.27	1	-1	1
1	1641	0.04	1	-1	-1
1	1695	0.64	-1	1	-1
1	1733	-0.27	1	-1	1
1	1744	0.94	-1	-1	1
1	1791	-0.57	1	1	1
0	1797	0.04	-1	-1	1
1	1810	0.74	-1	-1	1
0	1819	-0.77	1	-1	1
0	1826	-0.97	-1	-1	1
1	1858	-0.37	1	-1	1
0	1891	0.54	1	-1	1
0	1898	0.84	1	-1	1
1	1906	0.04	-1	-1	1
0	1926	-0.37	-1	-1	1
0	1932	-0.47	-1	-1	1
0	1960	-1.47	-1	-1	1
1	1961	0.64	-1	-1	1
1	1975	0.04	-1	-1	1
1	1978	-0.17	1	-1	1

1	1979	-0.37	-1	-1	-1
1	1990	0.54	-1	-1	1
1	2001	0.94	-1	-1	1
1	2005	-0.77	-1	-1	1
1	2032	0.24	1	-1	1
1	2057	0.04	1	-1	1
0	2121	1.04	-1	-1	1
1	2187	0.24	1	1	1
0	2190	-1.27	-1	-1	1
1	2193	0.54	-1	-1	1
1	2198	-1.27	-1	-1	1
0	2198	-0.17	1	-1	1
0	2203	0.44	1	-1	1
0	2205	0.44	-1	-1	1
0	2218	-0.47	1	-1	1
0	2563	-1.87	-1	-1	1
0	2596	0.04	1	-1	1
1	2688	1.04	1	-1	1
1	2706	0.34	-1	-1	1
0	2820	-0.27	1	-1	1
0	2903	-0.97	1	-1	1
1	2912	0.44	1	-1	1
0	2915	-2.07	1	-1	1
0	2919	-1.57	1	-1	1
0	2928	-0.47	-1	-1	1
0	2934	-0.77	1	-1	1
0	2952	-1.67	-1	-1	1
1	2968	-0.77	1	-1	1
1	3016	-0.27	1	-1	1
0	3037	1.14	1	-1	1
0	3078	-0.67	-1	-1	1
1	3094	-1.07	1	-1	1

0	3095	1.04	-1	-1	1
0	3120	-1.27	-1	-1	1
0	3141	-1.17	1	-1	1
1	3211	0.24	1	-1	-1
0	3265	-0.17	-1	-1	1
1	3273	-0.77	-1	-1	1
0	3273	-0.77	-1	1	1
0	3285	0.24	-1	-1	-1
0	3287	-0.07	1	-1	1
0	3299	-1.27	1	-1	1
0	3328	-0.27	-1	-1	1
0	3363	-0.87	-1	1	1
0	3440	-0.87	1	-1	1
0	3742	-0.87	-1	-1	1
0	4011	-0.27	-1	-1	1

Key:

Ascites:	-1	Ascites absent
	+1	Ascites present
Activity:	-1	Activity present
	+1	Activity absent
Treatment:	-1	Control
	+1	Treatment

* Standardised values obtained by subtraction of 59.58 and division by 9.51.

Choice of initial model

Schalm and Summerskill (1975) in discussing this trial and other similar investigations suggest that the effect of Prednisone treatment may depend on a patients status regarding activity and ascites. These considerations lead to a tentative model in which the hazard function for patient i is given by

$$\lambda_i(t) = \lambda_o(t) \exp(\beta x_i + \theta_j y_{ij}) \qquad 7.1$$

where $x_i' = (x_{i1}, x_{i2}, x_{i3})$ is the vector containing age, ascites and activity variables, as given in table 7.1., and

$$y_{i1} = \begin{cases} z_i & x_{i2} = -1,\ x_{i3} = -1, \\ 0 & \text{otherwise} \end{cases}$$

$$y_{i2} = \begin{cases} z_i & x_{i2} = -1,\ x_{i3} = +1, \\ 0 & \text{otherwise} \end{cases}$$

$$y_{i3} = \begin{cases} z_i & x_{i2} = +1,\ x_{i3} = -1, \\ 0 & \text{otherwise} \end{cases}$$

$$y_{i4} = \begin{cases} z_i & x_{i2} = +1,\ x_{i3} = +1, \\ 0 & \text{otherwise} \end{cases}$$

$$\text{with } z_i = \begin{cases} -1 \text{ if i in control group,} \\ +1 \text{ if i in treatment group.} \end{cases}$$

This model allows treatment comparisons to be made within each of the groups defined by ascites x activity.

Preliminary model checking methods of the type discussed in §6.1. assess the assumption that independent variables affect the hazard in this way. Figures 7.1., 7.2., 7.3. and 7.4. provide plots of log underlying cumulative hazard functions (obtained using the Kalbfleisch and Prentice method of §4.3.) having fitted model IV with each of age, ascites, activity and treatment defining strata in turn.

Fig. 7.1. Log underlying cumulative hazard functions to check inclusion of age. Model IV fitted with variables x_2, x_3 and y_j $j = 1,\ldots,4$.

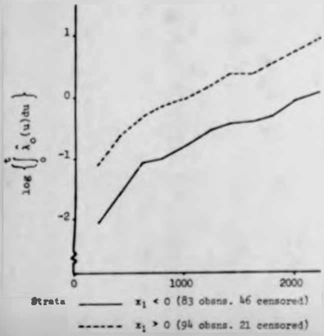


Fig. 7.2. Log underlying cumulative hazard functions to check inclusion of ascites. Model IV fitted with variables x_1, x_3 and y_j $j = 1,\ldots,4$.

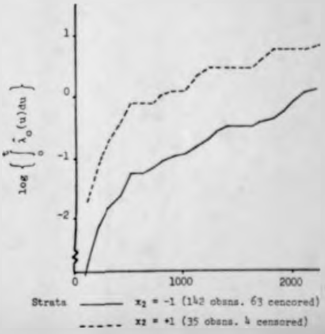
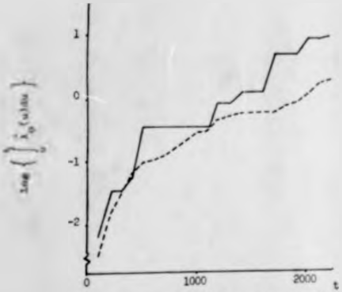


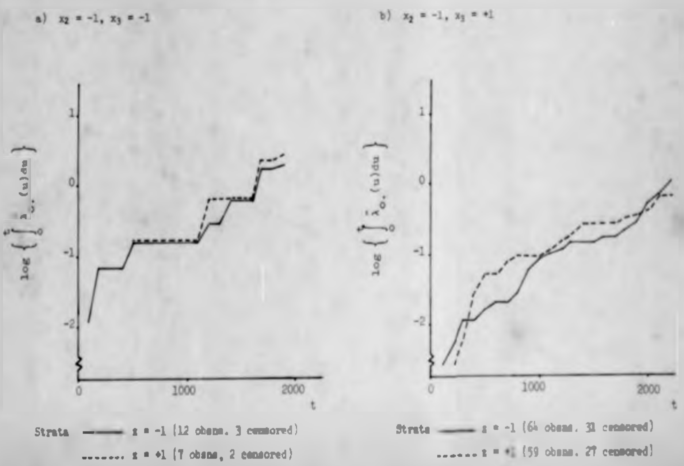
Fig. 7.3. Log underlying cumulative hazard functions to check inclusion of activity. Model IV fitted with variables x_1, x_2 and y_j, $j = 1, \ldots, 4$



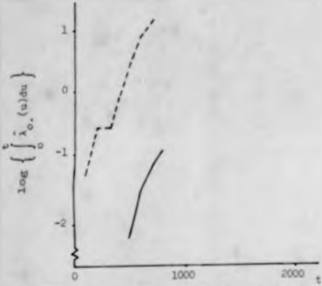
Strata ——— $x_3 = -1$ (26 obsns. 5 censored)

- - - - - $x_3 = +1$ (151 obsns. 62 censored)

Fig. 7.4. Log underlying hazard functions to check inclusion of treatment effect with each group defined by ascites = activity. Model IV fitted with variable age.



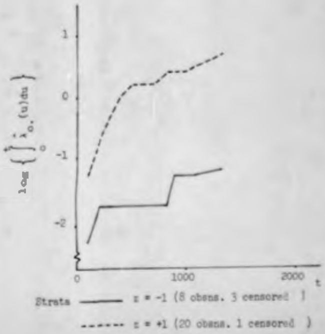
e) $x_2 = +1$, $x_3 = -1$



Strata ——— $z = -1$ (2 obsns. 0 censored)

- - - - - - $z = +1$ (5 obsns. 0 censored)

d) $x_2 = +1,\ x_3 = +1$



log $\left\{ \int_0^t \hat{\lambda}_{o.}(u)du \right\}$
1
0
-1
-2
0
1000
2000
t
Strata —— z = -1 (8 obsns. 3 censored)
------ z = +1 (20 obsns. 1 censored)

In figure 7.1., a dichotomy using the mean age allows this variable to be included in these checks. For the treatment variable, comparison has been performed by defining 8 strata according to y_{ij} ; $j = 1, \ldots, 4$ and fitting model IV with independent variable age. These plots provide no evidence to suggest that the multiplicative assumptions of the model at 7.1. are violated.

Parameter estimates, with standard errors, obtained by direct evaluation of the matrix of second partial derivatives, having fitted 7.1. are given in table 7.2.

Table 7.2. Parameter estimates and standard errors having fitted model I at 7.1.

Independent variable	Estimated value of coefficient	Standard error of estimator
x_1	0.6126	0.1214
x_2	0.4649	0.1265
x_3	-0.2009	0.1319
y_1	-0.0239	0.2861
y_2	0.0292	0.1239
y_3	0.4313	0.3826
y_4	0.5906	0.2529

Selection of significant effects

The methods of §4.3. have been employed to select those independent variables meriting inclusion in the model. Table 7.3. presents results of forward and backward stepwise selection procedures respectively.

Table 7.3. Selecting independent variables having significant effect on survival under model I.

a) Forward stepwise procedure.

Independent variables	Maximum value of log likelihood	Value of test statistic
None	-494.469	
x_1	-477.511	33.916*
x_2	-482.801	23.336
x_3	-491.655	5.628
y_1	-494.292	0.354
y_2	-494.466	0.006
y_3	-492.037	4.864
y_4	-486.467	16.004
x_1,x_2	-466.310	22.402*
x_1,x_3	-475.937	3.148
x_1,y_1	-477.496	0.030
x_1,y_2	-477.450	0.122
x_1,y_3	-476.132	2.758
x_1,y_4	-470.429	14.164
[illegible]	-465.397	1.826
x_1,x_2,y_1	-466.218	0.184
x_1,x_2,y_2	-466.272	0.076
x_1,x_2,y_3	-465.635	1.350
x_1,x_2,y_4	-463.703	5.814*
x_1,x_2,y_4,x_3	-462.368	2.670
x_1,x_2,y_4,y_1	-463.609	0.188
x_1,x_2,y_4,y_2	-463.666	0.074
x_1,x_2,y_4,y_3	-462.876	1.654

* stages in forward selection.

b) Backward selection procedure.

Independent variables	Maximum value of log likelihood	Value of test statistic
$x_1,x_2,x_3,y_1,y_2,y_3,y_4$	-461.682	
x_1,x_2,x_3,y_1,y_2,y_3	-464.810	6.256
x_1,x_2,x_3,y_1,y_2,y_4	-462.337	1.310
x_1,x_2,x_3,y_1,y_3,y_4	-461.709	0.054
x_1,x_2,x_3,y_2,y_3,y_4	-461.685	0.006*
x_1,x_2,y_1,y_2,y_3,y_4	-462.744	2.124
x_1,x_3,y_1,y_2,y_3,y_4	-467.010	10.656
x_2,x_3,y_1,y_2,y_3,y_4	-474.995	26.626
x_1,x_2,x_3,y_2,y_3	-464.821	6.272
x_1,x_2,x_3,y_2,y_4	-462.338	1.306
x_1,x_2,x_3,y_3,y_4	-461.713	0.056*
x_1,x_2,y_2,y_3,y_4	-462.841	2.312
x_1,x_3,y_2,y_3,y_4	-467.024	10.678
x_2,x_3,y_2,y_3,y_4	-475.049	26.728
x_1,x_2,x_3,y_3	-464.851	6.276
x_1,x_2,x_3,y_4	-462.368	1.310*
x_1,x_2,y_3,y_4	-462.876	2.326
x_1,x_3,y_3,y_4	-467.061	10.696
x_2,x_3,y_3,y_4	-475.050	26.674
x_1,x_2,x_3	-465.397	6.058
x_1,x_2,y_4	-463.703	2.670*
x_1,x_3,y_4	-468.174	11.612
x_2,x_3,y_4	-476.315	27.894
x_1,y_4	-466.310	5.214
x_1,x_2	-470.429	13.452
x_2,y_4	-479.313	31.220

* stages in backward selection

Both procedures result in a final model containing the independent variables x_1, x_2 and y_4. Parameter estimates and standard errors are given in table 7.4. for this final model, together with the full estimated covariance of the estimators obtained, as before, by direct evaluation of the matrix of second partial derivatives.

<u>Table 7.4</u>. Final version of model I with independent variables x_1, x_2 and y_4.

Independent variable	Estimated value of coefficient		Standard error of estimator
x_1	(β_1)	0.6427	0.1181
x_2	(β_2)	0.4992	0.1209
y_4	(θ_4)	0.5051	0.2326

Estimated covariance matrix

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}_4$
$\hat{\beta}_1$	0.0139		
$\hat{\beta}_2$	0.0003	0.0146	
$\hat{\theta}_4$	-0.0001	-0.0114	0.0541

Note that employing a selection procedure based on the approximate normality of parameter estimators and the results of table 7.2. would have led to the same final model.

<u>Functional form for $\lambda_0(t)$</u>

Using intervals $I_r = [100(r-1), 100r)$, $r = 1, \ldots ,30$ and $I_{31} = [3000, \infty)$, table 7.5 presents step estimates in the step function approximation (Kalbfleisch and Prentice, §4.3) of $\lambda_0(t)$ under model I having fitted the independent variables x_1, x_2 and y_4.

Table 7.5. Estimation of underlying hazard function $\lambda_o(t)$ in model I having fitted independent variables x_1, x_2 and y_4.

r	$\hat{\lambda}_o(t)$ $(x10^3)$, $t \epsilon I_r$	r	$\hat{\lambda}_o(t)$ $(x10^3)$, $t \epsilon I_r$
1	0.92	17	0.68
2	0.92	18	0.75
3	0.51	19	0.56
4	0.70	20	1.94
5	1.18	21	1.56
6	0.13	22	1.28
7	0.26	23	0
8	0.40	24	0
9	0.58	25	0
10	0.30	26	0
11	0.63	27	0.64
12	1.04	28	0.73
13	0.38	29	0
14	0.61	30	1.81
15	0	31	1.54
16	0.22		

Estimation of the log underlying cumulative hazard function, carried out using

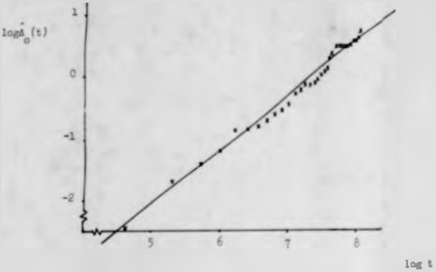
$$\hat{\Delta}_o(t) = \int_0^t \hat{\lambda}_o(u)du ,$$

is plotted against log t at t = 100r , r = 1,... ,31 in figure 7.5. and a straight line fitted by eye. The plot suggests that a relationship of the form

$$\log \Delta_o(t) = c\log t + d \qquad 7.2.$$

exists between $\Delta_o(t)$ and t.

Fig. 7.5 Plot of log underlying cumulative hazard function against log t having fitted model I with independent variables x_1, x_2 and y_4.



In addition the straight line fit yields approximate values 0.9 and -6.5 for c and d respectively. The expression 7.2. may be written equivalently as

$$\lambda_0(t) = \lambda \alpha t^{\alpha-1}$$

where $\alpha=c$ and $\lambda=e^d$, so that fitting a model of type II provides an appropriate 'smoothing' of the step function estimate of $\lambda_0(t)$. Table 7.6. gives details of this fit.

Table 7.6. Parameter estimates and standard errors having fitted model II with independent variables x_1, x_2 and y_4.

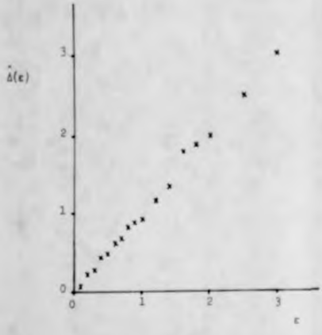
Independent variable		Estimated parameter value	Standard error of estimator
	$(\bar{\alpha})$	0.9283	0.0718
	$(\bar{\lambda})$	0.0011	0.0006
x_1		0.6436	0.1169
x_2		0.5172	0.1203
y_4		0.4895	0.2315

Bearing in mind standard errors, the estimated coefficients of x_1, x_2 and y_4 in model II are in close agreement with those obtained under model I. In addition $\bar{\alpha}$ and $\bar{\lambda}$ take values close to those obtained for c and e^d from the plot of log $\hat{\Lambda}_0(t)$ against log t.

Model check using residuals

The overall adequacy of the model has been checked using the (crude) residuals obtained through the model II fit as discussed in §6.2. Figure 7.6. presents a plot of the cumulative hazard function estimates $\hat{\Lambda}(\varepsilon)$ at points ε = 0(0.05)1, 1(0.2)2 , 2(0.5)3. The relationship is as expected, suggesting that the model II fit adequately describes the data. Note that a similar procedure, based on model I estimates and the estimated underlying hazard function to define residuals, could have been used.

Fig. 7.6. Plot of log cumulative hazard function for error quantity, using Altshuler's method. Crude residuals obtained from model II.



Discussion

Model fitting was carried out on an ICL 1904S computer. The Numerical Algorthms Group library routines E04EAF (models I and IV) and E04DCF (model II) (Short write-ups contained in Mark 4 version of Nag Mini Manual for ICL 1900* library) were used in the log likelihood maximizations. Routines to calculate the value of the log likelihood function and its first derivatives (and second derivatives in the case of models I and IV) at any point were supplied by the author.

The medical conclusions to be drawn from this analysis are clear. Only in the ascites present, activity absent group is there a significant treatment effect and treatment with Prednisone in this case is unfavourable as regards length of subsequent survival. For the data as a whole, younger patients tend to do better than older patients, as would be expected, and the presence of ascites has a detrimental effect on survival length.

The techniques used in checking model assumptions, that is log cumulative hazard plots prior to model fitting and residual plots after fitting, are clearly important points in any analysis. However, it is not clear what departures might be expected if some of the model assumptions are violated and more work in this area is required.

§7.2 A related area of study

Change in treatment status

A problem of the survival data type which has received recent attention in the literature concerns the situation in which a patient may switch treatment group during the course of the study. Turnbull, Brown and Hu (1974), using an illustration involving Heart-transplant data, have presented some theoretical techniques dealing with the two treatment group situation in which all patients enter the study in one of the groups and change in treatment status occurs, if at all, in a single direction. Considerations of this type arise elsewhere in the medical field. Spiers et.al.(1975) have reported a trial on patients with Chronic Granulocytic Leukaemia where an operation to remove the spleen (movement to group 2) takes place at some time after entry into the trial as a group 1 member.

Methods of analysis

One of the methods employed by the above authors is model III with a time-dependent covariate

$$x_i(t) = \begin{cases} 0 & \text{for patients remaining in group 1} \\ \delta(t-y_i) & \text{for patients in group 1 moving to group 2 at time } y_i \text{ after entry into study} \end{cases}$$

$$\text{where } \delta(u) = \begin{cases} 0 & u < 0 \\ 1 & u \geq 0 \end{cases}$$

Because of the 'no memory' property of the exponential distribution this is equivalent to the model III, two group case, in which t_i, $i \notin A$, $t_i - y_i$, $i \in A$ are treated as group 1 and group 2 observations (censored or exact) respectively and y_i, $i \in A$ are treated as group 1 censored observations, where A is the set of patients who change treatment.

Using a time-dependent covariate of this type in model I will result in the likelihood as in §6.2. Under model II, the likelihood will consist of a term for each $i \notin A$ of the form

$$\left[\lambda\alpha e^{\beta' z_i} t_i^{\alpha-1}\right]^{\delta_i} \exp\left(-\lambda e^{\beta' z_i} t_i^{\alpha}\right) ,$$

while for $i \in A$, the required quantity is

$$\left[\lambda\alpha e^{\theta+\beta' z_i} t_i^{\alpha-1}\right]^{\delta_i} \exp\left(-\lambda e^{\beta' z_i}\left\{y_i^{\alpha} + e^{\theta}\left(t_i^{\alpha} - y_i^{\alpha}\right)\right\}\right) .$$

Inferential procedures are carried out in the usual way.

APPENDIX A Expected values associated with models II, III, V and VI

A.1 The gamma function and its derivatives

The gamma function $\Gamma(a)$, defined for $a > 0$, by

$$\Gamma(a) = \int_0^\infty y^{a-1} e^{-y}\, dy \qquad \text{A1.}$$

may be expressed as an infinite product

$$\frac{1}{\Gamma(a)} = a e^{\omega a} \prod_{n=1}^{\infty} \left\{ \left(1 + \frac{a}{n}\right) e^{-a/n} \right\} \qquad \text{A2.}$$

where $\omega = \lim_{n \to \infty} \left[\sum_{k=1}^{n} \frac{1}{k} - \log n \right] = 0.577215\ldots$ is Euler's constant.

If $\Gamma^{(r)}(a)$ denotes the r^{th} derivative of $\Gamma(a)$ w.r.t. a it follows from A1. that

$$\Gamma^{(r)}(1) = \int_0^\infty (\log y)^r e^{-y}\, dy \qquad r = 0,1,2,\ldots$$

Using A2. and the well known property $\Gamma(a+1) = a\Gamma(a)$

$$-\log \Gamma(a+1) = \log \frac{1}{a\Gamma(a)} = \omega a + \sum_{n=1}^{\infty} \left\{ \log(a+n) - \log n - \frac{a}{n} \right\}$$

and successive differentiation yields

$$\frac{\Gamma'(a+1)}{\Gamma(a+1)} = -\omega + a \sum_{n=1}^{\infty} \frac{1}{(a+n)n} \qquad \text{A3.}$$

$$\frac{\Gamma''(a+1)\Gamma(a+1) - [\Gamma'(a+1)]^2}{[\Gamma(a+1)]^2} = \sum_{n=1}^{\infty} \frac{1}{n(a+n)} - a \sum_{n=1}^{\infty} \frac{1}{n(a+n)^2} \qquad \text{A4.}$$

$$\frac{\Gamma'''(a+1)[\Gamma(a+1)]^2 - 3\Gamma''(a+1)\,\Gamma'(a+1)\,\Gamma(a+1) + 2[\Gamma'(a+1)]^3}{[\Gamma(a+1)]^2}$$

$$= 2a \sum_{n=1}^{\infty} \frac{1}{n(a+n)^3} - 2 \sum_{n=1}^{\infty} \frac{1}{n(a+n)^2} \qquad \text{A5.}$$

Putting $a = 0$ in each of A3, A4 and A5 and using the result $\Gamma(1) = 1$ it follows that

$$\Gamma'(1) = -\omega,\ \Gamma''(1) = \omega^2 + S(2),\ \Gamma'''(1) = -\omega^3 - 3\omega S(2) - 2S(3) \qquad \text{A6.}$$

where $S(r) = \sum_{i=1}^{\infty} \frac{1}{i^r}$ is the Zeta-function

§A.2 Integrals of the form $\int_0^{\infty} y^a(\log y)^b\, e^{-y}\, dy$

It will be necessary for certain non-negative integers a and b to evaluate the integral

$$L(a,b) = \int_0^{\infty} y^a(\log y)^b\, e^{-y}\, dy \qquad \text{A7.}$$

Note that

$$L(a,0) = \Gamma(a+1) = a!\ ,\quad L(0,b) = \Gamma^{(b)}(1)\ ,\quad \begin{matrix} a=0,1,2,\ldots \\ b=0,1,2,\ldots \end{matrix} \qquad \text{A8.}$$

For $a,b = 1,2,\ldots$ integration of A7 yields

$$L(a,b) = a\,L(a-1,b) + b\,L(a-1,b-1) \qquad \text{A9.}$$

The recurrence relation A9 together with the initial values A8 may now be used to obtain values of $L(a,b)$. Numerical values are given in table A1.

Table A1. Values of $L(a,b)$ for $a,b = 0,1,2,3$.

b \ a	0	1	2	3
0	1	1	2	6
1	-0.577	0.523	2.523	8.523
2	1.978	3.379	7.804	23.177
3	-5.445	-14.357	-36.313	-85.527

A.3 Model II and Model V quantities

Under model II, $T_1, T_2, \ldots, T_n$ are independent random variables with T_i having p.d.f.

$$p_i(t/\underline{\beta},\lambda,\alpha) = \lambda\alpha t^{\alpha-1} e^{\underline{\beta}'\underline{x}_i} \exp(-\lambda t^{\alpha} e^{\underline{\beta}'\underline{x}_i}) \qquad t>0$$

Thus, for $a,b = 0,1,2,\ldots$

$$E\left\{(\log T_i)^b\, T_i^{a\alpha}\right\} = \int_0^{\infty} \lambda\alpha(\log u)^b u^{(a+1)\alpha-1} e^{\underline{\beta}'\underline{x}_i} \exp(-\lambda u^{\alpha} e^{\underline{\beta}'\underline{x}_i})\,du \qquad \text{A10.}$$

and using the change of variable $v = \lambda u^{\alpha} e^{\underline{\beta}'\underline{x}_i}$, this expression reduces to

$$\frac{e^{-a\underline{\beta}'\underline{x}_i}}{\alpha^b\lambda^a}\int_0^{\infty}(\log v - \log\lambda - \underline{\beta}'\underline{x}_i)^b v^a e^{-v}\,dv$$

$$= \frac{e^{-a\underline{\beta}'\underline{x}_i}}{\alpha^b\lambda^a}\sum_{k=0}^{b}\frac{(-1)^k b!}{k!(b-k)!}(\log\lambda + \underline{\beta}'\underline{x}_i)^k L(a,b-k) \qquad \text{A11.}$$

Similarly, under model V

$$E\left\{(\log T_{ji})^b \, T_{ji}^{\,a\alpha_j}\right\} = \frac{e^{-a\beta' z_{ji}}}{\alpha_j^{\,b}\lambda_j^{\,a}} \sum_{k=0}^{b} \frac{(-1)^k b!}{k!(b-k)!} (\log\lambda_j + \beta' z_{ji})^k I(a, b-k)$$

§A.4 Model III and Model VI quantities

When α = 1, Model II reduces to III. In this case quantities of the form

$$E(T_i^{\,a}) = \int_0^\infty \lambda u^a e^{\beta' z_i} \exp(-\lambda u\, e^{\beta' z_i})\,du \qquad \text{A12.}$$

for a = 1,2,... are required.

Putting α = 1 and b = 0 in A11. it follows that

$$E(T_i^{\,a}) = \frac{a!\,e^{-a\beta' z_i}}{\lambda^a} \qquad \text{A13.}$$

In addition, under model VI

$$E(T_{ji}^{\,a}) = \frac{a!\,e^{-a\beta' z_{ji}}}{\lambda_j^{\,a}} \qquad \text{A14.}$$

Appendix B. The asymptotic variance of $\hat{\theta}_{I1}$

§B.1 Taylor series expansion

For algebraic simplicity, let

$A = I^{-1}_{11}(\beta_1,\beta_2)$, $B = I^{-1}_{22}(\beta_1,\beta_2)$, $C = I^{-1}_{12}(\beta_1,\beta_2)$ and put

$X = AB - C^2$.

Then $V_I(\beta_1,\beta_2) = \frac{B}{X}$ and the Taylor series expansion of $\log V_I(\beta_1,\beta_2)$ about $(\beta_1,\beta_2) = (0,0)$ is given as

$$\log V_I(\beta_1,\beta_2) = \log \frac{B}{X}$$

$$= \left[\log \frac{B}{X}\right]_{(0,0)} + \beta_1\left[\frac{B^{(1)}}{B} - \frac{X^{(1)}}{X}\right]_{(0,0)} + \beta_2\left[\frac{B^{(2)}}{B} - \frac{X^{(2)}}{X}\right]_{(0,0)}$$

$$+ \frac{\beta_1^2}{2}\left[\frac{B^{(11)}}{B} - \frac{X^{(11)}}{X} + \frac{X^{(1)2}}{X^2} - \frac{B^{(1)2}}{B^2}\right]_{(0,0)} + \beta_1\beta_2\left[\frac{B^{(12)}}{B} - \frac{X^{(12)}}{X} - \frac{B^{(1)}B^{(2)}}{B^2} + \frac{X^{(1)}X^{(2)}}{X^2}\right]_{(0,0)}$$

$$+ \frac{\beta_2^2}{2}\left[\frac{B^{(22)}}{B} - \frac{X^{(22)}}{X} + \frac{X^{(2)2}}{X^2} - \frac{B^{(2)2}}{B^2}\right]_{(0,0)} + \cdots$$

where $B^{(i)} = \frac{\partial B}{\partial \beta_i}$, $i=1,2$, $\quad B^{(ij)} = \frac{\partial^2 B}{\partial \beta_i \partial \beta_j}$ $i,j=1,2,$

with $B^{(12)} = B^{(21)}$, and similarly for X functions.

Extending this notation in an obvious way to include A and C functions

$X^{(i)} = A^{(i)}B + AB^{(i)} - 2CC^{(i)}$

$X^{(ii)} = A^{(ii)}B + 2A^{(i)}B^{(i)} + AB^{(ii)} - 2C^{(i)2} - 2CC^{(ii)}$, $i=1,2$ and

$X^{(12)} = A^{(12)}B + A^{(1)}B^{(2)} + A^{(2)}B^{(1)} + AB^{(12)} - 2C^{(1)}C^{(2)} - 2CC^{(12)}$.

§B.2 A, B and C functions evaluated at $(s_1, s_2) = (0,0)$

$$(A)_{(0,0)} = \bar{I}_1(0,0) = E_p(g_{2,0}(0,0)).$$

$$(A^{(1)})_{(0,0)} = \left\{\frac{\partial \bar{I}_1}{\partial s_1}(s_1, s_2)\right\}_{(0,0)}$$

$$= E_p\left\{g_{3,0}(0,0) - g_{2,0}(0,0)\, g_{1,0}(0,0)\right\}.$$

$$(A^{(2)})_{(0,0)} = \left\{\frac{\partial \bar{I}_1}{\partial s_2}(s_1, s_2)\right\}_{(0,0)}$$

$$= E_p\left\{g_{2,1}(0,0) - g_{2,0}(0,0) g_{0,1}(0,0)\right\}.$$

$$(A^{(11)})_{(0,0)} = \left\{\frac{\partial^2 \bar{I}_1}{\partial s_1^2}(s_1, s_2)\right\}_{(0,0)}$$

$$= E_p[g_{4,0}(0,0) - 2g_{3,0}(0,0) g_{1,0}(0,0) - (g_{2,0}(0,0))^2 + g_{2,0}(0,0)(g_{1,0}(0,0))^2]$$

$$(A^{(12)})_{(0,0)} = \left\{\frac{\partial^2 \bar{I}_1}{\partial s_1 \partial s_2}(s_1, s_2)\right\}_{(0,0)}$$

$$= E_p(g_{3,1}(0,0) - g_{2,1}(0,0) g_{1,0}(0,0) - g_{2,0}(0,0) g_{1,1}(0,0)$$

$$- g_{3,0}(0,0)\, g_{0,1}(0,0) + g_{2,0}(0,0) g_{1,0}(0,0)\, g_{0,1}(0,0)).$$

$$(A^{(22)})_{(0,0)} = \left\{\frac{\partial^2 \bar{I}_1}{\partial s_2^2}(s_1, s_2)\right\}_{(0,0)}$$

$$= E_p[g_{2,2}(0,0) - 2g_{2,1}(0,0) g_{0,1}(0,0) - g_{2,0}(0,0) g_{0,2}(0,0)$$

$$+ g_{2,0}(0,0)\, (g_{0,1}(0,0))^2].$$

The B functions may be deduced directly from the above on considerations of symmetry.

$$(C)_{(0,0)} = I^{C}_{12}(0,0) = E_p\left\{g_{1,1}(0,0)\right\}$$

$$(C^{(1)})_{(0,0)} = \left\{\frac{\partial I^{C}_{12}}{\partial s_1}(s_1,s_2)\right\}_{(0,0)}$$

$$= E_p(g_{2,1}(0,0) - g_{1,1}(0,0)g_{1,0}(0,0))$$

$$(C^{(2)})_{(0,0)} = \left\{\frac{\partial I^{C}_{12}}{\partial s_2}(s_1,s_2)\right\}_{(0,0)}$$

$$= E_p(g_{1,2}(0,0) - g_{1,1}(0,0)g_{0,1}(0,0))$$

$$(C^{(11)})_{(0,0)} = \left\{\frac{\partial^2 I^{C}_{12}}{\partial s_1^2}(s_1,s_2)\right\}_{(0,0)}$$

$$= E_p\Big[g_{3,1}(0,0) - 2g_{2,1}(0,0)g_{1,0}(0,0) - g_{1,1}(0,0)\, g_{2,0}(0,0)$$

$$+ g_{1,1}(0,0)\left\{g_{1,0}(0,0)\right\}^2\Big]$$

$$(C^{(12)})_{(0,0)} = \left\{\frac{\partial^2 I^{C}_{12}}{\partial s_1 \partial s_2}(s_1,s_2)\right\}_{(0,0)}$$

$$= E_p\Big[g_{2,2}(0,0) - g_{2,1}(0,0)g_{0,1}(0,0) - \{g_{1,1}(0,0)\}^2 - g_{1,2}(0,0)g_{1,0}(0,0)$$

$$+ g_{1,1}(0,0)\, g_{0,1}(0,0)\, g_{1,0}(0,0)\Big]$$

$$(C^{(2,2)})_{(0,0)} = \left\{\frac{\partial^2 I^{C}_{12}}{\partial s_2^2}(s_1,s_2)\right\}_{(0,0)}$$

$$= E_p\Big[g_{1,3}(0,0) - 2g_{1,2}(0,0)g_{0,1}(0,0) - g_{1,1}(0,0)g_{0,2}(0,0)$$

$$+ g_{1,1}(0,0)\{g_{0,1}(0,0)\}^2\Big]$$

§B.3 g functions expressed as series expansions

$$g_{1,0}(0,0) = \sum_{i=1}^{n} \frac{1}{n-i+1} \sum_{j=i}^{n} z^*_{j1}$$

$$g_{2,0}(0,0) = \sum_{i=1}^{n} \frac{1}{n-i+1} \sum_{j=i}^{n} z^{*2}_{j1} - \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j1} z^*_{k1}$$

$$g_{1,1}(0,0) = \sum_{i=1}^{n} \frac{1}{n-i+1} \sum_{j=i}^{n} z^*_{j1} z^*_{j2} - \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j1} z^*_{k2}$$

$$g_{3,0}(0,0) = \sum_{i=1}^{n} \frac{1}{n-i+1} \sum_{j=i}^{n} z^{*3}_{j1} - 3 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j1} z^{*2}_{k1}$$

$$+ 2 \sum_{i=1}^{n} \frac{1}{(n-i+1)^3} \sum_{j=i}^{n} \sum_{k=i}^{n} \sum_{l=i}^{n} z^*_{j1} z^*_{k1} z^*_{l1}$$

$$g_{2,1}(0,0) = \sum_{i=1}^{n} \frac{1}{n-i+1} \sum_{j=i}^{n} z^{*2}_{j1} z^*_{j2} - \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j2} z^{*2}_{k1}$$

$$-2 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j1} z^*_{k1} z^*_{k2} + 2 \sum_{i=1}^{n} \frac{1}{(n-i+1)^3} \sum_{j=i}^{n} \sum_{k=i}^{n} \sum_{l=i}^{n} z^*_{j2} z^*_{k1} z^*_{l1}$$

$$g_{4,0}(0,0) = \sum_{i=1}^{n} \frac{1}{n-i+1} \sum_{j=i}^{n} z^{*4}_{j1} - 3 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^{*2}_{j1} z^{*2}_{k1}$$

$$-4 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j1} z^{*3}_{k1} + 12 \sum_{i=1}^{n} \frac{1}{(n-i+1)^3} \sum_{j=i}^{n} \sum_{k=i}^{n} \sum_{l=i}^{n} z^*_{j1} z^*_{k1} z^{*2}_{l1}$$

$$-6 \sum_{i=1}^{n} \frac{1}{(n-i+1)^4} \sum_{j=i}^{n} \sum_{k=i}^{n} \sum_{l=i}^{n} \sum_{m=i}^{n} z^*_{j1} z^*_{k1} z^*_{l1} z^*_{m1}$$

$$g_{3,1}(0,0) = \sum_{i=1}^{n} \frac{1}{n-i+1} \sum_{j=i}^{n} z^{*3}_{j1} z^*_{j2} - \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j2} z^{*3}_{k1}$$

$$-3 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j1} z^*_{j2} z^{*2}_{k1} - 3 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \sum_{j=i}^{n} \sum_{k=i}^{n} z^*_{j1} z^{*2}_{k1} z^*_{k2}$$

$$+ 6 \sum_{i=1}^{n} \frac{1}{(n-i+1)^3} \sum_{j=i}^{n} \sum_{k=i}^{n} \sum_{l=i}^{n} z^*_{j1} z^{*2}_{k1} z^*_{l2} + 6 \sum_{i=1}^{n} \frac{1}{(n-i+1)^3} \sum_{j=i}^{n} \sum_{k=i}^{n} \sum_{l=i}^{n} z^*_{j1} z^*_{k1} z^*_{l1} z^*_{l2}$$

$$-6\sum_{i=1}^{n}\frac{1}{(n-i+1)^4}\sum_{j=i}^{n}\sum_{k=i}^{n}\sum_{\ell=i}^{n}\sum_{m=i}^{n} z_{j1}^{*}\, z_{k1}^{*}\, z_{\ell 1}^{*}\, z_{m2}^{*} \,.$$

$$g_{2,2}(0,0) = \sum_{i=1}^{n}\frac{1}{n-i+1}\sum_{j=i}^{n} z_{j1}^{*2}\, z_{j2}^{*2} - 2\sum_{i=1}^{n}\frac{1}{(n-i+1)^2}\sum_{j=i}^{n}\sum_{k=i}^{n} z_{j2}^{*}\, z_{k1}^{*2}\, z_{k2}^{*}$$

$$-\sum_{i=1}^{n}\frac{1}{(n-i+1)^2}\sum_{j=i}^{n}\sum_{k=i}^{n} z_{j1}^{*2}\, z_{k2}^{*2} - 2\sum_{i=1}^{n}\frac{1}{(n-i+1)^3}\sum_{j=1}^{n}\sum_{k=i}^{n} z_{j1}^{*}\, z_{j2}^{*}\, z_{k1}^{*}\, z_{k2}^{*}$$

$$-2\sum_{i=1}^{n}\frac{1}{(n-i+1)^2}\sum_{j=i}^{n}\sum_{k=i}^{n} z_{j1}^{*}\, z_{k2}^{*2} z_{k1}^{*} + 2\sum_{i=1}^{n}\frac{1}{(n-i+1)^3}\sum_{j=i}^{n}\sum_{k=i}^{n}\sum_{\ell=i}^{n} z_{j2}^{*}\, z_{k2}^{*}\, z_{\ell 1}^{*2}$$

$$+8\sum_{i=1}^{n}\frac{1}{(n-i+1)^3}\sum_{j=i}^{n}\sum_{k=i}^{n}\sum_{\ell=i}^{n} z_{j2}^{*}\, z_{k1}^{*}\, z_{\ell 1}^{*}\, z_{\ell 2}^{*} + 2\sum_{i=1}^{n}\frac{1}{(n-i+1)^3}\sum_{j=i}^{n}\sum_{k=i}^{n}\sum_{\ell=i}^{n} z_{j1}^{*} z_{k1}^{*} z_{\ell 2}^{*2}$$

$$-6\sum_{i=1}^{n}\frac{1}{(n-i+1)^4}\sum_{j=i}^{n}\sum_{k=i}^{n}\sum_{\ell=i}^{n}\sum_{m=i}^{n} z_{j1}^{*}\, z_{k1}^{*}\, z_{\ell 2}^{*}\, z_{m2}^{*} \,.$$

By symmetry $g_{0,1}(0,0)$, $g_{0,2}(0,0)$, $g_{1,2}(0,0)$, $g_{0,3}(0,0)$, $g_{0,4}(0,0)$ and $g_{1,3}(0,0)$ may be deduced from the above.

1B.4 A, B, C and X functions at $(\beta_1,\beta_2) = (0,0)$ expressed in terms of population moments

i) $E_p\{g_{2,0}(0,0)\} = n\mu_{2,0} + o(n)$

$(A)_{(0,0)} = n\mu_{2,0} + o(n)$

ii) $E_p\{g_{3,0}(0,0)\} = n\mu_{3,0} + o(n)$

$E_p\{g_{2,0}(0,0)\, g_{1,0}(0,0)\} = n\mu_{3,0} + o(n)$

$(A^{(1)})_{(0,0)} = o(n)$

(iii) $E_p\{g_{2,1}(0,0)\} = n\mu_{2,1} + o(n)$

$E_p\{g_{2,0}(0,0)\, g_{0,1}(0,0)\} = n\mu_{2,1} + o(n)$

$(A^{(2)})_{(0,0)} = o(n)$

iv) $E_p\{g_{4,0}(0,0)\} = n\mu_{4,0} - 3n\mu^2_{2,0} + o(n)$

$E_p\{g_{3,0}(0,0)\, g_{1,0}(0,0)\} = n\mu_{4,0} - 3n\mu^2_{2,0} + o(n)$

$E_p[\{g_{2,0}(0,0)\}^2] = n^2\mu^2_{2,0} + n\mu_{4,0} - 2n\mu^2_{2,0}(\omega + \log n - \frac{1}{2}) + o(n)$

$E_p[g_{2,0}(0,0)\{g_{1,0}(0,0)\}^2] = 2n\mu_{4,0} - n^2\mu^2_{2,0}$ [illegible] $) + o(n)$

$(A^{(11)})_{(0,0)} = -2n\mu^2_{2,0} + o(n).$

v) $E_p\{g_{3,1}(0,0)\} = n\mu_{3,1} - 3n\mu_{1,1}\mu_{2,0} + o(n)$

$E_p\{g_{2,1}(0,0)\, g_{1,0}(0,0)\} = -3n\,\mu_{1,1}\mu_{2,0} + n\mu_{3,1} + o(n)$

$E_p\{g_{2,0}(0,0)\, g_{1,1}(0,0)\} = n\mu_{3,1}$ [illegible] $\mu_{1,1}\mu_{2,0} - n\mu_{1,1}\mu_{2,0}$ [illegible] $) + o(n)$

$E_p\{g_{3,0}(0,0)g_{0,1}(0,0)\} = -3n\mu_{1,1}\mu_{2,0} + n\mu_{3,1} + o(n)$

$E_p\{g_{2,0}(0,0)g_{1,0}(0,0)g_{0,1}(0,0)\} = 2n\mu_{3,1} + n^2\mu_{1,1}\mu_{2,0} - 4n\mu_{1,1}\mu_{2,0}$

$- 2n\mu_{1,1}\mu_{2,0}(\omega + \log n) + o(n)$

$(A^{(12)})_{(0,0)} = -2\,\mu_{1,1}\mu_{2,0} + o(n)$

vi) $E_p\{g_{2,2}(0,0)\} = n\mu_{2,2} - n\mu_{2,0}\mu_{0,2} - 2n\mu^2_{1,1} + o(n)$

$E_p\{g_{2,1}(0,0)g_{0,1}(0,0)\} = -n\mu_{2,2} - n\mu_{2,0}\mu_{0,2} - 2n\mu^2_{1,1} + o(n)$

$E_p\{g_{2,0}(0,0)g_{0,2}(0,0)\} = -n^2\mu_{2,0}\mu_{0,2} + n\mu_{2,2} - 2n(\omega + \log n - \frac{1}{2})\mu_{2,0}\mu_{0,2}$
$+ o(n)$

$E_p[g_{2,0}(0,0)\{g_{0,1}(0,0)\}^2] = 2n\mu_{2,2} + n^2\mu_{2,0}\mu_{0,2} - 4n\mu^2_{1,1}$
$- 2n\mu_{2,0}\mu_{0,2}(\omega + \log n) + o(n)$

$(A^{(22)})_{(0,0)} = -2n\mu^2_{1,1} + o(n).$

__B functions__

These may be obtained from the above by symmetry

i) $(B)_{(0,0)} = n\mu_{0,2} + o(n)$

ii) $(B^{(1)})_{(0,0)} = o(n)$

iii) $(B^{(2)})_{(0,0)} = o(n)$

iv) $(B^{(11)})_{(0,0)} = -2n\mu^2_{1,1} + o(n)$

v) $(B^{(12)})_{(0,0)} = -2n\mu_{1,1}\mu_{0,2} + o(n)$

vi) $(B^{(22)})_{(0,0)} = -2n\mu^2_{0,2} + o(n)$

C functions

i) $E_p\{g_{1,1}(0,0)\} = n\mu_{1,1} + o(n)$

$(C)_{0,0} = n\mu_{1,1} + o(n)$

ii) $E_p\{g_{2,1}(0,0)\} = n\mu_{2,1} + o(n)$

$E_p\{g_{1,1}(0,0)g_{1,0}(0,0)\} = n\mu_{2,1} + o(n)$

$(C^{(1)})_{(0,0)} = o(n)$

iii) $E_p\{g_{1,2}(0,0)\} = n\mu_{1,2} + o(n)$

$E_p\{g_{1,1}(0,0)\, g_{0,1}(0,0)\} = n\mu_{1,2} + o(n)$

$(C^{(2)})_{(0,0)} = o(n)$

iv) $E_p\{g_{3,1}(0,0)\} = n\mu_{3,1} - 3n\mu_{1,1}\mu_{2,0} + o(n)$

$E_p\{g_{2,1}(0,0)g_{1,0}(0,0)\} = -3n\mu_{1,1}\mu_{2,0} + n\mu_{3,1} + o(n)$

$E_p\{g_{1,1}(0,0)g_{2,0}(0,0)\} = n\mu_{3,1} + n^2\mu_{1,1}\mu_{2,0} - 2n\mu_{1,1}\mu_{2,0}(\log n + w - \tfrac{1}{2})$
$+ o(n)$

$E_p[g_{1,1}(0,0)\{g_{1,0}(0,0)\}^2] = 2n\mu_{3,1} + n^2\mu_{1,1}\mu_{2,0} - 2n\mu_{1,1}\mu_{2,0}$
$(\log n + w + 2) + o(n)$

$(C^{(11)})_{(0,0)} = -[illegible]\mu_{1,1}\mu_{2,0} + o(n)$

v) $E_p\{g_{2,2}(0,0)\} = n\mu_{2,2} - n\mu_{2,0}\mu_{0,2} - 2n\mu^2_{1,1} + o(n)$

$E_p\{g_{2,1}(0,0)g_{0,1}(0,0)\} = n\mu_{2,2} - n\mu_{2,0}\mu_{0,2} - 2n\mu^2_{1,1} + o(n)$

$E_p[\{g_{1,1}(0,0)\}^2] = n^2\mu^2_{1,1} + n\mu_{2,2} - 2n\mu^2_{1,1}(\log n + w - 1) + o(n)$

$E_p\{g_{1,2}(0,0)g_{1,0}(0,0)\} = n\mu_{2,2} - n\mu_{2,0}\mu_{0,2} - 2n\mu^2_{1,1} + o(n)$

$E_p\{g_{1,1}(0,0)g_{0,1}(0,0)g_{1,0}(0,0)\} = 2n\mu_{2,2} + n^2\mu^2_{1,1} - 2n\mu_{2,0}\mu_{0,2}$
$-2n\mu^2_{1,1}(\log n + w + 1) + o(n)$

$(C^{(12)})_{(0,0)} = -[illegible]\mu_{2,0}\mu_{0,2} - n\mu^2_{1,1} + o(n)$

vi) $E_p(g_{1,3}(0,0)) = n\mu_{1,3} - 3n\mu_{1,1}\mu_{0,2} + o(n)$

$E_p(g_{1,2}(0,0)g_{0,1}(0,0)) = n\mu_{1,3}$ [illegible] $\mu_{1,1}\mu_{0,2} + o(n)$

$E_p(g_{1,1}(0,0)g_{0,2}(0,0)) = n\mu_{1,3} + n^2\mu_{1,1}\mu_{0,2}$ [illegible] $+o(n)$

$E_p[g_{1,1}(0,0)(g_{0,1}(0,0))^2] = 2n\mu_{1,3} + n^2\mu_{1,1}\mu_{0,2} - 2n\mu_{1,1}\mu_{0,2}(\log n$ [illegible] $2) + o(n)$

$(C^{(12)})_{(0,0)} = -2n\mu_{1,1}\mu_{0,2} + o(n)$

X functions

i) $(X)_{(0,0)} = (AB-C^2)_{(0,0)} = n^2(\mu_{2,0}\mu_{0,2} - \mu^2_{1,1}) + o(n)$

ii) $(X^{(1)})_{(0,0)} = (A^{(1)}B + AB^{(1)} - 2CC^{(1)})_{(0,0)} = o(n^2)$

iii) similarly $(X^{(2)})_{(0,0)} = o(n^2)$

iv) $(X^{(11)})_{(0,0)} = (A^{(11)}B + 2A^{(1)}B^{(1)} + AB^{(11)} - 2C^{(1)2} - 2CC^{(11)})_{(0,0)}$

$= n^2(2\mu^2_{1,1}\mu_{2,0}$ [illegible] $\mu_{2,0}\mu_{0,2}) + o(n^2)$

v) $(X^{(12)})_{(0,0)} = (A^{(12)}B + A^{(1)}B^{(2)} + A^{(2)}B^{(1)} + AB^{(12)} - 2C^{(1)}C^{(2)} - 2CC^{(12)})_{(0,0)}$

$= n^2(2\mu^3_{1,1} - 2\mu$ [illegible] $)$

vi) $(X^{(22)})_{(0,0)} =$ [illegible] $+ AB^{(22)} - 2C^{(2)2} - 2CC^{(22)})_{(0,0)}$

$= n^2(2\mu^2_{1,1}\mu_{0,2}$ [illegible] $) + o(n^2)$

[illegible].5 Evaluation of terms in Taylor series

$\left(\log \frac{m}{X}\right)_{(0,0)} = \log\left\{\frac{n\mu_{0,2} + o(n)}{n^2(\mu_{2,0}\mu_{0,2} - \mu^2_{1,1}) + o(n^2)}\right\}$

$= \log\left\{\frac{\mu_{0,2}}{n(\mu_{2,0}\mu_{0,2} - \mu^2_{1,1})} + o\left(\frac{1}{n}\right)\right\}$

$$\theta_1\left\{\frac{B^{(1)}}{B}-\frac{X^{(1)}}{X}\right\}_{(0,0)} = \theta_1\left\{\frac{[illegible]}{n\mu_{0,2}+[illegible]} + \frac{[illegible]}{n^2(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)+o(n^2)}\right\}$$

$$= o(1)$$

Similarly

$$\theta_2\left\{\frac{B^{(2)}}{B}-\frac{X^{(2)}}{X}\right\}_{(0,0)} = o(1).$$

$$\frac{\theta_1^2}{2}\left\{\frac{B^{(11)}}{B}-\frac{X^{(11)}}{X}+\frac{X^{(1)^2}}{X^2}-\frac{B^{(1)^2}}{B^2}\right\}_{(0,0)}$$

$$= \frac{\theta_1^2}{2}\left\{\frac{[illegible]+o(n)}{n\mu_{0,2}+o(n)} - \frac{n^2(2\mu_{1,1}^2\mu_{0,2} - [illegible])+o(n^2)}{n^2(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)+o(n^2)} + o(1)\right\}$$

$$= \theta_1^2\,\frac{(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)}{\mu_{0,2}} + o(1).$$

$$\theta_1\theta_2\left\{\frac{B^{(12)}}{B}-\frac{X^{(12)}}{X}-\frac{B^{(1)}B^{(2)}}{B^2}-\frac{X^{(1)}X^{(2)}}{X^2}\right\}_{(0,0)}$$

$$= \theta_1\theta_2\left\{-\frac{[illegible]+o(n)}{n\mu_{0,2}+o(n)} - \frac{n^2([illegible])+o(n^2)}{n^2(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)+o(n^2)} + o(1)\right\}$$

$$= o(1).$$

Thus in the neighbourhood of $(\theta_1,\theta_2) = (0,0)$,

$$\log M_n(\theta_1,\theta_2) = \log\left\{\frac{\mu_{0,2}}{n(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)} + o\left(\frac{1}{n}\right)\right\}$$

$$+\ \theta_1^2\,\frac{(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)}{\mu_{0,2}} + o(1).$$

Appendix C   The asymptotic variance of $\hat{\beta}_{IV\,1}$

§C.1 Taylor series expansion

Put $A = I_{11}^{IV}\ (\beta_1, \beta_2)$, $B = I_{22}^{IV}\ (\beta_1, \beta_2)$, $C = I_{12}^{IV}\ (\beta_1, \beta_2)$ and $X = AB - C^2$. Again $V_{IV}(\beta_1, \beta_2) = \frac{B}{X}$ and expanding $\log V_{IV}\ (\beta_1, \beta_2)$ about $(\beta_1, \beta_2) = (0,0)$ gives terms as in §B.1.

§C.2 A, B and C functions evaluated at $(\beta_1, \beta_2) = (0,0)$

These may be obtained as sums over strata of corresponding model I quantities, (see §B.4).

A functions

$$(A)_{(0,0)} = n \sum_{j=1}^{s} q_j\ \mu_{(j)2,0} + \sum_{j=1}^{s} o(n_j) = n\mu_{2,0} + o(n)$$

$$(A^{(1)})_{(0,0)} = o(n) = (A^{(2)})_{(0,0)}$$

$$(A^{(11)})_{(0,0)} = -2n \sum_{j=1}^{s} q_j\ \mu^2_{(j)2,0} + o(n)$$

$$(A^{(12)})_{(0,0)} = -2n \sum_{j=1}^{s} q_j\ \mu_{(j)1,1}\ \mu_{(j)2,0} + o(n)$$

$$(A^{(22)})_{(0,0)} = -2n \sum_{j=1}^{s} q_j\ \mu^2_{(j)1,1} + o(n)$$

B functions

$$(B)_{(0,0)} = n\mu_{0,2} + o(n)$$

$$(B^{(1)})_{(0,0)} = o(n) = (B^{(2)})_{(0,0)}$$

$$(B^{(11)})_{(0,0)} = -2n \sum_{j=1}^{s} q_j\ \mu^2_{(j)1,1} + o(n)$$

$$(B^{(12)})_{(0,0)} = -2n \sum_{j=1}^{s} q_j\ \mu_{(j)1,1}\ \mu_{(j)0,2} + o(n)$$

$$(B^{(22)})_{(0,0)} = -2n \sum_{j=1}^{s} q_j\ \mu^2_{(j)0,2} + o(n)$$

C functions

$$(C)_{(0,0)} = n\mu_{1,1} + o(n)$$

$$(C^{(1)})_{(0,0)} = o(n) = (C^{(2)})_{(0,0)}$$

$$(C^{(11)})_{(0,0)} = -2n \sum_{j=1}^{g} q_j \, \mu_{(j)1,1} \, \mu_{(j)2,0} + o(n)$$

$$(C^{(12)})_{(0,0)} = -n \sum_{j=1}^{g} q_j \, \mu_{(j)2,0} \, \mu_{(j)0,2} + n \sum_{j=1}^{g} q_j \, \mu^2_{(j)1,1} + o(n)$$

$$(C^{(22)})_{(0,0)} = -2n \sum_{j=1}^{g} q_j \, \mu_{(j)1,1} \, \mu_{(j)0,2} + o(n)$$

§C.3 Evaluation of X functions at $(\theta_1, \theta_2) = (0,0)$

$$(X)_{(0,0)} = n^2 (\mu_{2,0} \mu_{0,2} - \mu^2_{1,1}) + o(n^2)$$

$$(X^{(1)})_{(0,0)} = o(n^2) = (X^{(2)})_{(0,0)}$$

$$(X^{(11)})_{(0,0)} = 2n^2 (2\mu_{1,1} (\sum_{j=1}^{g} q_j \, \mu_{(j)1,1} \, \mu_{(j)2,0}) - \mu_{0,2}(\sum_{j=1}^{g} q_j \, \mu^2_{(j)2,0})$$
$$- \mu_{2,0} (\sum_{j=1}^{g} q_j \, \mu^2_{(j)1,1})) + o(n^2) .$$

$$(X^{(12)})_{(0,0)} = 2n^2 (\mu_{2,0} (\sum_{j=1}^{g} q_j \, \mu_{(j)2,0} \, \mu_{(j)0,2}) + \mu_{1,1} (\sum_{j=1}^{g} q_j \, \mu^2_{(j)1,1})$$
$$- \mu_{0,2} (\sum_{j=1}^{g} q_j \, \mu_{(j)1,1} \, \mu_{(j)2,0}) - \mu_{2,0} (\sum_{j=1}^{g} q_j \, \mu_{(j)1,1} \, \mu_{(j)0,2})) + o(n^2)$$

$$(X^{(22)})_{(0,0)} = 2n^2 (2\mu_{1,1} (\sum_{j=1}^{g} q_j \, \mu_{(j)1,1} \, \mu_{(j)0,2}) - \mu_{2,0}(\sum_{j=1}^{g} q_j \, \mu^2_{(j)0,2})$$
$$- \mu_{0,2}(\sum_{j=1}^{g} q_j \, \mu^2_{(j)1,1})) + o(n^2) .$$

§C.4 Evaluation of terms in Taylor series

$$\left[\log \frac{B}{X}\right]_{(0,0)} + \log \left\{ \frac{n\mu_{0,2} + o(n)}{n^2(\mu_{2,0}\,\mu_{0,2} - \mu_{1,1}^2) + o(n^2)} \right\}$$

$$= \log \left\{ \frac{\mu_{0,2}}{n(\mu_{2,0}\,\mu_{0,2} - \mu_{1,1}^2)} + o\left(\frac{1}{n}\right) \right\}$$

$$\theta_1 \left(\frac{B^{(1)}}{B} - \frac{X^{(1)}}{X}\right)_{(0,0)} = o(1) = \theta_2 \left(\frac{B^{(2)}}{B} - \frac{X^{(2)}}{X}\right)_{(0,0)}$$

$$\theta_1^2 \left(\frac{B^{(11)}}{B} - \frac{X^{(11)}}{X} + \frac{X^{(1)2}}{X^2} - \frac{B^{(1)2}}{B^2}\right)_{(0,0)}$$

$$= \left\{ \mu_{0,2}^2 \Big(\sum_{j=1}^{s} q_j \mu_{(j)2,0}^2\Big) - 2\mu_{0,2}\mu_{1,1} \Big(\sum_{j=1}^{s} q_j\, \mu_{(j)1,1}\, \mu_{(j)2,0}\Big) + \mu_{1,1}^2 \Big(\sum_{j=1}^{s} q_j\, \mu_{(j)1,1}^2\Big) \right\} \Big/ \left\{ \mu_{0,2}(\mu_{2,0}\,\mu_{0,2} - \mu_{1,1}^2) \right\} + o(1)$$

$$\theta_1\theta_2 \left(\frac{B^{(12)}}{B} - \frac{X^{(12)}}{X} - \frac{B^{(1)}B^{(2)}}{B^2} - \frac{X^{(1)}X^{(2)}}{X^2}\right)_{(0,0)}$$

$$= 2\theta_1\theta_2 \left\{ \mu_{1,1}^2 \Big(\sum_{j=1}^{s} q_j \mu_{(j)1,1}\mu_{(j)0,2}\Big) + \mu_{0,2}^2 \Big(\sum_{j=1}^{s} q_j \mu_{(j)1,1}\mu_{(j)2,0}\Big) + \mu_{0,2}\mu_{1,1}\Big(\sum_{j=1}^{s} q_j \mu_{(j)2,0}\mu_{(j)0,2} + \sum_{j=1}^{s} q_j \mu_{(j)1,1}^2\Big) \right\} \Big/$$

$$\left\{ \mu_{0,2}(\mu_{2,0}\mu_{0,2} - \mu_{1,1}^2) \right\} + o(1).$$

$$\theta_2^{\,2}\left(\frac{B^{(22)}}{B}-\frac{X^{(22)}}{X}+\frac{X^{(2)2}}{X^2}-\frac{B^{(2)2}}{B^2}\right)_{(0,0)}$$

$$=\frac{\theta_2^{\,2}}{n}\Big\{\mu_{1,1}^2\Big(\sum_{j=1}^{s}q_j\mu^2_{(j)0,2}\Big)-2\mu_{0,2}\mu_{1,1}\Big(\sum_{j=1}^{s}q_j\mu_{(j)1,1}\mu_{(j)0,2}\Big)$$

$$+\mu_{0,2}^2\Big(\sum_{j=1}^{s}q_j\mu^2_{(j)1,1}\Big)\Big\}\Big/\Big\{\mu_{0,2}(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)\Big\}\qquad+o(1)$$

Thus in the neighbourhood of $(\theta_1,\theta_2)=(0,0)$

$$\log V_{IV}(\theta_1\theta_2)=\log\left\{\frac{\mu_{0,2}}{n(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)}+o\left(\frac{1}{n}\right)\right\}$$

$$+\Big\{\mu_{0,2}\sum_{j=1}^{s}q_j(\theta_1\mu_{(j)2,0}+\theta_2\mu_{(j)1,1})^2+\mu_{1,1}^2\sum_{j=1}^{s}q_j(\theta_1\mu_{(j)1,1}+\theta_2\mu_{(j)0,2})^2$$

$$-2\mu_{0,2}\mu_{1,1}\sum_{j=1}^{s}q_j(\theta_1\mu_{(j)1,1}+\theta_2\mu_{(j)0,2})(\theta_1\mu_{(j)2,0}+\theta_2\mu_{(j)1,1})+o(1)\Big\}\Big/$$

$$\Big\{\mu_{0,2}(\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2)\Big\}$$

§D.1 General results

Let $T_1, T_2, \ldots, T_n$ be independent random variables with T_i having p.d.f. $p_i(t/\underline{\theta})$, where $\underline{\theta} = (\theta_1, \ldots, \theta_q)'$ is a q x 1 vector of parameters. In addition suppose that $\exists$ functions $h_1(\cdot)$, $h_2(\cdot)$, $\ldots$, $h_n(\cdot)$, such that for $i = 1, \ldots, n$, $\varepsilon_i = h_i(T_i ; \underline{\theta})$ and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent and indentically distributed random variables. For $i = 1, \ldots, n$ let $R_i = h_i(T_i ; \hat{\underline{\theta}})$ denote the 'crude' residual corresponding to the random variable T_i where $\hat{\underline{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)'$ denotes the maximum likelihood estimator of $\underline{\theta}$. The results of Cox and Snell (1968) indicate that for $s = 1, \ldots, q$

$$b_s = E(\hat{\theta}_s - \theta_s) = \frac{1}{2} \sum_{r=1}^{q} I^{rs} \sum_{t=1}^{q} \sum_{u=1}^{q} I^{tu}(K_{rtu} + 2J_{t,ru}) + o(\tfrac{1}{n})$$

where $U_r^{(i)} = \dfrac{\partial \log p_i(T_i/\underline{\theta})}{\partial \theta_r}$,

$$V_{rs}^{(i)} = \frac{\partial^2 \log p_i(T_i/\underline{\theta})}{\partial \theta_r \, \partial \theta_s}, \quad W_{rst}^{(i)} = \frac{\partial^3 \log p_i(T_i/\underline{\theta})}{\partial \theta_r \, \partial \theta_s \, \partial \theta_t} \quad i = 1, \ldots, n,$$

and

$$I_{rs} = E\left(-\sum_{i=1}^{n} V_{rs}^{(i)}\right), \quad K_{rst} = E\left(\sum_{i=1}^{n} W_{rst}^{(i)}\right)$$

$$J_{r,st} = E\left(\sum_{i=1}^{n} U_r^{(i)} V_{st}^{(i)}\right), \quad r,s,t = 1, \ldots, q$$

where $\underset{q\times q}{\underline{I}^{-1}} = [I^{ij}]$ with $\underset{q\times q}{\underline{I}} = [I_{ij}]$.

In addition putting $H_r^{(i)} = \dfrac{\partial h_i(T_i;\underline{\theta})}{\partial \theta_r}$, $H_{rs}^{(i)} = \dfrac{\partial^2 h_i(T_i;\underline{\theta})}{\partial \theta_r \, \partial \theta_s}$

the above authors show that, to $o(\frac{1}{n})$.

$$E(R_i) = E(\epsilon_i) + \sum_{r=1}^{q} b_r E(H_r^{(i)}) + \sum_{r=1}^{q}\sum_{s=1}^{q} I^{rs} E(H_r^{(i)} U_s^{(i)} + \tfrac{1}{2} H_{rs}^{(i)}) \qquad \text{D 2.}$$

$$E(R_i^2) = E(\epsilon_i^2) + 2\sum_{r=1}^{q} b_r\, E(\epsilon_i H_r^{(i)})$$

$$+ 2\sum_{r=1}^{q}\sum_{s=1}^{q} I^{rs}\, E(\epsilon_i H_r^{(i)} U_s^{(i)} + \tfrac{1}{2} H_r^{(i)} H_s^{(i)} + \tfrac{1}{2}\epsilon_i H_{rs}^{(i)}) \qquad \text{D 3.}$$

$$E(R_i R_j) = \left\{E(\epsilon_i)\right\}^2 + (a_i + a_j)\, E(\epsilon_i)$$

$$+ \sum_{r=1}^{q}\sum_{s=1}^{q} I^{rs}\, E(\epsilon_i H_r^{(j)} U_s^{(i)} + \epsilon_j H_r^{(i)} U_s^{(j)} + H_r^{(i)} H_s^{(j)})$$

$$i \neq j \qquad \text{D 4.}$$

§D.2 Model III results

Evaluation of bias terms

Under model III, $q = p + 1$, $\underline{\theta} = (\beta_1, \ldots, \beta_p, \lambda)$ and

$\log p_i(t/\underline{\theta}) = \log\lambda + \underline{\beta}'\underline{z}_i - \lambda e^{\underline{\beta}'\underline{z}_i} t$ so that

$$U_j^{(i)} = z_{ij} - \lambda T_i z_{ij} e^{\underline{\beta}'\underline{z}_i} \qquad j=1, \ldots, p$$

$$U_{p+1}^{(i)} = \frac{1}{\lambda} - T_i e^{\underline{\beta}'\underline{z}_i}$$

$$V_{jk}^{(i)} = -\lambda T_i z_{ij} z_{ik} e^{\underline{\beta}'\underline{z}_i}$$

$$V_{j\,p+1}^{(i)} = -T_i z_{ij} e^{\underline{\beta}'\underline{z}_i} = V_{p+1\,j}^{(i)} \qquad j,k=1, \ldots, p$$

$$V_{p+1\,p+1}^{(i)} = -\frac{1}{\lambda^2}$$

$$W_{jk\ell}^{(i)} = -\lambda T_i z_{ij} z_{ik} z_{i\ell} e^{\underline{\beta}'\underline{z}_i}$$

$$W_{p+1\,jk}^{(i)} = -T_i z_{ij} z_{ik} e^{\underline{\beta}'\underline{z}_i} = W_{jp+1k}^{(i)} = W_{jk\,p+1}^{(i)}$$

$$W_{p+1\,p+1j}^{(i)} = 0 = W_{p+1\,j\,p+1}^{(i)} = W_{j\,p+1\,p+1}^{(i)} \qquad j,k,\ell = 1\ldots, p$$

$$W_{p+1\,p+1\,p+1}^{(i)} = \frac{2}{\lambda^3}$$

Using the results of §A.4 it follows that

$$\underline{I}^{-1} = [I^{ij}] = \left[\begin{array}{c:c} \left[\sum_{i=1}^{n} z_{ij}\, z_{ik}\right]^{-1} & \underset{p\times 1}{\underline{0}} \\ \hdashline \underset{1\times p}{\underline{0}} & \lambda^2/n \end{array}\right]$$

$$K_{jk\ell} = -\sum_{i=1}^{n} z_{ij}\, z_{ik}\, z_{i\ell}$$

$$K_{p+1\ jk} = -\frac{1}{\lambda}\sum_{i=1}^{n} z_{ij}\, z_{ik} = K_{j\ p+1\ k} = K_{jk\ p+1}$$

$$K_{p+1\ p+1\ j} = 0 = K_{p+1\ j\ p+1} = K_{j\ p+1\ p+1} \quad j,k,\ \ell=1,\dots,p$$

$$K_{p+1\ p+1\ p+1} = \frac{2n}{\lambda^3}$$

and $$J_{j,k\ell} = \sum_{i=1}^{n} z_{ij}\, z_{ik}\, z_{i\ell}$$

$$J_{j,k\ p+1} = \frac{1}{\lambda}\sum_{i=1}^{n} z_{ij}\, z_{ik} = J_{j,\ p+1\ k} = J_{p+1,\ jk}$$

$$J_{p+1,\ j\ p+1} = J_{p+1,p+1\ j} = J_{j,p+1\ p+1} = J_{p+1,\ p+1\ p+1} = 0$$

Thus from D 1. to $o(\frac{1}{n})$

$$b_j = E(\hat{\beta}_j - \beta_j) = \frac{1}{2}\sum_{r=1}^{p}\sum_{t=1}^{p}\sum_{u=1}^{p} I^{rj}\, I^{tu}\sum_{i=1}^{n} z_{ir}\, z_{it}\, z_{iu} \quad j=1,\dots,p$$

$$b_{p+1} = E(\hat{\lambda} - \lambda) = \frac{\lambda}{2n}(2 + p)$$

<u>Moments of crude residuals</u>

Putting $h_i(T_i;\ \underline{\theta}) = \lambda e^{\underline{\beta}'\underline{z}_i}\, T_i = \varepsilon_i \qquad i=1,\dots,n$

it follows that

$$E(H_r^{(i)}) = z_{ir} \quad r=1,\ldots,p, \quad E(H_{p+1}^{(i)}) = \frac{1}{\lambda}$$

$$E(H_{rs}^{(i)}) = z_{ir}\, z_{is}, \quad E(H_{r\,p+1}^{(i)}) = E(H_{p+1\,r}^{(i)}) = \frac{z_{ir}}{\lambda} \quad r=1,\ldots,p,$$

$$E(H_{p+1\,p+1}^{(i)}) = 0$$

and $E(H_r^{(i)} U_s^{(i)}) = -z_{ir}\, z_{is}$,

$$E(H_r^{(i)}\, U_{p+1}^{(i)}) = E(H_{p+1}^{(i)}\, U_r^{(i)}) = -\frac{z_{ir}}{\lambda} \quad r,s=1,\ldots,p,$$

$$E(H_{p+1}^{(i)}\, U_{p+1}^{(i)}) = -\frac{1}{\lambda^2}.$$

Thus using D 2. it follows that, to $o(\frac{1}{n})$, for $i=1,\ldots,n$

$$E(R_i) = 1 + \frac{p}{2n} + \sum_{r=1}^{p} b_r\, z_{ir} - \frac{1}{2}\left[\sum_{r=1}^{p}\sum_{s=1}^{p} I^{rs}\, z_{ir}\, z_{is}\right]$$

$$= 1 + a_i \qquad \text{D 5.}$$

In addition

$$E(\epsilon_i\, H_r^{(i)}) = 2\, z_{ir} \qquad E(\epsilon_i\, H_{p+1}^{(i)}) = \frac{2}{\lambda}$$

$$E(\epsilon_i\, H_r^{(i)} U_s^{(i)}) = -4\, z_{ir}\, z_{is} \quad E(\epsilon_i H_{p+1}^{(i)}\, U_s^{(i)}) = E(\epsilon_i H_s^{(i)}\, U_{p+1}^{(i)}) = -\frac{4}{\lambda^2} z_{is}$$

$$r,s=1,\ldots,p,$$

$$E(\epsilon_i\, H_{p+1}^{(i)}\, U_{p+1}^{(i)}) = -\frac{4}{\lambda^2}.$$

$$E(H_r^{(i)} H_s^{(i)}) = 2\, z_{ir}\, z_{is}, \quad E(H_{p+1}^{(i)}\, H_s^{(i)}) = \frac{2 z_{is}}{\lambda} \quad r,s=1,\ldots,p,$$

$$E(H_{p+1}^{(i)}\, H_{p+1}^{(i)}) = \frac{2}{\lambda^2}.$$

$$E(\epsilon_i H_{rs}^{(i)}) = 2\, z_{ir}\, z_{is}, \quad E(\epsilon_i H_{p+1\,s}^{(i)}) = E(\epsilon_i H_{s\,p+1}^{(i)}) = \frac{2 z_{is}}{\lambda} \quad r,s=1,\ldots,p,$$

$$E(\epsilon_i H_{p+1\,p+1}^{(i)}) = 0.$$

Thus from D 3. to $o(\frac{1}{n})$,

$$E(R_i^2) = 2 - \frac{2}{n} + \frac{2a_i}{n} + 4\left\{ \sum_{i=1}^{p} b_r z_{ir} - \sum_{r=1}^{p} \sum_{s=1}^{p} I^{rs} z_{ir} z_{is} \right\} \quad i=1,\ldots,n$$

Finally, for $i \neq j$

$$E(u_i R_r^{(j)} u_s^{(ii)}) = -z_{is} z_{jr},\ E(u_i R_{p+1}^{(j)} u_s^{(ii)}) = E(u_j R_r^{(ii)} u_{p+1}^{(j)}) = -\frac{z_{js}}{\lambda}$$

$$r,s=1,\ldots,p$$

$$E(u_i R_{p+1}^{(j)} u_{p+1}^{(ii)}) = -\frac{1}{\lambda^2}.$$

$$E(R_r^{(i)} R_s^{(j)}) = z_{ir} z_{js},\quad E(R_r^{(i)} R_{p+1}^{(j)}) = \frac{z_{ir}}{\lambda} \quad r,s=1,\ldots,p,$$

$$E(R_{p+1}^{(i)} R_{p+1}^{(j)}) = \frac{1}{\lambda^2}$$

so that, from D 4., to $o(\frac{1}{n})$,

$$E(R_i R_j) = 1 + (a_i + a_j) - \frac{1}{n} - \sum_{r=1}^{p} \sum_{s=1}^{p} I^{rs} z_{is} z_{jr} \quad i,j=1,\ldots,p \quad i \neq j,$$

and $\operatorname{cov}(R_i, R_j) = E(R_i R_j) - E(R_i)E(R_j)$

$$= 1 + (a_i + a_j) - \frac{1}{n} - \sum_{r=1}^{p} \sum_{s=1}^{p} I^{rs} z_{is} z_{jr} - (1+a_i)(1+a_j) + o(\frac{1}{n})$$

$$= -\frac{1}{n} - \sum_{r=1}^{p} \sum_{s=1}^{p} I^{rs} z_{is} z_{jr} + o(\frac{1}{n}).$$

§D.3 Model VI results

Evaluation of bias terms

Under model VI, $q = p+m$, $\underline{\theta} = (\beta_1,\ldots,\beta_p,\lambda_1,\ldots,\lambda_m)$ and

$$\log F_{ji}(t|\underline{\theta}) = \log\lambda_j + \underline{\beta}'\underline{z}_{ji} - \lambda_j e^{\underline{\beta}'\underline{z}_{ji}} t, \quad j=1,\ldots,m \quad i=1,\ldots,n_j.$$

N.B. $z_{jik} = x_{jik} - \bar{x}_{jk}$ where $\bar{x}_{jk} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{jik}$ $k=1,\ldots,p$.

Considerations as in model III yield

$$\underline{I}^{-1} = \left[\begin{array}{c:c} \left[\sum_{j=1}^{s} \sum_{i=1}^{n_j} z_{jik}\, z_{ji\ell}\right]^{-1} & \underset{p\times s}{\underline{0}} \\ \hdashline \underset{s\times p}{\underline{0}} & \operatorname{diag}_s \left[{}^{\lambda_j}/_{n_j}\right] \end{array}\right]$$

$$K_{k\ell m} = -\sum_{j=1}^{s} \sum_{i=1}^{n_j} z_{jik}\, z_{ji\ell}\, z_{jim}$$

$$K_{k\ell\ p+j} = -\frac{1}{\lambda_j} \sum_{i=1}^{n_j} z_{jik}\, z_{ji\ell} = K_{k\ p+j\ell} = K_{p+j\ k\ell}$$

$$K_{k\ p+j\ p+r} = 0 = K_{p+j\ k\ p+r} = K_{p+j\ p+r\ k} \qquad k,\ell,\ m=1,\ldots,p$$

$$K_{p+j\ p+r\ p+t} = \begin{cases} 2n_j/\lambda_j^3 & j=r=t \\ 0 & \text{otherwise,} \end{cases} \qquad r,t=1,\ldots,s.$$

$$J_{k,\ell m} = \sum_{j=1}^{s} \sum_{i=1}^{n_j} z_{jik}\, z_{ji\ell}\, z_{jim}$$

$$J_{k,\ell\ p+j} = \frac{1}{\lambda_j} \sum_{i=1}^{n_j} z_{jik}\, z_{ji\ell} = J_{k,p+j\ \ell} = J_{p+j,\ k\ell}$$

$$J_{k,\ p+j\ p+r} = 0 = J_{p+j,\ k\ p+r} = J_{p+j,\ p+r\ k} \qquad k,\ell,m = 1,\ldots,p$$

$$= J_{p+j,\ p+r\ p+t} \qquad j,r,t=1,\ldots,s$$

Thus to $o(\frac{1}{n})$

$$b_k = E(\hat{\beta}_k - \beta_k) = \frac{1}{2} \sum_{r=1}^{p} \sum_{t=1}^{p} \sum_{u=1}^{p} I^{rk} I^{tu} \left(\sum_{j=1}^{s} \sum_{i=1}^{n_j} z_{jir}\, z_{jit}\, z_{jiu} \right) \qquad k=1,\ldots,p$$

$$b_{p+j} = \frac{\lambda_j}{2n_j}\,(2+p) \qquad j=1,\ldots,s.$$

Moments of crude residuals

Putting $h_{ji}(T_{ji};\ \theta) = \lambda_j\, e^{\beta' z_{ji}}\, T_{ji} = \epsilon_{ji}$ $\qquad j=1,\ldots,s$, $i=1,\ldots,n_j$

calculations identical to those of §D .2 yield, to $o(\frac{1}{n})$

$$E(R_{ji}) = 1 + \frac{p}{2n_j} + \sum_{k=1}^{p} b_j\, z_{jik} - \frac{1}{2} \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{jik}\, z_{ji\ell}$$

$$= 1 + a_{ji}$$

and

$$E(R_{ji}^2) = 2 - \frac{2}{n_j} + \frac{2p}{n_j} + 4\left(\sum_{k=1}^{p} b_k\, z_{jik} - \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{jik}\, z_{ji\ell} \right).$$

For $ji \neq j_1 i_1$

$$E(\epsilon_{ji}\, H_k^{(j_1 i_1)} U_\ell^{(ji)}) = - z_{ji\ell}\, z_{j_1 i_1 k}$$

$$E(\epsilon_{ji}\, H_k^{(j_1 i_1)} U_{p+r}^{(ji)}) = \begin{cases} -z_{j_1 i_1 k}/\lambda_j & j=r \\ 0 & \end{cases}$$

$$= E(\epsilon_{j_1 i_1}\, H_{p+r}^{(ji)}\, U_k^{(j_1 i_1)}) \qquad k,\ t=1,\ldots,p$$

$$E(\epsilon_{ji}\, H_{p+r}^{(j_1 i_1)} U_{p+t}^{(ji)}) = \begin{cases} -1/\lambda_j \lambda_{j_1} & j=t,\ j_1=r \\ 0 & \text{otherwise} \end{cases}$$

$r, t=1,\ldots,s.$

$$E(H_k^{(ji)} H_\ell^{(j_1 i_1)}) = z_{jik}\, z_{j_1 i_1 \ell}$$

$$E(H_k^{(ji)} H_{p+r}^{(j_1 i_1)}) = \begin{cases} z_{jik}/\lambda_{j_1} & r=j_1 \\ 0 & \text{otherwise} \end{cases}$$

$k,\ \ell=1,\ldots,p$

$$E(H_{p+r}^{(ji)}\ H_{p+t}^{(j_1 i_1)}) = \begin{cases} 1/\lambda_j \lambda_{j_1} & j=r,\ j_1=t \\ 0 & \text{otherwise} \end{cases} \qquad r,t=1,\ldots,s$$

so that for $j = j_1$

$$E(R_{ji}\ R_{ji_1}) = 1 - \frac{1}{n_j} + (a_{ji} + a_{ji_1}) - \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{ji\ell}\ z_{ji_1k}$$

where if $j \neq j_1$

$$E(R_{ji}\ R_{j_1 i_1}) = 1 + (a_{ji} + a_{j_1 i_1}) - \sum_{k=1}^{p} \sum_{\ell=1}^{p} I^{k\ell} z_{ji\ell} z_{j_1 i_1 k}\ .$$

REFERENCES

ALTSHULER, B. (1970) 'Theory for the measurement of competing risks in animal experiments' Math. Biosciences, 6, 1-11.

BARNARD, G. A. (1963) 'Some logical aspects of the fiducial argument' J. R. Statist. Soc. B, 25, 111-114.

BRESLOW, N. (1970) 'A generalised Kruskall-Wallis test for comparing samples subject to unequal patterns of censoring' Biometrika, 57, 579-594.

BRESLOW, N. (1974) 'Covariance analysis of censored survival data' Biometrics, 30, 89-99.

BRESLOW, N. and CROWLEY, J. (1974) 'A large sample study of the life table and product limit estimates under random censorship' Ann. of Stat., 2, 437-453.

BYAR, D. P. and MANTEL, N. (1975) 'Some interrelationships among the regression coefficient estimates arising in a class of models appropriate to response-time data' Biometrics, 31, 943-947.

COX, D. R. (1953) 'Some simple approximate tests for Poisson variates' Biometrika, 40, 354-360.

COX, D. R. (1972) 'Regression models and life tables' J. R. Statist. Soc., B, 34, 187 - 220.

COX, D. R. (1975) 'Partial likelihood' Biometrika, 62, 269-276.

COX, D. R. and SNELL, E. J. (1968) 'A general definition of residuals' J. R. Statist. Soc., B, 30, 248-275.

CROWLEY, J. (1973) 'Non-parametric analysis of censored survival data with distribution theory for the k-sample generalised Savage statistic' Ph.D. Thesis, University of Washington.

CROWLEY, J. (1974) 'A note on some recent likelihoods leading to the log rank test' Biometrika, 61, 533-538.

CURNOW, R. N. (1972) Contribution to discussion of paper by R. Peto and J. Peto, J. R. Statist. Soc., A, 135. 199-200.

DIXON, W. J. and MOOD, A. M. (1946) 'The statistical sign test' J. Amer. Statist. Assoc., 41, 557-566.

EFRON, B. (1967) 'The two sample problem with censored data' Proc. Fifth Berkeley Symp., 4, 831-854.

FELDSTEIN, M. L. (1973) 'Distribution-free methods for comparing survival experiences and the use of the likelihood function in the analysis of clinical data on Breast Cancer' Ph.D. Thesis, State University of New York at Buffalo.

FIEGL, P. and ZELEN, M. (1965) 'Estimation of exponential survival probabilities with concomitant information' Biometrics, 21, 826-838.

FRASER, D. A. S. (1968) 'The structure of inference' New York: John Wiley and Sons, Inc.

FREIREICH, E. J. et. al. (1963) 'The effect of 6 mercaptopurine on the duration of steroid-induced remissions in acute leukaemia' Blood, [illegible], 699-716.

GEHAN, E. (1965a) 'A generalised Wilcoxon test for comparing arbitrarily single censored samples' Biometrika, 52, 203-223.

GEHAN, E. (1965b) 'A generalised two-sample Wilcoxon test for doubly censored data' Biometrika, 52, 650-653.

GEHAN, E. (1972) Contribution to discussion of paper by R. Peto and J. Peto, J. R. Statist. Soc., A, 135. 204-205.

GEHAN, E. and THOMAS, D. G. (1969) 'The performance of some two-sample tests in small samples with and without censoring' Biometrika, 56, 127-132.

GILBERT, J. P. (1962) 'Random censorship' Ph.D. Thesis, University of Chicago.

GLASSER, M. (1967) 'Exponential survival with covariance' J. Amer. Statist. Assoc., 62, 561-568.

GREENBERG, R. A., BAYARD, S. and BYAR, D. (1974) 'Selecting concomitant variables using a likelihood ratio step-down procedure and a method of testing goodness of fit in an exponential survival model' Biometrics, 30, 601-608.

HARTLEY, H. O. and HOCKING, R. R. (1971) 'The analysis of incomplete data' Biometrics, 27, 783-824.

HOLT, J. D. and PRENTICE, R. L. (1974) 'Survival analysis in twin studies and matched pair experiments' Biometrika, 61, 17-30.

JOHNSON, R. A. and MEHROTRA, K. G. (1972) 'Locally most powerful rank tests for the two-sample problem with censored data' An. of Math. Statist., 43, 823-831.

JUHL, E., WINKEL, P., TYGSTRUP, N. and the COPENHAGEN STUDY GROUP FOR LIVER DISEASES (1974) 'Sex, ascites and alcoholism in survival of patients with cirrhosis: Effect of Prednisone' N. Engl. J. Med., 291, 271-273.

KALBFLEISCH, J. D. (1974a) 'Some efficiency calculations for survival distributions' Biometrika, 61, 31-38.

KALBFLEISCH, J. D. (1974b) 'Some extensions and applications of Cox's regression and life model' Paper presented at the 'Joint Stat. Meeting in Tallahassee, Florida', March 1974.

KALBFLEISCH, J. D. and SPROTT, D. (... 'Application of likelihood methods to models involving lar... numbers of parameters' J.R. Statist. Soc., B, 32, 175...

KALBFLEISCH, J. D. and PRENTICE, R. ... 1972) Contribution to discussion of paper by D. R. Cox. J.R. St...t. Soc., B, 34, 215-216.

KALBFLEISCH, J. D. and PRENTICE, R. ... (1973) 'Marginal likelihoods based on Cox's regression and ... model' Biometrika, 60, 267-278.

KAPLAN, E. L. and MEIER, P. (1958) '...n-parametric estimation from incomplete observations' J. A... Statist. Assoc., 53, 457-481.

KENDALL, M. G. and STUART, A. (197... ...he Advanced Theory of Statistics, vol. II' London: Charles Griff... & Co.

LEE, E. T., DESU, M. M. and GEHA... ... 1975) 'A Monte Carlo study of the power of some two-sample ...s' Biometrika, 62, 425-432.

MANTEL, N. (1966) 'Evaluation of ... ...val data and two new rank order statistics arising in its con...ration' Cancer Chem. Rep., 50, 163-170.

MANTEL, N. (1967) 'Ranking proc... ...re for arbitrarily restricted observations' Biometrics, 23 ... 78.

MANTEL, N. and HANKEY, B. F. (197... A logistic regression analysis of response-time data where t... ...ard function is time dependent' Unpublished paper.

MANTEL, N. and MYERS, M. (1971) ...blems of convergence of maximum likelihood iterative procedures in multi-parameter situations' J. Amer. Statist. Assoc., 66, ...4-491.

MYERS, M. H., HANKEY, B. F. and MA... ... N. (1973) 'A logistic-exponential model for use with ... ...ime data involving regressor variables' Biometrics, ..., 257-269.

NELSON, W. and HAHN, G. J. (1972a) 'Linear estimation of a regression relationship from censored data - Part 1. Simple methods and their applications' Technometrics, 14, 247-269.

NELSON, W. and HAHN, G. J. (1972b) 'Linear estimation of a regression relationship from censored data - Part 2. Best linear unbiased estimation and theory' Technometrics, 15, 133-150.

OAKES, D. (1972) Contribution to discussion of paper by D. R. Cox, J.R. Statist. Soc., B, 34, 208.

PETO, R. (1972) 'Rank tests of maximal power against Lehmann-type alternatives' Biometrika, ..., 472-475.

PETO, R. (1973) 'Experimental survival curves for interval censored data' J.R. Statist. Soc., C, 22, 86-91.

PETO, R. and PETO, J. (1972) 'Asymtotically efficient rank invariant test procedures' J.R. Statist. Soc., A, 135, 185-206.

PRENTICE, R. L. (1973) 'Exponential survivals with censoring and explanatory variables' Biometrika, 60, 279-284.

SAMPFORD, M. R. (1954) 'The estimation of response time distribution, III : Truncation and survival' Biometrics, 10, 531-561.

SAMPFORD, M. R. and TAYLOR, J. (1959) 'Censored observations in randomized block experiments' J.R. Statist. Soc., B, 21, 214-237.

SCHALM, S. W. and SUMMERSKILL, M. D. (1975) 'Treatment of Cirrhosis with Prednisone' N. Engl. J. Med., 292, 1030-1031.

SPIERS, A. S. D., BAIKIE, A. G. et.al. (1975) 'Chronic Granulocytic Leukaemia : Effect of Elective Splenectomy on the course of disease' B.M.J., 1, 175-179.

THOMAS, D. R. (1971) 'On a generalised Savage statistic for comparing two arbitrarily censored samples' Unpublished paper.

TURNBULL, B. W., BROWN, B. W. and HU, M. (1974) 'Survivorship analysis of Heart transplant data' J. Amer. Statist. Assoc., 69, 74-80.

WILCOXON, F. (1945) 'Individual comparisons by ranking methods' Biometrics, 1, 80-83.

ZELEN, M. (1971) 'The analysis of several 2x2 contingency tables' Biometrika, 58, 129-137.

ZELEN, M. (1973) 'Contingency tables and the evaluation of survival data' Unpublished paper.

ZIPPEN, C. and ARMITAGE, P. (1966) 'Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter' Biome[illegible] 22, 665-672.