SOME ASPECTS OF SURVIVAL ANALYSIS

Robert Dan Mensah

A Thesis

in

The Department

of

Mathematics

Presented in Partial Fulfillment of the Requirements
for the degree of Master of Science at
Concordia University
Montréal, Québec, Canada

August, 1981

ABSTRACT

SOME ASPECTS OF SURVIVAL ANALYSIS

Robert Dan Mensah

Survivability (length of life studies) theory is desired to make probability statements about survival or remission times of acutely ill patients or animals undergoing biological experimentation. This thesis is an expository survey of the survival analysis literature, developments and advances made in solving some of the complex problems in this field in a systematic manner. It presents methodologies appropriate in analyzing survival data either from laboratory studies on animals or survival studies of acutely ill humans. We draw quite extensively on reliability theory, showing how some of the estimation developed for problems in reliability can be applied to corresponding problems with biomedical application.

The exposition emphasizes the newer, research aspects of survivability theory. A number of new classes of life distributions arising naturally in survival models are treated systematically. Parametric and nonparametric approaches in the presence of censored and uncensored survival times are considered.

Suppose that in a life-testing situation the failure of

an individual can be classified into one of $(k > 1)$ mutually exclusive classes, usually causes of failure. It is assumed that associated with each cause of failure there is a characteristic life distribution belonging to a specific class of distributions. The estimation of parameters in such compound models have been investigated.

Concomitant information on a patient's condition often accompanies survival time information. This has led to recent introduction of dealing with survival data using regression models. We examine some of these models.

ACKNOWLEDGEMENTS

The author wishes to express his indebtedness to Dr. Y.P. Chaubey for his guidance, encouragement and support in the preparation of this thesis.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

SURVIVAL DISTRIBUTIONS AND THEIR APPLICATIONS

1.1 INTRODUCTION

Survivability theory is concerned with the measurement of length of life. This length of life could be that of a mechanism, a human being, or an animal. A very important problem in survival analysis is the comparison of survival times of patients or experimental animals receiving two or more different treatments. Similar problems also occur in reliability theory when it is necessary to compare life distributions before and after mechanical changes. However, reliability theory differs with survivability theory in an important aspect. In reliability theory the common problem is the optimization of the system reliability by varying the number and arrangements of different components in the system. However, in survivability theory this aspect is absent because the system under consideration, in this case is living and functioning thus making the rearrangements of organs an impractical alternative.

Survival analysis has been developed in several different disciplines as well: epidemiology, demography, and actuarial science. Because of this diversity of origin survival data can

take so many different forms, ranging from results of small-scale laboratory tests to massive records from long-term clinical trials. Thus each case may require different criterion of evaluation and may require different procedures of analysis.

The survival time of a patient, say with cancer is a major criterion for evaluating the treatment the patient received. In carcinogenesis experiments, the time to development of tumour is an important end point in the analysis of animal experiments in which potentially carcinogenic agents are administered. In the evaluation of individuals or combinations of agents in transplanted animal systems, the survival time of the animal is a major (and sometimes the only) end point in the analysis of the study. Hence it is worthwhile to consider methodology for analyzing survival times. Thus "survival time" is meant in the broad sense, so that for example, the times may be: length of response, time to recurrence of the disease, time from start of treatment to first response time, or some other function of response. Since the modeling and analysis of such data have as a principal terminal point, the time an event occurs, we refer to such events as failures; thus in the literature we shall use failure time or survival time interchangeably.

There has been a remarkable increase of activity in the statistical analysis of survival data over the last two decades, largely stimulated by problems arising in the analysis of

clinical trials which has resulted in a considerable literature on the topic. Most of the literature has endeavoured to set out general principles to be used in each particular case, although they do not rule out the possibility of the researcher developing his own methodology.

In Section 1.2, we shall consider notations and definitions used in describing survival time data as well as its mathematical formulation and Section 1.3 gives a general outline of the thesis.

1.2 NOTATIONS AND MATHEMATICAL FORMULATION

Survival data consists of measurements of times to the death of individuals. If we let T be a non-negative random variable denoting the survival time of an individual from a homogeneous population, we can characterise the probability distribution of T by three equivalent functions which are useful in survival analysis.

1.2.1 The Death Density Function

The death density function, $f(t)$ is the probability that an individual dies during the time interval $t < T < t + \Delta t$, for $\Delta t \to 0$. This is a probability density function, where the random variable is time, T. The death density function is sometimes called the unconditional failure rate, which is mathematically expressed as:

$$f(t) = \lim_{\Delta t \to 0^+} \frac{\Pr\{t < T < t + \Delta t\}}{\Delta t} . \qquad (1.2.1)$$

The function, $f(t)$ possesses the following properties:

(i) $f(t) \geq 0, \quad \forall t .$

(ii) $\int_{-\infty}^{\infty} f(t)dt = 1 .$

Since survival time is only measured for positive values of t,

$$f(t) = 0 \quad \text{for} \quad t < 0$$

$$\therefore \int_{0}^{\infty} f(t)dt = 1$$

1.2.2 Survivorship Function

The survivorship function, $S(t)$ is the probability that an individual survives for at least time $t(t > 0)$.

That is,

$$S(t) = P(T \geq t), \quad 0 < t < \infty \qquad (1.2.2)$$

is called the Survivorship function.

The cumulative distribution function $F(t)$ is defined as

$$F(t) = \int_{-\infty}^{t} f(T)dT ,$$

and since $t > 0$

$$F(t) = \int_{0}^{t} f(T)dT .$$

Thus

$$S(t) = 1 - F(t) = \int_{t}^{\infty} f(T)dT .$$

Assuming $F(t)$ is differentiable in t, where $t \geq 0$, then the death density function is given by

$$f(t) = \lim_{\Delta t \to 0^+} \left[\frac{\text{Prob. individual dies in } (t,t+\Delta t)}{\Delta t}\right]$$

$$= F'(t)$$

$$= -\frac{dS(t)}{dt} = -S'(t) . \qquad (1.2.3)$$

It is evident from the definition of $S(t)$ that $S(0) = 1$ and

$$\lim_{t \to \infty} S(t) = 0 .$$

$S(t)$ is also a monotone, non-increasing, left-continuous function.

1.2.3 Hazard Function, $\lambda(t)$

The probability that an individual dies in the time interval $(t,t+\Delta t)$ no matter how small Δt is, given he has survived up to the time t is given by $\lambda(t)\Delta t$.

That is, the hazard function specifies the instantaneous rate of failure at $T = t$ conditional upon survival up to time t. It is also referred to as the failure rate, the instantaneous death rate, or the force of mortality. Mathematically, it is defined as:

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{\Pr\{t \leq T < t + \Delta t | T \geq t\}}{\Delta t} . \qquad (1.2.4)$$

We also define $\lambda(T|t)$ to be a conditional death density function for an individual who dies at time $T > t$, given that he has survived up to time t, for a fixed value of t. Since $\lambda(T|t)$ is a death density function,

$$\int_t^\infty \lambda(T|t)dT = 1 . \qquad (1.2.5)$$

Again, since $\lambda(T|t)$ is a death density function, it is related to the unconditional death density function as follows:

$$\lambda(T|t) = \alpha(t)f(T) \qquad \text{if } T \geq t$$
$$= 0 \qquad \text{if } T < t$$

where

$\alpha(t)$ is a constant of proportionality depending only on t.

From Equation (1.2.5), it follows that

$$\alpha(t) = \frac{1}{S(t)}$$

by observing that

$$\frac{1}{S(t)} \int_t^{\infty} f(T)\, dT = \frac{S(t)}{S(t)} = 1. \qquad (1.2.6)$$

If we observe also that $\lambda(t)$, the conditional failure rate at time t, is the same as $\lambda(T|t)$, it then follows that

$$\lambda(t) = \frac{f(t)}{S(t)} \qquad (1.2.7)$$

$$\therefore \quad \lambda(t) = \lim_{\Delta t \to 0^+} \frac{Pr\{t < T \leq t + \Delta t | T > t\}}{\Delta t} = \frac{f(t)}{S(t)}.$$

We could therefore draw up a relationship among $S(t)$, $f(t)$ and $\lambda(t)$.

Rewrite Equation (1.2.7) as:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt} \qquad (1.2.8)$$

Integrating and using $S(0) = 1$ we obtain

$$S(t) = \exp(-\int_0^t \lambda(u)du) . \qquad (1.2.9)$$

Thus, the death density function is expressed as:

$$f(t) = \lambda(t) \exp(-\int_0^t \lambda(u)du) . \qquad (1.2.10)$$

We have therefore three connecting relationships among $\lambda(t)$, $f(t)$, and $S(t)$, which are

$$\left.\begin{array}{ll} \text{(i)} & f(t) = -S'(t), \\ \text{(ii)} & \lambda(t) = \dfrac{f(t)}{S(t)}, \\ \text{(iii)} & S(t) = \exp[-\int_0^t \lambda(u)du]. \end{array}\right\} \qquad (1.2.11)$$

We also sometimes use the notation:

$$\Lambda(t) = \int_0^t \lambda(u)du = -\log S(t)$$

which is called the cumulative hazard function (c.h.f.).

Even though any of the three functions $\lambda(t)$, $f(t)$ or $S(t)$ uniquely defines a specific survival distribution, each provides the researcher a different view of the data. For example, the shape of $\lambda(t)$ gives an indication of the type of risk to which the population under study is exposed as a function of time. If $\lambda(t)$ is increasing, we know that there is an aging process that increases the rate of death in the

population. The death density function, f(t), may be used to assess the peak period of death in the population. If the researcher requires the time when, say 60% of his sample survives, he would use $\hat{S}(t)$, the estimated survival probability.

Example:

Suppose $\lambda(t) = \lambda$ independent of t, then it implies from (iii) of Equation (1.2.11), that

$$S(t) = \exp(-\lambda t) . \qquad (1.2.12)$$

Thus, if the hazard rate is constant as a function of time, the survival distribution is the negative exponential distribution.

So far, we have only considered the case when the random variable T is continuous.However, T can be a discrete random variable which takes values $y_1 < y_2 < y_3 < \ldots\ldots$ The death density function associated with each y_i is given as

$$f(y_i) = P(T = y_i) \quad \text{for} \quad i = 1,2,3,\ldots$$

Define the indicator function U(.) as

$$U(t-a) = \begin{cases} 0 & \text{if} \quad t < a \\ 1 & \text{if} \quad t \geq a . \end{cases}$$

Then the survivorship function is given as

$$S(t) = \sum_{i \mid y_i \geq t} f(y_i)$$

$$= \sum f(y_i)\, U(y_i - t) .$$

The conditional probability of failure at y_i is also defined as

$$\lambda_i = Pr\{T = y_i \mid T \geq y_i\}$$

$$= \frac{f(y_i)}{S(y_i)} \qquad \text{for} \qquad i = 1,2,3,\ldots$$

Analogous to Equation (1.2.11) (iii), the survivorship function could be written as

$$S(t) = \prod_{i \mid y_i < t} (1-\lambda_i) .$$

And also from [(1.2.11),(iii)] the death density function will be expressed as

$$f(y_i) = \lambda_i \prod_{j=1}^{i-1} (1-\lambda_j) . \qquad (1.2.13)$$

In a more general case, if the distribution of T has both discrete and continuous parts then the death density function is expressed as a sum of the discrete and continuous components.

In a special case, if λ_k denotes the hazard function

for the continuous part and the discrete points occur at $y_1 < y_2 < y_3 < \ldots$, then the overall survivorship function can be written as

$$S(t) = \exp[-\int_0^t \lambda_k(u)du] \prod_{i|y_i<t} (1-\lambda_i)$$

We can therefore combine the discrete, continuous and mixed cases. The hazard function could be written as

$$\lambda(t)dt = \lambda_k(t)dt + \Sigma\lambda_i\delta(t-y_i)dt$$

where

$\delta(t-y_i)$ is the Dirac delta function defined as

$$\delta(y)dy = \begin{cases} 1 & \text{if } y = 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that if T is discrete, then

$$\lambda(t) = \Sigma\lambda_i\delta(t-y_i)$$

where

$\delta(t-y_i)$ is as defined above and $\lambda_t = \Pr\{T=t|T\geq t\}$

We define the cumulative hazard function as

$$\Lambda(t) = \int_0^t \lambda(u)du = \int_0^t \lambda_k(u)du + \sum_{i|y_i<t} \lambda_i$$

where the Dirac delta components give the discrete contributions to the integral. By using the product law of probability, Kalbfleisch and Prentice (1980) have derived the survivorship function in the discrete, continuous, or mixed cases as

$$S(t) = {}_0P^t\,[I - d\Lambda(u)]$$

where they define the product integral ${}_0P^t$ as

$${}_0P^t\,[1 - d\Lambda(u)] = \lim_{r\to\infty} \prod_{1}^{r} \{1 - [\Lambda(u_k) - \Lambda(u_{k-1})]\}$$

with $0 = u_0 < u_1 < \ldots < u_r = t$ and $u_k - u_{k-1} \to 0$ as $r \to \infty$.

We could also write $S(t)$ as

$$S(t) = {}_0P^t\,[1 - \lambda(u)du]$$

where the Dirac delta function takes care of the discrete contributions.

1.3 OUTLINE OF THE THESIS

The main purpose of this thesis is to explore some statistical models and methods used in analyzing survival or failure time data. Though such data evolve also from industrial and engineering applications my attention has been directed to problems coming from the medical sciences. Thus the survival time studies considered are planned studies (e.g.

clinical trials or laboratory experiments) in which a primary purpose is to characterize and compare the survival experience following the administration of two or more treatments. Analysis of a consecutive series of cases with a particular disease is also considered in which possible aims of the study are to characterize the survival time and to delineate patient characteristics related to long survival. We may also give analogous example from engineering and industrial application whenever it helps to explain a point.

In clinical, and other experimental trials, measurements of further characters beyond just time (or age) and mortality, are often obtained. Some, if not all, of these may be expected to have some association with failure rates. The variables corresponding to such characters are often referred to as concomitant or explanatory variables, or briefly covariates. For example, the elevated blood sugar of diabetic patients, the high level of cholesterol in cardiovascular diseases and many more cases like this have motivated the use of regression models in the analysis of survival data. We shall discuss one such model as introduced by Cox (1972) in detail as well as analysis of multivariate failure time data and competing risks.

Chapter II is concerned primarily, with parametric and non-parametric estimation of survival data. When there is no prior experience for a particular survival study, the first

objective is to characterize the data obtained. Secondly, there is usually interest in choosing a survival time model that represents the main features of the data with estimates of parameters of the model. Examples of models considered are Weibull, exponential, gamma, normal, Gompertz and lognormal. We shall investigate the exponential, lognormal and the Weibull in detail. We also estimate the parameters of these distributions in the presence and absence of censoring. In addition to these parametric survival models we introduce non-parametric procedures. Those considered are the Kaplan and Meier (1958) estimates for censored data, life table methods for analyzing survival data and the graphical procedures for estimating the cumulative hazard rate, developed by Nelson (1972).

In Chapter III, we discuss multivariate failure time models with more emphasis on the bivariate failure model.

Death is not a repetitive event and it is usually attributed to a single cause, however, various risks competing for the life of an individual must be considered in any survival data. In Chapter IV, we look at the case of competing risks.

In Chapter V, we introduce the use of regression models in analyzing survival data. We shall consider the Cox (1972) model in detail.

CHAPTER II

SOME SURVIVAL MODELS

2.1 INTRODUCTION

In this Chapter we consider some statistical distributions that are used to describe survival times. Although by survival we usually signify time to death, these distributions are useful for describing other data. For example, we could be describing the length of stay in a mental hospital, wherein "birth" is entry into the hospital and "death" is discharge. We may also consider the length of remission in acute leukemia; here "birth" is the time at which the patient goes into remission, and "death" is the time at which the individual relapses.

The distributions discussed here are the exponential, gamma, Weibull, normal, lognormal and the Gompertz, of which the exponential Weibull and the lognormal will be considered in more detail. Our main tool for estimating the parameters of the distributions is the method of maximum likelihood for the following reasons:

(i) Conceptually it is a simple procedure, although the computational problems may not always be simple.

(ii) The asymptotic properties of maximum likelihood

estimators (under fairly general conditions) make their use desirable, and

(iii) Maximum likelihood estimation affords a rather general method of estimation of parameters of survival distributions; even when observations are censored, for example, one can in most instances obtain the maximum likelihood estimators of the parameters of the survival distribution.

For some models explicit solution of the Maximum Likelihood estimators (MLE's) may be possible and it will be presented where feasible. However, for other models solutions cannot be obtained explicitly. For this reason two computational methods of finding MLE and their comparison are obtained in Section 2.2.

Section 2.3 considers different parametric models and the maximum likelihood estimation of parameters and their properties. Often it is not known prior to data collection what survival distribution is appropriate. To investigate this, model fitting is an important aspect in survival analysis. This is considered in Section 2.4.

In Section 2.5 , we present non-parametric methods of estimation of the survival curve. Essentially, our task is to present methods appropriate for:

(i) Small sample data, with each individual time of death recorded effectively exactly;

(ii) Large sample data, with times of death usually grouped into fixed intervals.

The purpose of this Chapter is also to review methods that have been proposed for estimating the survivorship function from incomplete data - that is, data arising from experiments in which censoring of the data occurs due to withdrawals from the study.

2.2 COMPUTATIONAL ASPECTS OF MLE

We briefly review the method of maximum likelihood (ML). Suppose $t_1, t_2, \ldots, t_n$ are n independent survival times whose death density function is a function of the parameters $\theta_1, \theta_2, \ldots, \theta_k$, $n > k$. We define the likelihood function (LF), $L(\theta_1, \theta_2, \ldots, \theta_k)$ given the sample $t_1, t_2, \ldots, t_n$ as

$$L(\theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} f(t_i; \theta_1, \theta_2, \ldots, \theta_k) \quad (2.2.1)$$

where

$f(t_i; \theta_1, \ldots, \theta_k)$ is the death density function.

The maximum likelihood estimators $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ of $\theta_1, \theta_2, \ldots, \theta_k$, respectively, based on the sample $t_1, t_2, \ldots, t_n$ are such that if $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_k$ is any other set of estimators,

then

$$L(\hat{\theta}_1,\ldots,\hat{\theta}_k) \geq L(\tilde{\theta}_1,\ldots,\tilde{\theta}_k)$$

Under regularity conditions outlined in Cramér (1946) the MLE's $\hat{\theta}_1,\ldots,\hat{\theta}_k$ of $\theta_1,\ldots,\theta_k$ are obtained as the solution to the $k \times k$ system of equations.

$$\left.\frac{\partial \log L(\theta_1,\ldots,\theta_k)}{\partial \theta_i}\right|_{\substack{=0 \\ \theta_i = \hat{\theta}_i}} \qquad i=1,2,\ldots,k \qquad (2.2.2)$$

The estimators $(\hat{\theta}_1,\ldots,\hat{\theta}_k)$ are asymptotically normally distributed with mean $(\theta_1,\ldots,\theta_k)$ and variance-covariance matrix $V_{\underset{\sim}{\hat{\theta}}}$ [see Rao (1952)], where

$$V_{\underset{\sim}{\hat{\theta}}} = \begin{bmatrix} -E\left(\frac{\partial^2 \log L}{\partial \theta_1^2}\right) & \cdots & -E\left(\frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_k}\right) \\ \vdots & & \vdots \\ -E\left(\frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_k}\right) & \cdots & -E\left(\frac{\partial^2 \log L}{\partial \theta_k^2}\right) \end{bmatrix} \qquad (2.2.3)$$

that is,

$$\begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix} \sim N\left[\begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}; \; V_{\underset{\sim}{\hat{\theta}}}\right] \qquad (2.2.4)$$

In cases where it is not possible to obtain explicit MLE's for the parameters of the survival distributions we use numerical techniques. There are two cases of the numerical techniques:

(i) No constraints on values the parameters assume, and

(ii) Parameters are subject to constraint.

For example, in Case 2, in maximizing the logarithm of likelihood functions, the constraint, if there is one, is typically as follows: The value of a parameter must lie interior to a particular region (e.g., $0 < p < 1$) and must not lie on the boundary of that region. Numerical procedures that do not allow for constraints can be used as long as successive maximum likelihood estimator lies in the interior of the region.

We describe here maximizing techniques that do not consider constraints. These techniques can be put into two classes - direct and indirect. In the direct class, a starting value is determined at what is thought to be a good approximation to the desired value. One then proceeds in a stepwise fashion upward toward the maximum. An example of this class is the method of steepest ascent or gradient method of Cauchy. In the indirect method class, one obtains first the derivatives of the logarithm of the likelihood function with respect to each parameter. The resulting equations are set equal to zero, and attempt is made to find the values of the parameters

(in terms of the observations) that simultaneously satisfy these equations. Two examples of this class are the Newton-Raphson procedure and the method of scoring which we investigate below. Rao (1965) and Kale (1961,1962) have also discussed this indirect procedure. Although more sophisticated and modern methods [see Kennedy and Gentle (1980)] are available for solving non-linear equations of the type which confronts us, the Newton-Raphson procedure and the method of scoring, are very practical to implement and conceptually are not difficult.

2.2.1 The Newton-Raphson Method

The Newton-Raphson method is a widely used and often-studied method for minimization. We illustrate this technique by solving first for a single θ and then presenting the case for $\theta_i, i=1,2,\ldots,k$. Let

$$g(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} \qquad (2.2.5)$$

where

$$L(\theta) = \prod_{i=1}^{n} f(t_i;\theta).$$

The problem is then to find the value of θ, say $\hat{\theta}$, such that

$$g(\hat{\theta}) = 0 \qquad (2.2.6)$$

Thus $\hat{\theta}$ is the requisite MLE of θ.[1] If $\hat{\theta}$ cannot be obtained explicitly from solving Equation (2.2.6), we may attempt a solution by means of the Newton-Raphson procedure for which the function $g(\hat{\theta})$ is approximated by the first two terms in a Taylor series expansion; (we ignore the contribution of the higher order terms, since their deviations from the solution value $\hat{\theta}$ are raised to higher powers, hence diminish rapidly.) Let $\hat{\theta}_o$ be an initial estimate of $\hat{\theta}$, then $\hat{\theta}_\nu$, the $\nu^{\underline{th}}$ iteration or approximation of $\hat{\theta}$, is given by:

$$\hat{\theta}_\nu = \hat{\theta}_{\nu-1} - \frac{g(\hat{\theta}_{\nu-1})}{g'(\hat{\theta}_{\nu-1})} \tag{2.2.7}$$

where

$$g'(\hat{\theta}_{\nu-1}) = \left.\frac{dg(\theta)}{d\theta}\right|_{\theta=\hat{\theta}_{\nu-1}} \tag{2.2.8}$$

Equation (2.2.7) follows from the rearrangement of the first two terms in the Taylor series. Thus, Newton-Raphson method consists of solving, at each iteration, the equation

$$g(\hat{\theta}_\nu) + (\theta-\hat{\theta}_\nu)g'(\hat{\theta}_\nu) = 0$$

---

[1] If there is more than one value $\tilde{\theta}$ such that $g(\hat{\theta}) = 0$, judicious choice of an initial value is very important. In most cases, the initial value obtained is in the neighbourhood of the MLE. When in doubt, try several different initial values.

For the next iterate $\hat{\theta}_{\nu+1}$, the solution takes the form of Equation (2.2.7) by replacing $\hat{\theta}_\nu$ by $\hat{\theta}_{\nu+1}$ and $\hat{\theta}_{\nu-1}$ by $\hat{\theta}_\nu$.

Note that this amounts to selecting $\hat{\theta}_{\nu+1}$ as the point where the tangent line to $g(\theta)$ at θ_ν intersects the θ axis.

The following Figure 2.1 is a graphical representation of the Newton-Raphson method which converges. Observe that since $g'(\hat{\theta}_o) \neq 0$, the tangent line is not parallel to the axis. Where this line crosses the θ axis, we find our next approximation $\hat{\theta}_1$, and so on. Normally, we stop the iteration procedure when log $L(\theta)$ stops increasing appreciably. The value of log $L(\theta)$ should be calculated at each step, which permits us to monitor the stopping procedure.

Suppose $f(t;\theta_1,\ldots,\theta_k)$ is a death density function containing k parameters $\theta_1,\theta_2,\ldots,\theta_k$, $k \geq 2$. Furthermore, suppose the maximum Likelihood estimators $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$ of $\theta_1,\theta_2,\ldots,\theta_k$ (respectively) are found by differentiating the logarithm of the likelihood function, with respect to $\theta_1,\theta_2,\ldots,\theta_k$, setting these derivatives equal to zero and solving the resulting equations in terms of $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$. This often leads to a system of k equations in k unknowns which cannot be solved directly. We therefore extend the Newton-Raphson method to k dimensions.

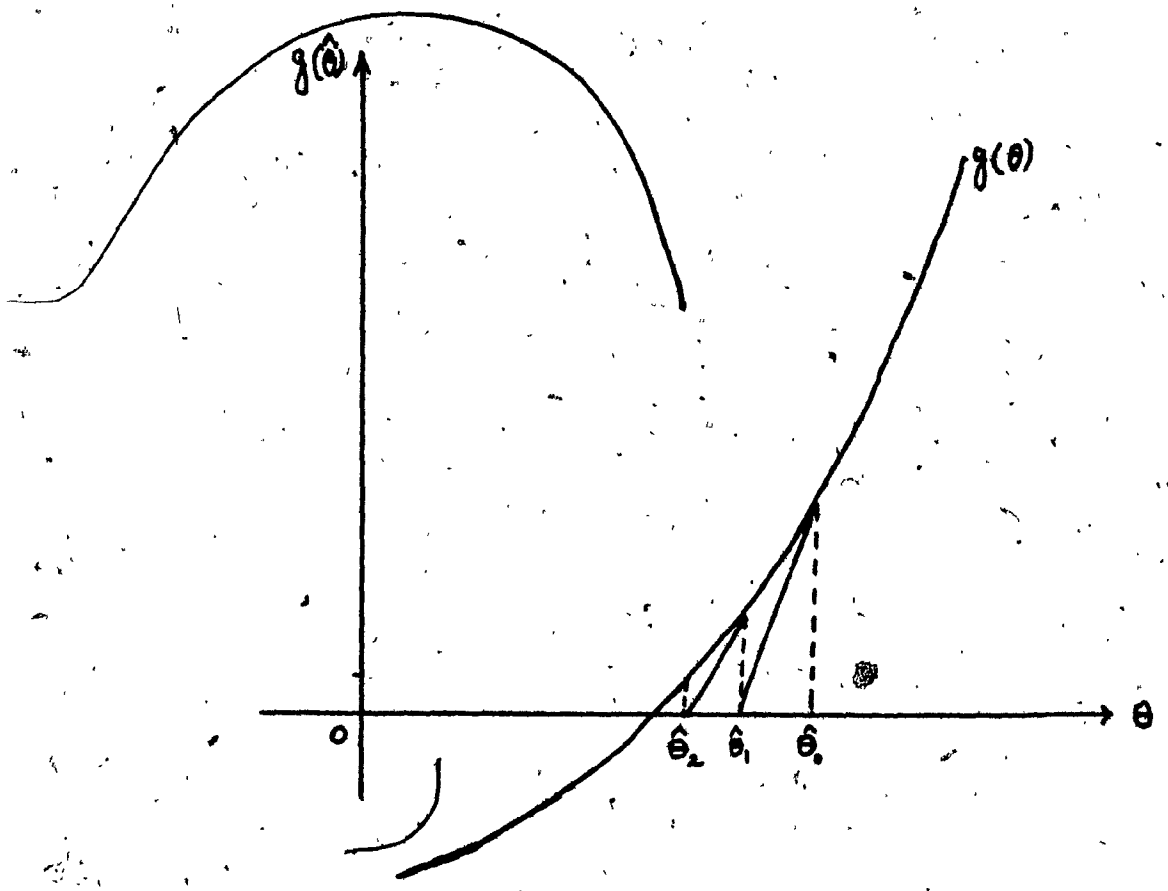

FIG. 2.1 GRAPHICAL REPRESENTATION OF THE NEWTON-RAPHSON TECHNIQUE WHICH CONVERGES

Let $L(\theta_1,\theta_2,\ldots,\theta_k)$ be the likelihood function and let us suppose that the MLE's $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$ are found by solving simultaneously the vector equation.

$$\underset{\sim}{g}'(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k) = \underset{\sim}{0}' \qquad (2.2.9)$$

in terms of $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$, where

$$\underset{\sim}{g}'(\hat{\theta}_1,\ldots,\hat{\theta}_k) = (g_1(\hat{\theta}_1,\ldots,\hat{\theta}_k),\ldots,g_k(\hat{\theta}_1,\ldots,\hat{\theta}_k))$$

and

$$g_i(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k) = \left.\frac{\partial \log L(\theta_1,\ldots,\theta_k)}{\partial \theta_i}\right|_{\theta_j=\hat{\theta}_j}, \quad i,j=1,2,\ldots,k \qquad (2.2.10)$$

Suppose $\hat{\theta}_{10},\hat{\theta}_{20},\ldots,\hat{\theta}_{ko}$ is the set of initial estimates of $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$, respectively. Then the $\nu^{\underline{th}}$ iteration $\theta_{1\nu},\theta_{2\nu},\ldots,\theta_{k\nu}$ to the solution $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$ is

$$\underset{\sim}{\hat{\theta}}' = \underset{\sim}{\hat{\theta}}'_{\nu-1} - \underset{\sim}{g}'^{(\nu-1)} ||v_{ij}||^{-1}_{(\nu-1)} \qquad (2.2.11)$$

where

$$\underset{\sim}{\hat{\theta}}'_{\nu} = (\hat{\theta}_{1\nu},\hat{\theta}_{2\nu},\ldots,\hat{\theta}_{k\nu}), \quad \underset{\sim}{\hat{\theta}}'_{\nu-1} = (\hat{\theta}_{1,\nu-1},\hat{\theta}_{2,\nu-1},\ldots,\hat{\theta}_{k,\nu-1})$$

$$g'^{(\nu-1)} = (g_1^{(\nu-1)}, g_2^{(\nu-1)}, \ldots, g_k^{(\nu-1)}),$$

$$g_i^{(\nu-1)} = g_i(\hat{\theta}_{1,\nu-1}, \hat{\theta}_{2,\nu-1}, \ldots, \hat{\theta}_{k,\nu-1}), \quad i=1,2,\ldots,k$$

(Note: The prime notation indicates the transpose of a matrix) and $||V_{ij}||_{(\nu-1)}$ is the $k \times k$ matrix whose ij^{th} element is

$$V_{ij}^{(\nu-1)} = \left.\frac{\partial g_i(\theta_1, \ldots, \theta_k)}{\partial \theta_j}\right|_{(\theta_1,\ldots,\theta_k) = (\hat{\theta}_{1,\nu-1},\ldots,\hat{\theta}_{k,\nu-1})}$$

$$i,j=1,2,\ldots,k \qquad (2,2.12)$$

or the second derivative of the log likelihood.

The choice of the initial estimates $\hat{\theta}_{10}, \hat{\theta}_{20}, \ldots, \hat{\theta}_{ko}$ is very important, because it is possible that the Newton-Raphson procedure will converge to a value that is not the maximum of $L(\theta_1, \theta_2, \ldots, \theta_k)$. Often several sets of initial values are looked at, and each time the convergent values $\hat{\theta}_{1\nu}, \hat{\theta}_{2\nu}, \ldots, \hat{\theta}_{k\nu}$ are obtained, these values are substituted into $L(\theta_1, \ldots, \theta_k)$. The overall maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are taken as that set of convergent values which maximizes $L(\theta_1, \ldots, \theta_k)$. Gross and Clark (1975) noted that although this method does not always guarantee a maximum, it is a safeguard in that more than one set of initial values is considered, which means that any peculiarities in convergence can be uncovered.

2.2.2 The Scoring Method

The method of scoring [Rao(1952)] is similar to the Newton-Raphson procedure for obtaining maximum likelihood estimates of parameters of distributions when the requisite system of equations for solutions is non-linear. The only difference between these procedures is that the matrix of second derivatives used in the Newton-Raphson technique is replaced by the matrix of the expected values of second derivatives in the method of scoring.

Suppose that $L(\theta_1,\theta_2,\ldots,\theta_k)$ is the likelihood function whose maximum likelihood estimators $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$ are found by solving Equation (2.2.9). Suppose $\hat{\theta}_{10},\hat{\theta}_{20},\ldots,\hat{\theta}_{k0}$ are the set of initial estimates of $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$, respectively. Then the $\nu^{\underline{th}}$ iteration $\hat{\theta}_{1\nu},\hat{\theta}_{2\nu},\ldots,\hat{\theta}_{k\nu}$ to the solution $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$ is:

$$\hat{\underset{\sim}{\theta}}'_{\nu} = \hat{\underset{\sim}{\theta}}'_{\nu-1} - \underset{\sim}{g}'^{(\nu-1)} ||E(V_{ij})||^{-1} \qquad (2.2.13)$$

where

$\hat{\underset{\sim}{\theta}}'_{\nu},\hat{\underset{\sim}{\theta}}'_{\nu-1}$ and $\underset{\sim}{g}'^{(\nu-1)}$ are defined in Equation (2.2.11) and

$||E(V_{ij})||$ is the matrix whose $ij^{\underline{th}}$ element is $E(V_{ij})$, evaluated at $\hat{\theta}_{\nu-1}$ where V_{ij} is $\partial^2 \log L/\partial\theta_i\partial\theta_j$.

That is the expected value of each element in the matrix is with respect to the sample values $t_1,t_2,\ldots,t_n$. For example, in the

continuous case:

$$E(V_{ij}) = E\left(\frac{\partial^2 \log L}{\partial\theta_i \, \partial\theta_j}\right) \qquad (2.2.14)$$

However, $L = \prod_{\nu=1}^{n} f(t_\nu;\underset{\sim}{\theta}')$. Hence $\log L = \sum_{\nu=1}^{n} \log f(t_\nu;\underset{\sim}{\theta}')$.

It then follows, since $t_1, t_2, \ldots, t_n$ are identically distributed, that

$$E\left(\frac{\partial^2 \log L}{\partial\theta_i \, \partial\theta_j}\right) = nE\left(\frac{\partial^2 \log f(t;\underset{\sim}{\theta}')}{\partial\theta_i \, \partial\theta_j}\right) \quad \forall \; i,j=1,2,\ldots,k$$

That is

$$E(V_{ij}) = n \int_{\theta}^{\infty} \frac{\partial^2 \log f(t;\underset{\sim}{\theta}')}{\partial\theta_i \, \partial\theta_j} \, f(t;\underset{\sim}{\theta}')dt \qquad (2.2.15)$$

A similar argument with the sum replacing the integral in Equation (2.2.15) holds in the discrete case. For some distributions, determining Equation (2.2.15) is quite simple and the method of scoring is useful. For others the integration is difficult (involving series expansions, numerical integration, etc.,), and in those cases the Newton-Raphson technique is usually employed. As with the Newton-Raphson method, a judicious choice of the starting value $\hat{\theta}_{10}, \ldots, \hat{\theta}_{k0}$ is quite important. A discussion of the Newton-Raphson method and the method of scoring is given by Kale (1961, 1962).

2.3 PARAMETRIC MODELS FOR SURVIVAL TIME STUDIES

Here, we consider some theoretical models for survival time distribution. Two general situations may be distinguished.

(i) There is detailed knowledge of the way in which failures take place so that a model is known.

(ii) Some survival time data are available and one objective of the analysis is to choose a survival time model.

In the former situation when data are available, the model is fitted and the "goodness of fit" is tested. In the latter case, the model is chosen after analysis of the data, and the validity of the model is tested by fitting the model to several sets of other data.

There are several reasons why fitting a survival time model might be desirable. Firstly, if a good model can be found, the survival experience of groups of individuals can be characterized economically in terms of a few parameters which, in turn, may offer some insight into the mechanism of failure. Secondly, when survival studies are being performed in sequence, such as clinical trials during successive periods of time by a cooperative group, information about early failures in the second and later studies would be easier to inter-

pret if a survival model had been fitted to the data from the first study. More generally, data from a current study could be compared with that predicted based on the results of a past study. Thirdly, in studies in which two or more survival time distributions are compared, it is generally true that a more powerful test of the difference between distributions can be made if a model for the survival distributions is assumed known. Finally, in some situations, it may be possible to relate the parameters of a survival distribution to prognostic characteristics of the individuals in a study. The latter possibility was considered by Fiegl and Zelen (1965) and Breslow (1974) when the survival time distribution is exponential. More often, the mathematical distribution may help to elucidate the nature of the phenomenon under investigation.

2.3.1 The Exponential Distribution

The one parameter exponential distribution could be used as a survival model if we get the hazard function to be a constant, that is $\lambda(t) = \lambda > 0$, over the range of the random variable T, representing the failure time with t being a point in its range. The probability density function and the survivorship function are respectively:

$$f(t) = \lambda \exp(-\lambda t) \qquad (2.3.1.1)$$

and

$$S(t) = \exp(-\lambda t) \qquad (2.3.1.2)$$

The constant hazard rate gives an indication of an instantaneous failure rate which does not depend on t so that the conditional probability of failure in a time interval of any given length is the same irrespective of how long the individual has been on trial. However, it is reasonable to assume an exponential death density function, if the cause of death occurs according to a Poisson process with a constant failure rate (see Feller Vol.II). If,for instance, an individual is subjected to random events, such as blood clot or thrombosis that causes the body to "fail" if and, only if that event occurs, we would expect the exponential death density to govern the length of life of the individual. We can always check the appropriateness of the use of the exponential model through the techniques given in Section 2.4.

To estimate the survival parameter λ we could use the ML method described in Section 2.2. The likelihood function, is:

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda t_i)$$

Maximizing the logarithm of $L(\lambda)$ we have

$$\text{Log } L(\lambda) = n \log\lambda - \lambda \sum_{i=1}^{n} t_i$$

Differentiating with respect to λ and equating to zero, we have:

$$\hat{\lambda} = \left(\frac{\sum_{i=1}^{n} t_i}{n}\right)^{-1} = \bar{t}^{-1}$$

which is the ML estimator of λ - the constant hazard rate. If $\mu = \lambda^{-1}$ is the mean time to death, then since the maximum likelihood estimates are invariant under one-to-one transformations,

$$\hat{\mu} = \bar{t}$$

is the maximum likelihood estimator of μ.

We can easily show that the minimum variance unbiased estimator (MVUE) for μ is, say, $\tilde{\mu}$. First, we construct $\tilde{\mu}$ in such a way that it is an unbiased linear estimator of μ. That is,

$$\tilde{\mu} = \sum_{i=1}^{n} a_i t_i$$

where $\sum_{i=1}^{n} a_i = 1$. Then $a_1, \ldots, a_n$ must be chosen so that $\mathrm{Var}(\tilde{\mu})$ achieves a minimum subject to the constraint $\sum_{i=1}^{n} a_i = 1$. We easily find that $a_i = \frac{1}{n}$ for all i; thus:

$$\tilde{\mu} = \bar{t}$$

That is, the maximum likelihood and minimum variance unbiased estimator of μ coincide when the parent density function is the exponential. In estimation theory terminology, $\bar{t}$ is the best linear unbiased estimator (BLUE) of μ. However, $\bar{t}^{-1}$

is not the BLUE for λ, since BLUEs are not generally preserved under inversing.

$\hat{\mu}$ is also the sufficient statistic for μ since

$$L(\mu) = \left(\frac{1}{\mu^n}\right)\exp\left(\frac{-n\bar{t}}{\mu}\right)$$

$$= g(t,\mu)h(t_1,\ldots,t_n)$$

where

$$h(t_1,\ldots t_n) \equiv 1$$

Therefore, $\bar{t}$ is the maximum likelihood, MVUE, as well as the sufficient statistic for μ when $t_1,\ldots,t_n$ is a random sample of size n from a population with hazard rate a constant.

Suppose patients with some serious illness (e.g., carcinoma of the lung) are on study to determine their average survival time. However, the study lasts only for a limited period. Hence, it is likely that not all patients are dead at the end of the study. There are two cases to be considered here:

(1) The patients enter the study independently at different points in time. Suppose we have n patients on study. Let T_i be the maximum time for which the $i^{\underline{th}}$ patient can be observed, i=1,2,...,n. Thus, if the study ends at time Z_i, then $T_i = T - Z_i$ (since we

assume a constant hazard function λ, we can proceed to estimate λ under the assumption that $T_i = T - Z_i$).

The survival time of the $i^{\underline{th}}$ patient t_i is known only if $t_i \leq T_i$. Thus, given a sample of n patients, the information available is the set of maximum times $T_1, T_2, \ldots, T_n$ and those times to death t_j, such that $t_j \leq T_j$. The sample size n is fixed in advance of the study, but d, the total number of deaths observed, is a random variable, because the study is assumed to end at the fixed time T.

(2) The n patients who enter the study are assumed to enter at the same time. For example, suppose we wish to study the length of survival of individuals who are exposed to excessive amounts of radiation at the same time (e.g., persons who survived the atomic bomb attack at Hiroshima.) In such a study, the survival times arrive in a natural ordering, meaning that the individual with the shortest survival time is recorded first, the one with the next shortest survival time is recorded second, and so on. If the survival times are recorded for the first $d \leq n$ individuals who die, then

$$t_{(1)} \leq t_{(2)} \leq, \ldots, \leq t_{(d)}$$

We discuss first progressively censored data and secondly, singly censored data.

Let us assume that each patient has the same death density function

$$f(t) = \lambda \exp(-\lambda t) \qquad \lambda > 0, t \geq 0$$

Following Bartholomew's (1957) development for the progressively censored case, the probability, the $i^{\underline{th}}$ individual dies while he is on study is

$$1 - S(T_i) = \int_0^{T_i} \lambda \exp(-\lambda t)\, dt = 1 - \exp(-\lambda T_i) = Q_i$$

Let $\mu = \frac{1}{\lambda}$. To estimate μ (hence λ) we use the Maximum Likelihood (ML) method.

The contribution due to the $i^{\underline{th}}$ individual, is

$$f(t_i) = \begin{cases} \mu^{-1} \exp(-\mu^{-1} t_i) & 0 \leq t_i \leq T_i \\ \exp(-\mu^{-1} T_i) & t_i > T_i \end{cases} \qquad (2.3.1.3)$$

Equation (2.3.1.3) follows because an individual either dies at $t_i \leq T_i$ with density $\mu^{-1}\exp(-\mu^{-1}t_i)$ or he survives beyond time T_i. Should the latter situation occur, we can only measure the probability of his survival, which is $\exp(-\mu^{-1}T_i)$ or $S(T_i)$.

The likelihood function is given by

$$L(\mu) = \prod_{i=1}^{n} [\mu^{-1}\exp(-\mu^{-1}t_i)]^{\delta_i}[\exp(-\mu^{-1}T_i)]^{1-\delta_i} \qquad (2.3.1.4)$$

Similarly, for any survival distribution $S(T)$ and corresponding death density function $f(t)$, we can write

$$L(\mu) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i}[S(T_i)]^{1-\delta_i}$$

where

$$\delta_i = \begin{cases} 1 & \text{if the } i^{\underline{th}} \text{ patient dies in the interval } 0 \leq t_i \leq T_i \\ 0 & \text{if he does not die in the interval } 0 \leq t_i \leq T_i \end{cases}$$

Taking logarithms, we find that

$$\log L = -\sum_{i=1}^{n} [\delta_i(\log\mu + \mu^{-1}t_i) + (1-\delta_i)\mu^{-1}T_i]$$

Setting $\left.\frac{\partial \log L}{\partial \mu}\right|_{\mu=\hat{\mu}} = 0$, we have

$$\sum_{i=1}^{n} [\delta_i(-\hat{\mu}^{-1} + \hat{\mu}^{-2}t_i) + (1-\delta_i)\hat{\mu}^{-2}T_i] = 0 \qquad (2.3.1.5)$$

The value $\hat{\mu}$ that satisfies Equation (2.3.1.5), thus maximizes Equation (2.3.1.4), is

$$\hat{\mu} = d^{-1}\sum_{i=1}^{n} (\delta_i t_i + (1-\delta_i)T_i) \qquad (2.3.1.6)$$

where

$d = \sum_{i=1}^{n} \delta_i$ is assumed to be greater than zero.

The $\hat{\mu}$ given by Equation (2.3.1.6) is thus the sum of the times to death of patients who die while on study plus the sum of the patient study times for those patients surviving the duration of the study, divided by the number of patients who die on study. Since maximum likelihood estimators are invariant under one-to-one transformations

$$\hat{\lambda} = d\left[\sum_{i=1}^{n} (\delta_i t_i + (1-\delta_i) T_i)\right]^{-1}$$

is the MLE for λ.

If $d = 0$, we define

$$\hat{\mu} = \sum_{i=1}^{n} T_i$$

However, there will be few, if any, practical situations for which $d = 0$. Bartholomew (1957) obtained an approximation for the variance of $\hat{\mu}$ which is given as

$$V(\hat{\mu}) = \frac{\hat{\mu}^2}{\sum_{i=1}^{n} \hat{Q}_i}$$

where

$$\hat{Q}_i = 1 - \exp(-\lambda T_i) \qquad i = 1,2,\ldots,n$$

Example:

We consider the example in Bartholomew (1957), assuming that the data represent 10 survival times (in days) of patients with advanced lung cancer and that the study is terminated at a particular point in time.

TABLE 2.1 SURVIVAL TIMES OF PATIENTS WITH ADVANCED LUNG CANCER

Patient Number	1	2	3	4	5	6	7	8	9	10
Survival Time t_i) (Days)	2		51		33	27	14	24	4	
Number of Days Until End of Period T_i	81	72	70	60	41	31	31	30	29	21

From Table 2.1, we see that $\delta_2 = \delta_4 = \delta_{10} = 0$, and all other values of δ_i equal unity. Furthermore

$$d = 7, \quad \sum_{i=1}^{10} \delta_i t_i = 155 \text{ and } \sum_{i=1}^{10} (1-\delta_i) T_i = 153$$

Thus

$$\hat{\mu} = \frac{(155+153)}{7} = 44 \text{ days}$$

and

$$\hat{\lambda} = \frac{1}{44} = 0.023 \text{ death per day}$$

The approximate variance of $\hat{\mu}$ is

$$V(\hat{\mu}) \simeq \frac{\mu^2}{\sum_{i=1}^{n} Q_i} = \frac{44^2}{6.15} = 314.8$$

Again, if we assume that each patient has the same exponential death density function and all patients have the same point of entry into the study and also the study is terminated after the survival time of the $d^{\underline{th}}$ patient (out of $n \geq d$ patients in all) has been recorded; n is fixed, and d is assumed to be fixed. Thus $t_{(d)}$, the survival time of the $d^{\underline{th}}$ patient, is assumed to be a random variable. This case is considered by Halperin (1952) and Epstein and Sobel (1953).

The likelihood function $L(\underset{\sim}{\theta}')$ for the general k parameter case, where

$\underset{\sim}{\theta}' = (\theta_1, \ldots, \theta_k)$ is

$$L(\underset{\sim}{\theta}') = \frac{n!}{(n-d)!} \prod_{i=1}^{d} f(t_{(i)}; \underset{\sim}{\theta}')[S(t_{(d)}; \underset{\sim}{\theta}')]^{n-d}$$

where

$$S(t_{(d)};\theta') = \int_{t_{(d)}}^{\infty} f(t);\theta')dt .$$

For the exponential case, when $\lambda = \mu^{-1}$ is the parameter of interest, Halperin (1952) shows

$$L(\lambda) = \frac{n!}{(n-d)!}\lambda^d \exp\{-\lambda[\sum_{i=1}^{d-1} t_{(i)} + (n-d+1)t_{(d)}]\} .$$

The maximum likelihood estimator obtained by the standard procedure is

$$\hat{\lambda} = \frac{d}{\sum_{i=1}^{d-1} t_{(i)} + (n-d+1)t_{(d)}} = \frac{d}{y} \text{ (say)} . \quad (2.3.1.7)$$

Halperin also obtains the mean and variance of $\hat{\lambda}$, which are respectively,

$$E(\hat{\lambda}) = \frac{d\lambda}{d-1} \quad (2.3.1.8)$$

and

$$\text{Var } \hat{\lambda} = \frac{\lambda^2}{d-1} . \quad (2.3.1.9)$$

The above equations follow because for fixed d, the variable $2\lambda y$ has a χ^2-distribution with $2d$ degrees of freedom.

Epstein and Sobel consider μ, the mean time to death as the parameter of interest *vis-à-vis* the exponential survival distribution. They obtain $\hat{\mu}$, the maximum likelihood estimator of μ, which is $\hat{\lambda}^{-1}$ where $\hat{\lambda}$ is given by Equation (2.3.1.7). They then show that $\hat{\mu}$ is the minimum variance unbiased estimator of μ, using the fact that $2y/\mu$ has a χ^2-distribution with $2d$ degrees of freedom. It then follows that

$$\text{Var}(\hat{\mu}) = \frac{\mu^2}{d}$$

Finally, Epstein and Sobel obtain the expected value and variance of the time to the $d^{\underline{th}}$ death:

$$E(t_{(d)}) = \mu \sum_{j=1}^{d} [n-j+1]^{-1}$$

and

$$\text{Var}(t_{(d)}) = \mu^2 \sum_{j=1}^{d} [n-j+1]^{-2}$$

There are distributions for which a survival parameter's maximum likelihood cannot be obtained by the method described in Section 2.2. Suppose the death density of t involves only a location parameter θ; that is, if

$$f(t) = \exp[-(t-\theta)] \qquad t \geq \theta \qquad (2.3.1.10)$$

Equation (2.2.2) cannot be used to obtain $\hat{\theta}$. In this case, if

$t_{(1)} < t_{(2)} < \ldots < t_{(n)}$ are n ordered survival times whose death density is given by Equation (2.3.1.1), the maximum likelihood estimator $\hat{\theta} = t_{(1)}$. This says that if the survival times of n individuals are guaranteed to survive at least a period of length θ, then the ML estimator of θ is the minimum of the recorded survival times, that is

$$\hat{\theta} = \min_i t_i$$

When the parameters of a death density function are shape parameters, the ML estimates of these parameters and their large sample variance-covariance matrix are obtained through methods discussed in Section 2.2.

2.3.2 The Weibull Model

In this model the hazard function depends on time and is given by

$$\lambda(t) = \lambda p(t-\xi)^{p-1} \qquad (2.3.2.1)$$

where

λ, p, ξ are non-negative.

The Weibull hazard function has been often found to yield an excellent fit to survival data. We usually assume that ξ, the location parameter, is known or is zero. When ξ is known,

we can make a transformation $t' = t - \xi$. Since ξ must be less than $\min_i t_i$, if one has small values of t_i, taking $\xi = 0$ would be a sensible choice.

If we drop the prime notation, the Weibull hazard function, $\lambda(t)$ could be written as

$$\lambda(t) = \lambda p t^{p-1} \qquad (2.3.2.2)$$

Plotting this hazard function against timt t (See Figure 2.2) we observe that this hazard is monotone decreasing for $p < 1$, increasing for $p > 1$ and reduces to the constant exponential hazard if $p = 1$. The density function and the survivorship function corresponding to this hazard function are, respectively:

$$f(t) = \lambda p t^{p-1} \exp(-\lambda t^p) \quad t \geq 0, p>0, \lambda>0 \qquad (2.3.2.3)$$

and

$$S(t) = \exp(-\lambda t^p) \qquad (2.3.2.4)$$

Since it is often difficult to distinguish between the Weibull and gamma density function, both being generalizations of the exponential density function, in practice it is usually helpful if the researcher could pinpoint the method in which his data are obtained. Because of its versatility in fitting time-to-failure distributions of a rather extensive variety of complex mechanisms, the Weibull distribution has, in recent

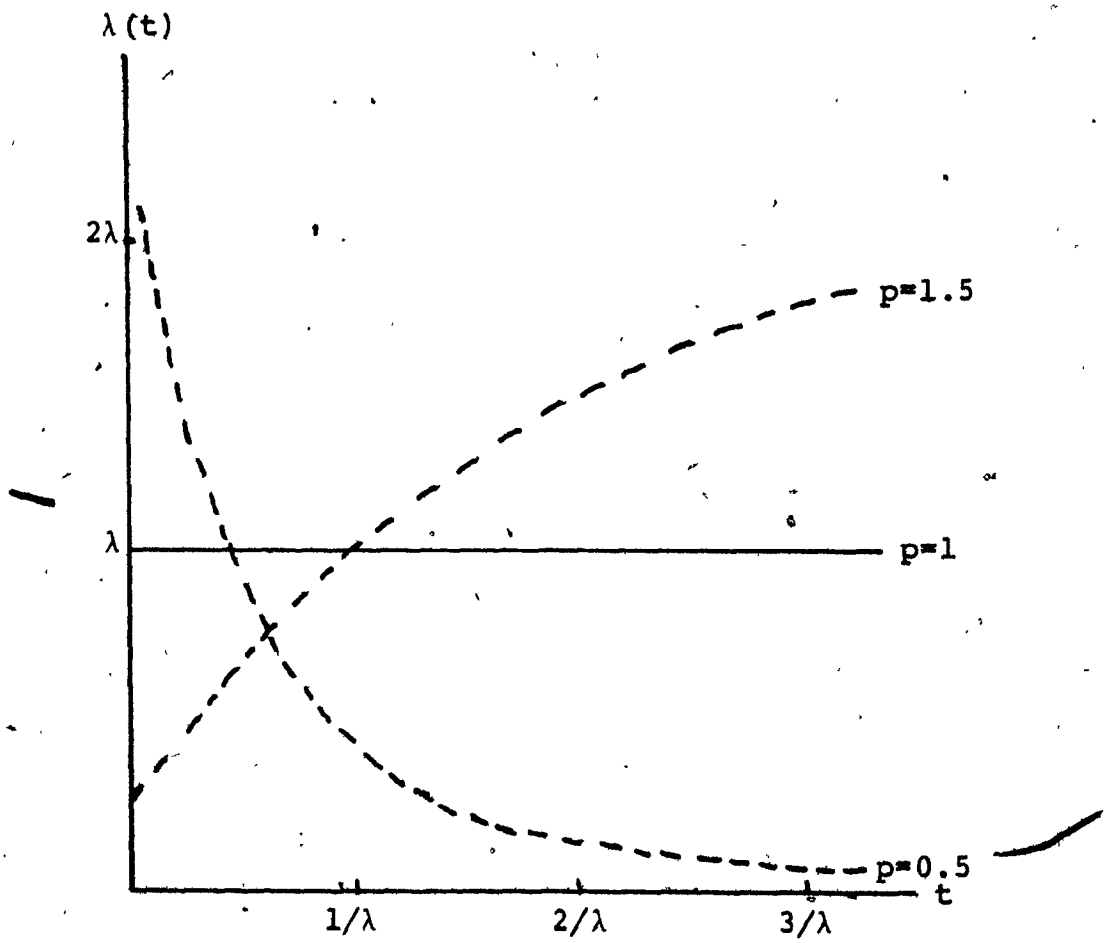

FIG. 2.2 HAZARD FUNCTIONS FOR THE TWO PARAMETER WEIBULL MODEL

years, assumed a position of importance in the field of reliability and life-testing. Various problems associated with this distribution have been considered by Dubey (1963), Menon (1963), Esary and Proschan (1963), Kao (1959), Lehman (1963) and Proschan (1963). A major deterrent to wider usage of the Weibull distribution has been the difficulty in estimating its parameters. Unfortunately, the calculations involved are not always simple.

Consider the case of a complete sample and define the Weibull density function as

$$f(x) = (p/\theta)x^{p-1}\exp(-x^p/\theta), \quad x\geq 0, p>0, \theta>0 \qquad (2.3.2.5)$$

This particular form in which Equation (2.3.2.5) is written was chosen for the purpose of simplifying derivations of the maximum likelihood estimating equations.

Consider a random sample consisting of n observations when Equation (2.3.2.5) is the applicable density function. The likelihood function of this sample is

$$L(x_1,\ldots,x_n;p,\theta) = \prod_{i=1}^{n}(p/\theta)x_i^{p-1}\exp(-x_i^p/\theta) \qquad (2.3.2.6)$$

The maximum likelihood equations are given by

$$\left.\begin{aligned} \frac{\partial \ell n L}{\partial p} &= \frac{n}{p} + \sum_{i=1}^{n} \ell n\, x_i - \frac{1}{\theta} \sum_{i=1}^{n} x_i^p \ell n\, x_i = 0 \\ \frac{\partial \ell n L}{\partial \theta} &= -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{1}^{n} x_i^p = 0 \end{aligned}\right\} \qquad (2.3.2.7)$$

By eliminating θ between these two equations and simplifying we have

$$\left[\frac{\sum_{1}^{n} x_i^p \ell n\, x_i}{\sum_{1}^{n} x_i^p} - \frac{1}{p}\right] = \frac{1}{n} \sum_{1}^{n} \ell n\, x_i \qquad (2.3.2.8)$$

from which MLE of p can be obtained with the aid of standard iterative procedures discussed in Section 2.2. In most cases, a simple trial-and-error approach will suffice. Once two values p_1 and p_2 have been found within a suffiently narrow interval such that $p_1 < \hat{p} < p_2$, linear interpolation will yield the required solution. With $\hat{p}$ thus determined, θ is estimated from the second equation of Equation (2.3.2.7) as

$$\hat{\theta} = \sum_{1}^{n} \frac{x_i^{\hat{p}}}{n} \qquad (2.3.2.9)$$

Anti-neoplastic drugs are of importance to the cancer chemotherapist as treatment protocols increasingly call for the administration of multiple-drug combinations. Analysis of survival experiments involving animal models can provide valuable information on the potency of the drug. Consider N animal models placed under observation in a laboratory and as each

failure occurs, the time is noted. Finally, at some predetermined fixed time x_o, or after some predetermined fixed number of sample animal models fail, the test is terminated. In both of these cases the data collected consist of observations $x_1, x_2, \ldots, x_n$ plus the information that (N-n) animal models survived beyond the time of termination x_o in the former case, and x_n in the latter.) We use the standard terminology as employed by Cohen (1959); when x_o is fixed and thus n is a random variable, censoring is said to be of Type I; when n is fixed and the time of termination x_n is a random variable, censoring is said to be of Type II.

In both Type I and Type II censoring, the likelihood function may be written as

$$L = \frac{N!}{(N-n)!} \left[\prod_{i=1}^{n} (p/\theta) x_i^{p-1} \exp(-x_i^p/\theta)\right][1 - F(x_T)]^{N-n} \quad (2.3.2.10)$$

where for Type I censoring, the time of termination $x_T = x_o$, and in Type II, censoring $x_T = x_n$. The distribution function F(x) from Equation (2.3.2.5) is given by

$$F(x) = \int_0^x pt^{p-1} \exp(-t^p/\theta)\, dt/\theta$$

$$= 1 - \exp(-x^p/\theta). \quad (2.3.2.11)$$

We obtain the estimating equations as

$$\left.\begin{aligned} \frac{\partial \ell n L}{\partial p} &= \frac{n}{p} + \sum_{I}^{n} \ell n\, x_i - \frac{1}{\theta} \Sigma^{*}\, x_i^p\, \ell n\, x_i = 0 \\ \frac{\partial \ell n L}{\partial \theta} &= -\frac{n}{\theta} + \frac{1}{\theta^2} \Sigma^{*}\, x_i^p = 0 \end{aligned}\right\} \qquad (2.3.2.12)$$

where

Σ^{*} signifies that the summation extends over the entire sample with the $(N-n)$ survivors assigned the value x_T; that is, x_o or x_n.

In particular, Cohen (1965) noted that

$$\left.\begin{aligned} \Sigma^{*}\, x_i^p\, \ell n\, x_i &= \sum_{1}^{n} x_i^p\, \ell n\, x_i + (N-n)\, x_T^p\, \ell n\, x_T \\ \Sigma^{*}\, x_i^p &= \sum_{1}^{n} x_i^p + (N-n)\, x_T^p \end{aligned}\right\} \qquad (2.3.2.13)$$

In the above form, estimating Equations (2.3.2.12) are fully analogous to Equations (2.3.2.7) for complete samples, and on eliminating θ between the two Equations of (2.8.2.12), Cohen showed that

$$\left[\frac{\Sigma^{*}\, x_i^p\ n\, x_i}{\Sigma^{*}\, x_i^p} - \frac{1}{p}\right] = \frac{1}{n} \sum_{1}^{n} \ell n\, x_i \qquad (2.3.2.14)$$

We solve for $\hat{p}$ employing the same techniques suggested for use in solving Equation (2.3.2.8) in the case of a complete sample. With $\hat{p}$ thus determined, it then follows from the second equation of (2.3.2.12) that

$$\hat{\theta} = \Sigma^* \frac{x_i^p}{n} \qquad (2.3.2.15)$$

In many life-testing situations, the initial censoring results in withdrawal of only a portion of the survivors, with some remaining on test and therefore continuing under observation until ultimate failure or until a subsequent stage of censoring is performed. For sufficiently large samples, censoring may be progressive through several stages. Suppose that censoring occurs progressively in k stages at times T_i where $T_i > T_{i-1}$, $i = 1,2,\ldots,k$, and that at the $i^{\underline{th}}$ stage of censoring r_i sample specimens selected randomly from the survivors at time T_i are removed (censored) from further observation. If we let N designate the total sample size as in the case of singly censored observation, and n the number of animal models which fail and therefore provide completely determined life spans, it follows that

$$N = n + \sum_{1}^{k} r_i \qquad (2.3.2.16)$$

In Type I progressive censoring where the T_i are fixed, the likelihood function may be written as

$$L = C \prod_{i=1}^{n} f(x_i) \prod_{i=1}^{k} [1 - F(T_i)]^{r_i} \qquad (2.3.2.17)$$

where

C is a constant

$f(x)$ is the density function, and

$F(x)$ is the distribution function.

With $f(x)$ given by Equation (2.3.2.5) and $F(x)$ by Equation (2.3.2.11), the logarithm of the likelihood function becomes

$$\ell nL = n\ell n\ p - n\ell n\theta + (p-1)\sum_{1}^{n} \ell n\ x_i - (\frac{1}{\theta})\sum_{1}^{n} x_i^p - (\frac{1}{\theta})\sum_{1}^{k} r_i\ T_i^p + \ell n\ C \qquad (2.3.2.18)$$

from which estimating equations are obtained as

$$\left.\begin{aligned} \frac{\partial \ell n\ L}{\partial p} &= \frac{n}{p} + \sum_{1}^{n} \ell n\ x_i - \frac{1}{\theta}\,\Sigma^{**}\ x_i^p\ \ell n\ x_i = 0 \\ \frac{\partial \ell n\ L}{\partial p} &= -\frac{n}{\theta} + \frac{1}{\theta^2}\,\Sigma^{**}\ x_i^p = 0 \end{aligned}\right\} \qquad (2.3.2.19)$$

where

Σ^{**} signifies summation over the entire sample with the r_i observations censored at time T_i assigned the value $x_i = T_i$.

More specifically, Cohen noted that

$$\left.\begin{aligned} \Sigma^{**}\ x_i^p\ \ell n\ x_i &= \sum_{1}^{n} x_i^p\ \ell n\ x_i + \sum_{1}^{k} r_i\ T_i^p\ \ell n\ T_i \\ \Sigma^{**}\ x_i^p &= \sum_{1}^{n} x_i^p + \sum_{1}^{k} r_i\ T_i^p \end{aligned}\right\} \qquad (2.3.2.20)$$

In the above form, estimating Equations (2.3.2.19) are fully analogous with Equations (2.3.2.7) for complete samples and with Equations (2.3.2.12) for singly censored samples. Equations (2.3.2.7) and (2.3.2.12) may be considered as special cases of Equations (2.3.2.19). Accordingly on eliminating θ between the two Equations of (2.3.2.19) we have

$$\left[\frac{\Sigma^{**} x_i^p \ln x_i}{\Sigma^{**} x_i^p} - \frac{1}{p}\right] = \frac{1}{h}\sum_1^n \ln x_i \qquad (2.3.2.21)$$

to be solved for $\hat{p}$ in the same manner as previously suggested for the complete and the single censored sample. With $\hat{p}$ thus determined, it follows from the second Equation of (2.3.2.19) that

$$\hat{\theta} = \Sigma^{**} \frac{x_i^p}{n} \qquad (2.3.2.22)$$

As noted by Cohen (1963) intermediate steps in the derivation of estimating equations for Type II progressively censored samples differ from corresponding steps in the case of Type I censoring. The end result, however, is the same in both cases, and the estimating equations given here are applicable for both sample types. We should note that the times T_i are the times at which withdrawals are actually made.

The asymptotic variance-covariance matrix of $(\hat{p}, \hat{\theta})$ is obtained by using the formulae given in Section 2.2. In the present situation, it seems appropriate to approximate the

expected values by their ML estimates. Accordingly, we have as the appropriate variance-covariance matrix

$$\begin{bmatrix} -\frac{\partial^2 \ell nL}{\partial p^2}\Big|_{\hat{p},\hat{\theta}} & -\frac{\partial^2 \ell nL}{\partial p \partial \theta}\Big|_{\hat{p},\hat{\theta}} \\ -\frac{\partial^2 \ell nL}{\partial \theta \partial p}\Big|_{\hat{p},\hat{\theta}} & -\frac{\partial^2 \ell nL}{\partial \theta 2}\Big|_{\hat{p},\hat{\theta}} \end{bmatrix}^{-1} \approx \begin{bmatrix} V(\hat{p}) & cov(\hat{p},\hat{\theta}) \\ cov(\hat{p},\hat{\theta}) & V(\hat{\theta}) \end{bmatrix}$$

The elements of the information matrix on the left-hand side of the above equation are obtained by differentiating Equation (2.3.2.7) for complete samples, Equation (2.3.2.12) for singly censored samples, and Equation (2.3.2.19) for progressively censored samples as follows:

For complete samples:

$$\left.\begin{aligned} -\frac{\partial^2 \ell nL}{\partial p^2}\Big|_{\hat{p},\hat{\theta}} &= \frac{n}{\hat{p}^2} + \frac{1}{\hat{\theta}} \sum_1^n x_i^{\hat{p}} (\ell n\, x_i)^2 \\ -\frac{\partial^2 \ell nL}{\partial p \partial \theta}\Big|_{\hat{p},\hat{\theta}} &= -\frac{\partial^2 \ell nL}{\partial \theta \partial p}\Big|_{\hat{p},\hat{\theta}} = -\frac{1}{\hat{\theta}^2} \sum_1^n x_i^{\hat{p}} \ell n\, x_i \\ -\frac{\partial^2 \ell nL}{\partial \theta^2}\Big|_{\hat{p},\hat{\theta}} &= -\frac{n}{\hat{\theta}^2} + \frac{2}{\hat{\theta}^3} \sum_1^n x_i^{\hat{p}} \end{aligned}\right\} \quad (2.3.2.23)$$

For singly censored samples:

$$\left.\begin{aligned}
-\frac{\partial^2 \ell nL}{\partial p^2}\bigg|_{\hat{p},\hat{\theta}} &= \frac{n}{\hat{p}^2} + \frac{1}{\hat{\theta}} \Sigma^* \, x_i^{\hat{p}} \, (\ell n \, x_i)^2 \\
-\frac{\partial^2 \ell nL}{\partial p \partial \theta}\bigg|_{\hat{p},\hat{\theta}} &= -\frac{\partial^2 \ell nL}{\partial \theta \partial p}\bigg|_{\hat{p},\hat{\theta}} = -\frac{1}{\hat{\theta}^2} \Sigma^* \, x_i^{\hat{p}} \, \ell n \, x_i \\
-\frac{\partial^2 \ell nL}{\partial \theta^2}\bigg|_{\hat{p},\hat{\theta}} &= -\frac{n}{\hat{\theta}^2} + \frac{2}{\hat{\theta}^3} \Sigma^* \, x_i^{\hat{p}}
\end{aligned}\right\} \qquad (2.3.2.24)$$

For progressively censored samples:

$$\left.\begin{aligned}
-\frac{\partial^2 \ell nL}{\partial p^2}\bigg|_{\hat{p},\hat{\theta}} &= \frac{n}{\hat{p}^2} + \frac{1}{\hat{\theta}} \Sigma^{**} \, x_i^{\hat{p}} \, (\ell n \, x_i)^2 \\
-\frac{\partial^2 \ell nL}{\partial p \partial \theta}\bigg|_{\hat{p},\hat{\theta}} &= -\frac{\partial^2 \ell nL}{\partial \theta \partial p}\bigg|_{\hat{p},\hat{\theta}} = \frac{1}{\hat{\theta}^2} \Sigma^{**} \, x_i^{\hat{p}} \, \ell n \, x_i \\
\frac{\partial^2 \ell nL}{\partial \theta^2}\bigg|_{\hat{p},\hat{\theta}} &= -\frac{n}{\hat{\theta}^2} + \frac{2}{\hat{\theta}^3} \Sigma^{**} \, x_i^{\hat{p}}
\end{aligned}\right\} \qquad (2.3.2.25)$$

Although the foregoing results are valid in a strict sense only for large samples, they may be relied upon to provide reasonable approximations to estimate variances and covariances for moderate size samples. In small samples, it must be recognized that errors due to bias sometimes greatly exceed the errors induced by large estimates of variances. This is an area which requires further investigation with respect to the Weibull distribution. Some linear combination of order statistics providing

unbiased estimates may be preferred over the estimates considered here. At least the maximum likelihood estimates are consistent and hence we are assured that the bias diminishes as the sample size becomes large.

Cohen (1965) has given as an illustrative example of the application of these results for n = 20 from a data given by Menon (1963). The sample comes from a population in which p = 0.5 and $\theta = \sqrt{e} \doteq 1.649$. For comparison, the maximum likelihood estimates are listed below with corresponding estimates based on Menon's results, along with the population values and moment estimates. The moment estimates of θ was calculated using

$$\mu_k' = \theta^{k/p}\, \Gamma[(k/p) + 1]$$

where

μ_k' is the k^{th} non-central moment which readily follows from Equation (2.3.2.5).

TABLE 2.2 COMPARISON OF DIFFERENT ESTIMATES USING MENON'S DATA

	POPULATION VALUES	MOMENT ESTIMATES	MENON'S ESTIMATES	ML ESTIMATES
θ	1.649	1.230	1.400	1.363
p	0.500	0.430	0.570	0.506

The Variance-Covariance matrix is

$$\begin{bmatrix} 0.007 & 0.014 \\ 0.014 & 0.123 \end{bmatrix}.$$

Other estimators of θ and p have been studied by Menon (1963), Miller and Freund (1965), and Gumbel (1958).

Bain and Antle (1967) performed Monte Carlo simulations to compare the biases and the variance among four of the five estimators that are potential substitutes for the maximum likelihood estimators of the Weibull parameters. These estimators which are compared in samples of sizes 5, 10, 20, 25, and 30 are the Bain-Antle choice 1 estimators, the Menon estimators, the Miller and Freund estimators, and the Gumbel estimators. The conclusion of the simulation was that all these estimators are quite good and the differences among them are small. Menon's estimators improved as the sample size increased, however, they cannot be used in the presence of censoring.

2.3.3 The Lognormal Model

The hazard function for this model is given by

$$\lambda(t) = \frac{1}{\sqrt{(2\Pi\lambda_o)t}} \exp\{-\frac{[\log(\lambda_1 t)]^2}{2\lambda_o}\} \frac{1}{1 - \Phi[\frac{\log(t\lambda_1)}{\sqrt{\lambda_o}}]} \quad (2.3.3.1)$$

where

$$\Phi(y) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{y} \exp(-\frac{x^2}{2})\, dx \quad (2.3.3.2)$$

The probability density function is

$$f(t) = \frac{1}{\sqrt{(2\Pi\lambda_o)t}} \exp\{-\frac{[\log(\lambda_1 t)]^2}{2\lambda_o}\} \quad (2.3.3.3)$$

and the cumulative distribution function is

$$F(t) = \frac{1}{\sqrt{(2\Pi\lambda_o)}} \int_{o}^{t} \exp\{-\frac{[\log(\lambda_1 y)]^2}{2\lambda_o}\} \frac{dy}{y} \quad (2.3.3.4)$$

By writing $x = \frac{[\log(\lambda_1 y)]}{\sqrt{\lambda_o}}$ one can identify $F(t)$ as

$$\Phi[\frac{\log t + \log \lambda_1}{\sqrt{\lambda_o}}].$$

This shows that the random variable log T is normally distributed with mean $-\log \lambda_1$ and variance λ_o.

Boag (1949) gives an extensive application of the lognormal survival distribution to patients with cancer. He compares fitting patient survival to both the lognormal and exponential survival distributions. As a practical example of the use of the lognormal survival distribution, length of times to recovery from injuries or surgery often follows a lognormal distribution. It can be shown that the hazard function increases initially to a maximum value and then decreases to zero as t approaches infinity; that is, there is an early period of positive aging followed by negative aging. A common choice of survival distribution among research workers in the life sciences is the lognormal. Since many probability density functions of survival time are skewed to the right, it is natural to take logarithms to obtain a more symmetrical distribution. Gehan (1969) showed that a lognormal distribution is difficult to distinguish from an exponential distribution. Hence it is suggested that if a distribution is to be chosen empirically and the hazard function is nearly constant, the exponential distribution is a reasonable choice because of its simplicity.

Feinleib (1960) has also proposed a method of analyzing lognormally distributed survival data with incomplete follow-up. Lea (1945), Osgood (1952, 1958) have observed that the distributions of duration of survival in several diseases can be rather closely approximated by a lognormal curve. They have pointed out that although the frequency distributions of

survival times in these diseases are markedly skewed to the right, the logarithms of the survival times are approximately normally distributed. Thus, if one were to plot the cumulative distribution of the survival times of patients with these diseases on lognormal graph using the methods discussed in Section 2.4, a straight line should be obtained. An inspection of the graphs presented by these authors often showed, however, that the observed distribution was markedly convex with respect to the predicted straight line. A similar disparity was observed in Feinleib et al (1960) of a study of survival patients suffering from chronic leukemia. (See Figure 2.3).

Gaddum (1945) has shown that such deviations can be corrected by subtracting an appropriate constant from the survival times. The transformation results in a three-parameter or translated lognormal distribution. That is, the variable t defined by

$$t = \frac{1}{c} \log \left(\frac{x-\alpha}{b}\right) \qquad (2.3.3.5)$$

is a standard normal variate with mean zero and variance one, where

x is the duration of survival measured from the date of diagnoses, onset of symptoms, or some other point in the course of the disease,

α is the appropriate constant, and

logs b and c are the mean and standard deviation of the variable log $(x-\alpha)$, respectively.

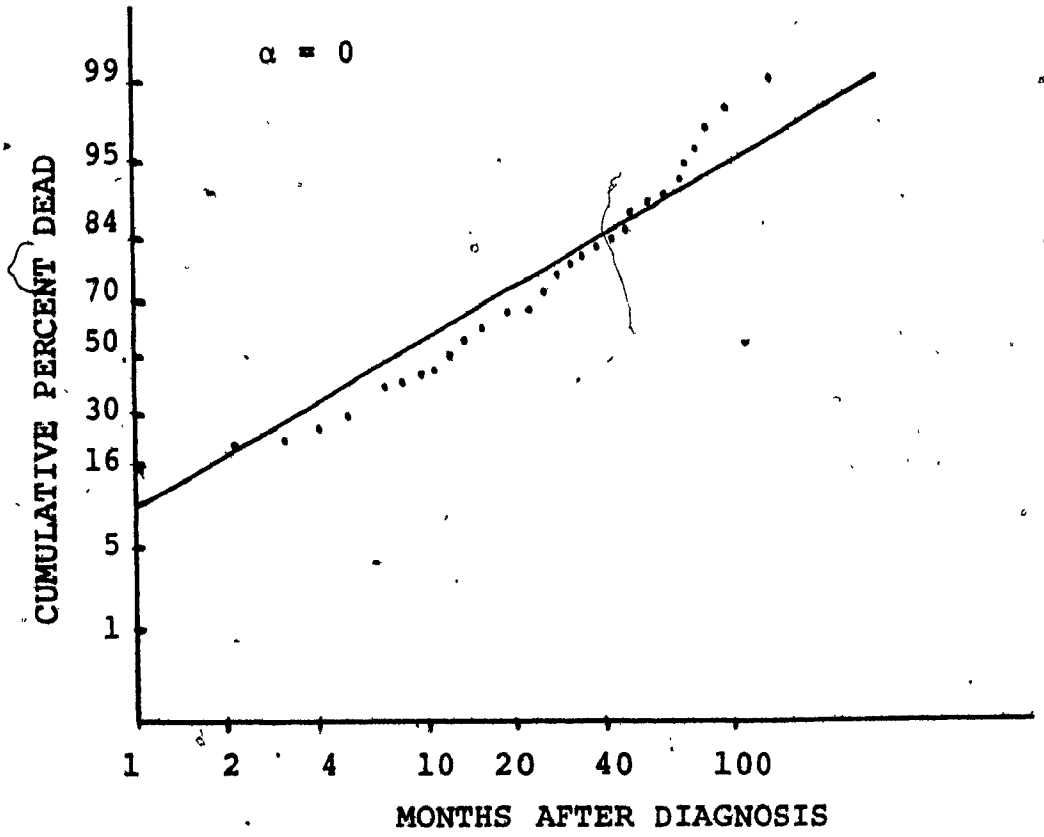

FIG. 2.3 DURATION OF SURVIVAL FROM DATE OF DIAGNOSIS OF 234 PATIENTS WITH CHRONIC LYMPHOCYTIC LEUKEMIA DURATION OF SURVIVAL IS PLOTTED ALONG THE LOGARITHMIC SCALE

To-date, no method has been proposed for fitting a translated lognormal distribution to incomplete data, although maximum likelihood solutions for complete data have been provided by Cohen (1951). Feinleib (1960), gave modifications of Cohen's equations as applied to life data. However, it has not yet been ascertained whether the values obtained by this modified maximum likelihood procedure are in complete agreement with the exact maximum likelihood solution for the case of incomplete data. He noted that if the survival data are complete, the procedure proposed is strictly equivalent to the maximum likelihood solution for grouped data.

Under the conditions that $\log(x-\alpha)$ is normally distributed, the density function of x would be given by

$$f(x) = \begin{cases} \dfrac{1}{c(x-\alpha)\sqrt{2\pi}} \exp\left[-\dfrac{1}{2c^2}\log^2\left(\dfrac{x-\alpha}{b}\right)\right] & \text{if } x > \alpha \\ 0 & \text{if } x \leq \alpha \end{cases} \qquad (2.3.3.6)$$

where

$$c > 0, \; -\infty < \alpha < \infty, \; b > 0$$

If the duration of each patient were known, the parameters, of this distribution could be obtained from the ML estimators described by Cohen (1951). He found out that the maximum likelihood estimates of $\log b$ and c^2 are

$$\log \hat{b} = \frac{1}{n} \sum_{i=1}^{n} \log(x_i - \hat{\alpha}) \qquad (2.3.3.7)$$

and

$$\hat{c}^2 = \frac{1}{n} \sum_{i=1}^{n} \log^2(x_i - \hat{\alpha}) - \left[\frac{1}{n} \sum_{i=1}^{n} \log(x_i - \hat{\alpha})\right]^2 \qquad (2.3.3.8)$$

while the MLE of α is the solution, $\hat{\alpha}$, of the equation

$$\lambda(\alpha) = \left[\sum_{i=1}^{n} \frac{1}{x_i - \alpha}\right]\left[n \sum_{i=1}^{n} \log(x_i - \alpha) - n \sum_{i=1}^{n} \log^2(x_i - \alpha) + \left(\sum_{i=1}^{n} \log(x_i - \alpha)\right)^2\right] - n^2 \sum_{i=1}^{n} \frac{\log(x_i - \alpha)}{x_i - \alpha} = 0 \qquad (2.3.3.9)$$

where

n is the number of patients observed and x_i is the duration of survival of the i^{th} patient. Equation (2.3.3.9) may be solved for $\hat{\alpha}$ by iterative procedure discussed in Section 2.2 and the solution subsequently substituted in Equations (2.3.3.7) and (2.3.3.8) to obtain the estimates $\log \hat{b}$ and $\hat{c}^2$.

Adjustments for these estimates were made by Feinleib (1960) for incomplete follow-up survival times. In human survival studies, the duration of survival is generally not known for each individual. Some patients may still be alive at the termination of the study, while others may withdraw from the study prior to their deaths and be lost to follow-up.

For such patients the value of x_i would be unknown.

One method of making allowances for these patients is the actuarial or life table method. The life table method is based on the calculation of conditional probabilities of dying during short intervals of the period under consideration. A commonly used expression for calculating the probabilities of dying during any interval, is known as the "½-rule" and is given by the following equation[2]

$$q_i = \frac{D_i}{L - \frac{1}{2} W_i} \qquad (2.3.3.10)$$

where

y_i = the beginning of the $i^{\underline{th}}$ interval;

D_i = the number of individuals known to have died during the $i^{\underline{th}}$ interval;

L = number of individuals alive and under observation at the beginning of the $i^{\underline{th}}$ interval;

W_i = number of individuals withdrawing during the $i^{\underline{th}}$ interval;

---

[2]The usual life table notation has been slightly modified to allow for summation over the index for intervals rather than directly over time. If the number of intervals in the life table is p, the initial point of the time scale, y_1, is set equal to the (unknown) parameter α, y_1 is chosen to be positive and larger than α, and the upper end-point of the last interval y_{p+1}, is set equal to plus infinity, while the remaining p-2 points are placed at any convenient distance between y_1 and y_{p+1}. Since the value of α is not known at the outset of the calculations, y_1 may be taken to be the origin from which survival is measured and set equal to zero.

q_i = the conditional probability of dying during the $i^{\underline{th}}$ interval, given survival to y_i.

Having calculated the values of q_i for each interval, these are then applied to a hypothetical number of individuals, ℓ_1, alive at the beginning of the first interval, e.g.,1,000, to yield ℓ_i, the number of individuals of the original 1,000 who are alive at the beginning of the $i^{\underline{th}}$ interval, and d_i, the number of deaths that occur during the $i^{\underline{th}}$ interval, for i ranging from 1 to p. The values of the ℓ_i and the d_i may be obtained from the following relations:

$$d_i = q_i \ell_i \qquad (2.3.3.11)$$

$$\ell_{i+1} = \ell_i - d_i \qquad (2.3.3.12)$$

The adjusted survival curve of a group of patients may next be obtained by plotting the proportion who have died as a function of time. This is done by plotting the cumulative sum of d_i/ℓ_i in the life table against the end point of the last interval in each sum.

An intuitive approach toward estimating the parameters of the lognormal would be to treat the numbers dying in such interval as calculated in the life table, the d_i, as if they had arisen from an ordinary random sample. This was the method adopted by Feinleib et al (1960) in the study on chronic leukemia.

Before substituting the values obtained from the life table into the usual maximum likelihood equations, certain adjustments were made for the grouping.

When using grouped data, the assumption is usually made that all the values within any small interval can be considered concentrated at the mean of the portion of the distribution over that interval. That is, in terms of the transformed variable, t, it is desired to find the point t_i^* for each interval, such that

$$t_i^* \int_{t_i}^{t_{i+1}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \int_{t_i}^{t_{i+1}} \frac{1}{\sqrt{2\pi}}\, t\, e^{-t^2/2}\, dt \qquad (2.3.3.13)$$

where

$$t_i = \frac{1}{c} \log\left(\frac{y_i-\alpha}{b}\right) \qquad (2.3.3.14)$$

For all intervals but the first and the last, the usual assumption may be made that the desired point is the mid-point of the transformed interval, i.e.,

$$t_i^* = \tfrac{1}{2}(t_i+t_{i+1}) \quad i=2,3,\ldots,p-1 \qquad (2.3.3.15)$$

The point in each interval on the original time-scale which is mapped into the t_i^* is given by the inverse relationship.

$$t_i^* = \frac{1}{c} \log\left(\frac{y_i^*-\alpha}{b}\right) \qquad (2.3.3.16)$$

Substituting Equation (2.3.3.16) into Equation (2.3.3.15), the following simplification is obtained:

$$\log(y_i^*-\alpha) = \tfrac{1}{2}[\log(y_i-\alpha) + \log(y_{i+1}-\alpha)], \quad i\ 2,3,\ldots,p-1 \tag{2.3.3.17}$$

For the first and last intervals equations (2.3.3.15) and (2.3.3.17) do not hold since $y_1 = \alpha$, $y_{p+1} = \infty$; $t_1 = -\infty$ and $t_{p+1} = \infty$. Estimates of the means of these intervals can be made from the data in the life table by making the assumptions that

$$\frac{d_1}{\ell_1} = \int_{-\infty}^{t_2} \frac{1}{\sqrt{2\Pi}} e^{-t^2/2}\, dt \tag{2.3.3.18}$$

and

$$\frac{d_p}{\ell_1} = \int_{t_p}^{\infty} \frac{1}{\sqrt{2\Pi}} e^{-t^2/2}\, dt \tag{2.3.3.19}$$

Equation (2.3.3.13) can now be solved explicitly to yield

$$t_1^* = \frac{\ell_1}{d_1}\left[-\frac{e^{-t_2^2/2}}{\sqrt{2\Pi}}\right] \tag{2.3.3.20}$$

and

$$t_p^* = \frac{\ell_1}{d_p}\left[\frac{e^{-t_p^2/2}}{\sqrt{2\Pi}}\right] \tag{2.3.3.21}$$

where

t_2 and t_p are determined from Equations (2.3.3.18) and (2.3.3.19), respectively.

We observe that the values of y_i^* $(i=2,3,\ldots,p-1)$ as determined by Equation (2.3.3.17) are independent of b and c, and depend only on α. The values of y_1^* and y_p^*, obtained by substituting the values of t_1^* and t_p^* found from equations (2.3.3.20) and (2.3.3.21) into equation (2.3.3.16) depend on all three parameters.

A method of determining y_1^* and y_p^* which is independent of the parameter b is provided by the relations

$$\log(y_1^*-\alpha) = \log(y_2-\alpha) - c(t_2-t_1^*) \qquad (2.3.3.22)$$

and

$$\log(y_p^*-\alpha) = \log(y_p-\alpha) + c(t_p^*-t_p) \qquad (2.3.3.23)$$

where

t_2, t_p, t_1^* and t_p^* are determined by equations (2.3.3.18) through (2.3.3.21), respectively.

The above relations provide convenient estimates for the y_i^* to be used in the following modifications of Cohen's maximum likelihood equations which are to be used in the estimation of the parameters of the lognormal distribution from the grouped data of the life table:

$$\log \hat{b} = \frac{1}{\ell_1} \sum_{i=1}^{p} d_i \log(y_i - \hat{\alpha})$$

$$\hat{c}^2 = \frac{1}{\ell_1} \sum_{i=1}^{p} d_i \log^2(y_i^* - \hat{\alpha}) - \left[\frac{1}{\ell_1} \sum_{i=1}^{p} d_i \log(y_i^* - \hat{\alpha})\right]^2$$

$$\lambda(\alpha) = \left[\sum_{i=1}^{p} \frac{d_i}{x_i^* - \alpha}\right]\left[\ell_1 \sum_{i=1}^{p} d_i \log(y_i^* - \alpha) - \ell_1 \sum_{i=1}^{p} d_i \log^2(y_i^* - \alpha) + \right.$$

$$\left. \left(\sum_{i=1}^{p} d_i \log(y_i^* - \alpha)\right)^2\right] - \ell_1^2 \sum_{i=1}^{p} \frac{d_i \log(y_i^* - \alpha)}{y_i^* - \alpha} = 0$$

Feinleib (1960) has also given a short-cut procedure for finding the initial estimates of the parameters α, b, and c.

2.3.3.1 Validity and Interpretation of the Log-Normal Distribution

The first problem to be considered is the validity of assuming that survival data follow a translated lognormal distribution. This problem can be resolved only in terms of empirical observations. Figure 2.4 shows that the subtraction of an appropriate constant may improve the approximation (compare Figures 2.3 and 2.4). Hence, empirically it seems valid to assume that the translated lognormal distribution is appropriate.

The second consideration is whether the use of a parameter, the constant to be subtracted, appreciably improves the goodness-of-fit. A comparison of Figures 2.3 and 2.4 in our

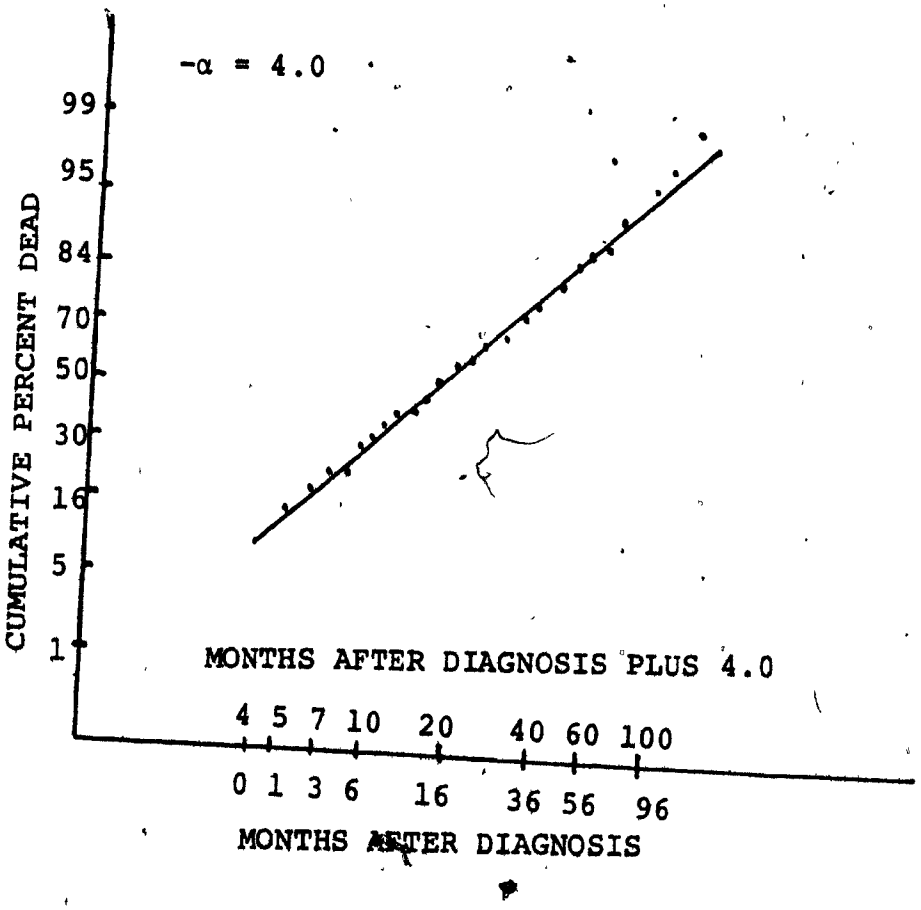

FIG. 2.4 DURATION OF SURVIVAL FROM DATE OF DIAGNOSIS OF 234 PATIENTS WITH CHRONIC LYMPHOCYTIC LEUKEMIA. DURATION OF SURVIVAL IN MONTHS PLUS 4.0 IS PLOTTED ALONG THE LOGARITHMIC SCALE

case shows that the improvement is appreciable. Thus, it seems justifiable in certain cases, to use the more general three-parameter distribution in preference to the two-parameter distribution. (i.e., where α is assumed *a priori* to be zero.)

The third problem is the interpretation to be given to the three parameters and the logarithmic transformation. Any venture at a biological interpretation of a mathematical phenomenon is hazardous. In general, however, a translated log-normal distribution might be interpreted as meaning that a multitude of factors affect the variable under consideration, that the effect of each factor depends upon the accumulative effect of the factors which have preceded it, and that the factors apparently began to act at the point α in the original time-scale. The point α, however, is not necessarily the time of onset of the disease, nor need it be any other clinically recognizable point in the natural history of the disease. It is strictly-speaking nothing more than a mathematical estimation of a parameter which enables us to fit a theoretical distribution to an empirical one. If one desires to interpret it biologically, it might be described as the apparent origin in time at which the cumulative factors which lead to death began to act. One should note that these interpretations may have no real basis in observable facts.

2.3.4 Other Models

2.3.4.1 The normal model

Use of the normal distribution as a model for the survival data is limited. One difficulty is that if the random variable t is assumed to follow a normal density function with mean μ and variance σ^2, then

$$f(t) = \frac{1}{\sigma\sqrt{2\Pi}} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] \qquad -\infty<\mu<\infty,\ -\infty<t<\infty,\ \sigma>0. \tag{2.3.4.1}$$

Theoretically both the observed time to death and the mean time to death could take on negative values; however, if $\mu > 0$ and $\frac{\mu}{\sigma} \geq 3$, it is virtually impossible for t to be negative. Nonetheless, the normal distributions do find applications in reliability theory when it is assumed that failure or death is due to accumulated wear. Thus, if K failures in total are required for a death, then for K large (as a consequence of the central limit theorem), the death density would approach normality. Or, if the amount of some substance needed by the body is normally distributed and failure occurs because it is totally consumed, we would expect a normal death density.

The survivorship function then is

$$S(t) = \frac{1}{\sigma\sqrt{2\Pi}} \int_t^{\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \tag{2.3.4.2}$$

The hazard rate, $\lambda(t)$ is

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\exp\{-\frac{(t-\mu)^2}{2\sigma^2}\}}{\int_t^\infty \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}dx} \qquad (2.3.4.3)$$

2.3.4.2 The Gompertz model

The Gompertz model is widely used in actuarial work. The hazard function is given by $\exp(\lambda_o+\lambda_1 t)$, which reduces to a constant when $\lambda_1 = 0$; hence the exponential distribution is a special case. If $\lambda_1 > 0 (< 0)$, there is positive (negative) aging starting from $\exp(\lambda_o)$. The survivorship function is

$$S(t) = \exp\{-\frac{\exp(\lambda_o)}{\lambda_1}[\exp(\lambda_1 t) - 1]\}$$

and the probability density function is easily written. The mean of the distribution is given by

$$E(T) = \exp(\lambda_o) G[\frac{\exp(\lambda_o)}{\lambda_1}]$$

where

$$G(x) = e^x \int_x^\infty y^{-1} e^{-y}\, dy$$

and is tabulated in Broadbent (1958).

2.3.4.3 The Gamma model

Another two parameter generalizations of the exponential model is the gamma distribution with density function

$$f(t) = \frac{\lambda^k t^{k-1} \exp(-\lambda t)}{\Gamma(k)}, \text{ where } k > 0, \lambda > 0 \quad (2.3.4.4)$$

Suppose $t_1, t_2, \ldots, t_n$ are independent survival times for n individuals, each of whom has a constant hazard rate λ.

Let

$$y = \sum_{i=1}^{n} t_i$$

That is, y is the total survival time for all n individuals. We can determine the death density function, hence the hazard rate and survival distribution of y, since the average time to death $\bar{t}$, is y/n.

Using mathematical induction if $n = 2$, the joint death density function for t_1 and t_2 is:

$$f(t_1,t_2) = \lambda^2 \exp[-\lambda(t_1+t_2)] \quad t_1 \geq 0, t_2 \geq 0 \quad (2.3.4.5)$$

Letting $y = t_1 + t_2$, we see that

$$f(t_1,y) = \lambda^2 \exp(-\lambda y) \quad 0 \leq t \leq y, y \geq 0 \quad (2.3.4.6)$$

Integrating out t_1, we have

$$f(y) = \int_0^y \lambda^2 \exp(-\lambda y)dt_1 = \lambda^2 y \exp(-\lambda y), y \geq 0 \quad (2.3.4.7)$$

Assume now for n = k, that

$$f(y_k) = \frac{\lambda^k y_k^{k-1} \exp(-\lambda y_k)}{(k-1)!}, \quad y \geq 0 \quad (2.3.4.8)$$

That is, $f(y_k)$ is a ssumed (under the induction hypothesis) death density function for $y_k = \sum_{i=1}^{k} t_i$. Equations (2.3.4.7) and (2.3.4.8) are special cases for the gamma density function whose general form is

$$f(t) = \frac{\lambda^\gamma t^{\gamma-1} \exp(-\lambda t)}{\Gamma(\gamma)}, t \geq 0 \quad (2.3.4.9)$$

where

$\lambda > 0$, and $\gamma > 0$ are the scale and shape parameters, respectively.

Now

$$f(y_k, t_{k+1}) = \frac{\lambda^{k+1} y_k^{k-1} \exp[-\lambda(y_k+t_{k+1})]}{(k-1)!}, y_k \geq 0, t_{k+1} \geq 0 \quad (2.3.4.10)$$

Letting $t = y_k + t_{k+1}$, we observe that

$$f(t_{k+1}, t) = \frac{\lambda^{k+1}(t-t_{k+1})^{k-1} \exp(-\lambda t)}{(k-1)!}, 0 \leq t_{k+1} \leq t, \ t \geq 0 \quad (2.3.4.11)$$

Integrating out t_{k+1}, we finally have

$$f(t) = \frac{\lambda^{k+1} t^{k} \exp(-\lambda t)}{k!}, \quad t \geq 0 \qquad (2.3.4.12)$$

Thus we have shown that if $t = t_1 + t_2 + \ldots + t_n$, where each time t_i has an exponential death density function $\lambda \exp(-\lambda t_i)$ and the times $t_1, t_2, \ldots, t_n$ are mutually independent, the death density function of t is given by

$$f(t) = \frac{\lambda^{n} t^{n-1} \exp(-\lambda t)}{(n-1)!}, \quad t \geq 0 \qquad (2.3.4.13)$$

which is also a special case of the gamma density function defined by Equation (2.3.4.9).

For the special case when $\gamma = n$, an integer, it is easy to show by successive integration by parts that $S(t)$, the gamma survival distribution, is given by

$$S(t) = \sum_{\nu=0}^{n-1} \frac{(\lambda t)^{\nu}}{\nu!} \exp(-\lambda t), \quad t \geq 0 \qquad (2.3.4.14)$$

Therefore, in this special case the hazard rate is given by

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda^{n} t^{n-1}/(n-1)!}{\sum_{\nu=0}^{n-1} (\lambda t)^{\nu}/\nu!} \qquad (2.3.4.15)$$

which is an increasing hazard rate.

An application of this hazard rate, Equation (2.3.4.15) to a real-life situation can be seen as follows - Suppose a researcher is studying a group of patients suffering from a

kidney disease in which the failure rate of each kidney is constant and equals λ. For a patient to die from the disease, both kidneys must fail. The hazard rate of these patients with this disease is found by substituting $n = 2$ in Equation (2.3.4.15). Thus

$$\lambda(t) = \frac{\lambda^2 t}{1 + \lambda t}$$

which is easily seen to be an increasing function of time for $\lambda > 0$.

Harter and Moore (1965), Wilk et al (1962), have both considered methods of estimating the parameters of the gamma density function for complete and censored samples. Actually for t continuous the gamma hazard function is

$$\lambda(t) = \frac{t^{\gamma-1} \exp(-\lambda t)}{\int_t^\infty x^{\gamma-1} \exp(-\lambda x)dx}, \quad t \geq 0$$

2.3.4.4 The generalized Pareto distribution

Another distribution which has recently been considered for progressively censored survival data is the generalized Pareto distribution. This three-parameter model discussed by Davis and Feldstein (1979) is general enough so as to include increasing, decreasing, and constant failure rate distributions. They indicated that the constant hazard rate, giving

rise to the exponential distribution may not be a realistic assumption when modelling chronic disease populations. Also, when the origin corresponds to an event which may dramatically influence survival, such as surgery, the objection to constant hazard rate may be especially strong.

In situations where constant hazard is not a reasonable assumption, monotone hazard usually is. The Weibull and most other parametric families used to model survival data have monotone hazard rates. Another family of monotone hazard rates considered by Davis and Feldstein is given by

$$\lambda(t) = \theta + \frac{\gamma}{(t+\phi)} \qquad (2.3.4.16)$$

with a survivorship function

$$S(t) = (1+t/\phi)^{-\gamma} e^{-\theta t} \qquad (2.3.4.17)$$

which they found to be the survivorship function for the Pareto Type III, $P_3(\theta,\gamma,\phi)$ distribution. The restriction on the parameters are

$$(\theta \geq 0,\ \phi > 0,\ \gamma \geq -\theta\phi) \qquad (2.3.4.18)$$

It is obvious from Equations (2.3.4.16) and (2.3.4.18) that the value of γ determines whether the hazard rate is increasing, $\gamma<0$, constant, $\gamma=0$ or decreasing, $\gamma>0$. In all cases the hazard rate begins at $\theta + \gamma/\phi$ and tends monotonically to θ. Examining Equation (2.3.4.18) we see that there are only two logical

boundary families, (i) $\theta = 0$ and (ii) $\gamma = -\theta\phi$. When $\theta = 0$, the P_3 distribution reduces to the Lomax or Pareto Type II, $P_2(\phi,\gamma)$, distribution which has its hazard rate and survivorship function respectively as

$$\lambda(t) = \gamma/(t+\phi) \qquad (2.3.4.19)$$

and

$$S(t) = (1 + \frac{t}{\phi})^{-\gamma} \qquad (2.3.4.20)$$

The case of $\gamma = -\theta\phi$ yields hazard rate and survivorship function respectively as

$$\lambda(t) = \frac{\theta t}{(t+\phi)} \qquad (2.3.4.21)$$

and

$$S(t) = (1 + \frac{t}{\phi})^{\theta\phi}\, e^{-\theta t} \qquad (2.3.4.22)$$

The P_3 family has the following closure property which is similar to that of Weibull random variables with a common shape parameter.

Theorem:

If an item can fail from any of k independent causes, the hazard rate for the i^{th} cause being $P_3(\theta_i,\gamma_i,\phi)$, and if failure results from the first occurrence of any one of the causes then time to failure is distributed as $P_3(\Sigma\theta_i,\Sigma\gamma_i,\phi)$.

The theorem follows because under independence the hazard rate associated with time to failure is the sum of the hazard rates of the k possible causes. Davis and Feldstein (1979) have dealt with the derivation, characterization and application of maximum likelihood methods to the P_3 family in the case of progressively censored data when ϕ is assumed known.

2.3.4.5 Piecewise survival model

There are other more complex types of distributions that can be considered. For example, we could have a J-shaped hazard rate characterized by a period of decreasing risk followed by a relatively constant failure rate. This corresponds to a high risk of death in early infancy, a fairly constant and low rate from childhood through middle age, and an increasing rate in old age.

Finally, we consider the death density function whose hazard rate is constant except for jump changes. That is,

$$\lambda(t) = \begin{cases} \lambda_1 & 0 \le t < t_1 \\ \lambda_2 & t_1 \le t < t_2 \\ \vdots & \\ \lambda_{k-1} & t_{k-2} \le t < t_{k-1} \\ \lambda_k & t \ge t_{k-1} \end{cases} \qquad (2.3.4.23)$$

where

$t_1, t_2, \ldots, t_{k-1}$ are known points of change of the parameter.

There is an application of Equation (2.3.4.23) to life tables, namely, the hazard rate of people is assumed to be constant in each time interval defined in the table. However, the overall death rate is not constant in life tables in general. Expressions for $S(t)$ and $f(t)$, are obtained as follows from $\lambda(t)$:

$$S(t) = \begin{cases} \exp(-\lambda_1 t) & 0 \leq t < t_1 \\ \exp[-\lambda_1 t_1 - \lambda_2 (t_2 - t_1)] & t_1 \leq t < t_2 \\ \vdots & \\ \exp[-\lambda_1 t_1 - \lambda_2 (t_2 - t_1) - \ldots - \lambda_k (t - t_{k-1})] & t \geq t_{k-1} \end{cases} \quad (2.3.4.24)$$

and

$$f(t) = \begin{cases} \lambda_1 \exp(-\lambda_1 t) & 0 \leq t < t_1 \\ \lambda_2 \exp[-\lambda_1 t_1 - \lambda_2 (t - t_1)] & t_1 \leq t < t_2 \\ \vdots & \\ \lambda_k \exp[-\lambda_1 t_1 - \lambda_2 (t_2 - t_1) - \ldots - \lambda_k (t - t_{k-1})] & t \geq t_{k-1} \end{cases} \quad (2.3.4.25)$$

This death density is referred to as the piecewise exponential death density function.

Plotted data sometimes indicate that it is not possible to fit a single distribution function of a simple known mathe-

matical form over the whole range of data. For example, cancer data often indicate rather high mortality rates in the first few years of diagnosis, but the patients who survive this critical period seem to experience subsequent mortality similar to that in the general population of the same age. Let $0<x_1<x_2 < \ldots < x_k<\infty$ be k fixed points at which changes of survival function might reasonably be expected. These points are often suggested by graphical displays, even if no additional information on the disease process is available. We denote the survival function in the successive intervals by $S_0(t)$, $S_1(t)$, ..., $S_k(t)$. Generally, $S_i(t)$ and $S_j(t)$, $(i\neq j)$ can even belong to different families; in practice, they are usually different members of the same family. For convenience, let

$$p_0(t|x_i) = \frac{S_i(t)}{S_i(x_i)}, \text{ for } x_i \leq t < x_{i+1}, i=0,1,\ldots,k,$$

denote the conditional probability ${}_{t-x_i}p_{x_i}$ of surviving until time (age)t, given alive at time x_i, (We denote $x_0=0$, and $S_0(0)=1$ and $x_{k+1}=\infty$.)

In particular, we have

$$p_0(x_1|0) = S_0(x_1) \text{ and } p_i(x_{i+1}|x_i) = \frac{S_i(x_{i+1})}{S_i(x_i)} \text{ for } i=1,2,\ldots,k-1$$

The unconditional probability of surviving until time t, where $x_i \leq t < t_{i+1}$ is

$$p_o(x_1|0)\cdot p_1(x_2|x_1)\cdots p_{i-1}(x_i|x_{i-1})p_i(t|x_i) =$$

$$= S_o(x_1)\cdot\frac{S_1(x_2)}{S_1(x_1)}\cdot\frac{S_2(x_3)}{S_2(x_2)}\cdots\frac{S_i(t)}{S_i(x_i)} \quad (2.3.4.26)$$

Therefore, the piecewise survival function $S(t)$ over k successive periods is

$$S(t) = \begin{cases} S_o(t) & \text{for } 0\le t<x_1 \\ S_o(x_1)\cdot\dfrac{S_1(t)}{S_1(x_1)} & \text{for } x_1\le t<x_2 \\ S_o(x_1)\cdot\dfrac{S_1(x_2)}{S_1(x_1)}\cdot\dfrac{S_2(t)}{S_2(x_2)} & \text{for } x_2\le t<x_3 \\ \vdots & \\ S_o(x_1)\cdot\dfrac{S_1(x_2)}{S_1(x_1)}\cdot\dfrac{S_2(x_3)}{S_2(x_2)}\cdots\dfrac{S_k(t)}{S_k(x_k)} & \text{for } x_k\le t \end{cases} \quad (2.3.4.27)$$

The corresponding death density function is

$$f(t) = \begin{cases} f_o(t) & \text{for } 0\le t<t_1 \\ S_o(x_1)\cdot\dfrac{f_1(t)}{S_1(x_1)} & \text{for } x_1\le t<x_2 \\ S_o(x_1)\cdot\dfrac{S_1(x_2)}{S_1(x_1)}\cdot\dfrac{f_2(t)}{S_2(x_2)} & \text{for } x_2\le t<x_3 \\ \vdots & \\ S_o(x_1)\cdot\dfrac{S_1(x_2)}{S_1(x_1)}\cdot\dfrac{S_2(x_3)}{S_2(x_2)}\cdots\dfrac{f_k(t)}{S_k(x_k)} & \text{for } x_k\le t \end{cases}$$

When the parametric form of $S_i(t)$ is specified, simple graphical methods discussed in Section 2.4 can be used to estimate parameters of each piece.

2.4 METHODS OF FITTING PARAMETRIC DISTRIBUTIONS

We have so far assumed that we know the parametric form in which the survival data follow; but suppose we don't know as often occurs in practice. There are two broad groups of methods for fitting distributions - graphical and analytical.

Among the analytical methods are the minimum Chi-square method and the least squares method which are regarded as mathematically sound in its application. However, their elegancy depends on the assumptions on which they are based. Many computer programs are available for fairly general use of these methods.

Most graphical methods rely on plotting some functions of the hazard rate function, or cumulative hazard function against some functions of t. The functions are usually so chosen that if the parametric form of the survival function is reasonably appropriate, an approximately straight-line plot will be obtained. The plotting can be facilitated by the use of hazard papers or probability papers. The relationship between the cumulative probability function $F(t)$ and the cumulative hazard $\Lambda(t)$ for a distribution can be written as

$$F(t) = 1 - \exp(-\Lambda(t)) \qquad (2.4.1)$$

This basic relationship can be seen on all hazard papers. Hazard plotting papers have been developed for five commonly used theoretical distributions: the exponential, normal, log-normal, extreme value, and Weibull. Nelson (1972) has discussed the parametric form of these distributions, and their hazard plotting paper. Especially useful in survival analysis are cumulative hazard papers on which $[t,\Lambda(t)]$ can be entered directly to give a straight line plot. We discuss this method for the case of the exponential distribution.

As pointed out earlier the hazard rate frequently presents the best means of identification. In the case of the exponential distribution, this is a constant. However, the effect of individual data values sometimes makes the plotted hazard rate so irregular that it is difficult to identify it. One technique for overcoming this problem is to look at the cumulation of the hazard rate as given by Nelson (1972). This is directly analogous to the use of normal probability paper (see Dixon and Massey (1967)) so scaled that the cumulative normal curve $F(x)$ is a straight line if the data are normally distributed. In the case of the exponential distribution, the cumulative hazard rate is simply

$$\Lambda(t) = \int_0^t \lambda dx = \lambda t \qquad t \geq 0 \qquad (2.4.2)$$

a linear function passing through zero. Thus the exponential hazard paper is merely regular graph paper with values of t on the vertical axis and $\Lambda(t)$ on the horizontal axis. We

have to resolve the problem as to how to choose the plotting position for $\Lambda(t)$. Since $\Lambda(t)$ is the theoretical cumulative hazard function, it must be estimated in applications. Suppose the times to failure are ordered $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(n)}$. It can be shown that

$$E(\Lambda(t_{(i)})) = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-i+1}$$

Thus the surrogate plotting position for $\Lambda(t_{(i)})$, which is unknown, is taken to be $E(\Lambda(t_{(i)}))$ which is known. We should note that when the data are censored, the plotting position, $E[\Lambda(t_{(i)})]$ may not be exact. Nelson points out, however, that the procedure is still satisfactory. The values $t_{(i)}$ are plotted on the vertical axis and the values $E[\Lambda(t_{(i)})]$ (which is the sum of the reverse ranks at the i^{th} order failure) are plotted on the horizontal axis. The general procedure for displaying the data are as follows:

(1) Order the n observations from smallest to largest, regardless of whether they are t_i or T_i (time to death or censored time, respectively) marking the T_i times with check marks.

(2) Calculate the hazard value for each time as $100/k$, where k is its reverse rank. The hazard value is the observed conditional probability of survival. Cumulate the hazard values downward. The cumulative

can be larger than 100%.

(3) Mark the vertical scale from 0 to the longest length of survival and the horizontal scale from 0 to the largest value of the cumulative hazard values.

(4) Plot each survival time vertically against its corresponding cumulative hazard value.

(5) By eye, fit a straight line through 0 and the points in the centre of the distribution.

If the points tend not to cluster around this straight line (particularly the central 80%), it is not reasonable to assume an exponential distribution. When the assumption of exponentiality is not satisfied, it is possible at times to transform the data, permitting an exponential distribution to be fitted to the transformed data. When a transformation is not possible another distribution should be considered. If the points fit the straight line, μ can be estimated by reading t from the vertical axis corresponding to 100%.

Occasionally individual times to death are not available to the researcher for analysis. Instead, only the number of deaths in each of a set of mutually exclusive time intervals can be obtained. The set of time intervals taken together covers the entire non-negative time axis. As an example, consider a group of 118 arthritic patients, each of

whom receives an analgesic to relieve his discomfort. Table 2.3 shows the number of patients receiving relief (in minutes) for 16 mutually exclusive time intervals. These data are a fictitious adaptation of data appearing in Table 3 of Scheuer (1968). We can fit an exponential death density function to this grouped data.

Suppose survival data on n individuals occur as in Table 2.3. More generally, we suppose the non-negative time axis to be divided into the mutually exclusive and exhaustive intervals $I_j: T_{j-1} \leq t < T_j$, $j=1,2,\ldots,k$, with $T_o = 0$ and $T_k = \infty$.

The data are the number of deaths observed in each interval: f_j deaths occuring in interval $I_j, j=1,2,\ldots,k$, $\sum_{j=1}^{k} f_j = n$. We wish to test the hypothesis that these data have an exponential death density function

$$f(t;\lambda) = \lambda\exp(-\lambda t) \qquad t \geq 0, \quad \lambda > 0$$

and to estimate λ.

When the hypothesis of exponentiality is true, the expected number of deaths in I_j is $np_j(\lambda)$, where $p_j(\lambda)$ is obtained by

$$p_j(\lambda) = \int_{T_{j-1}}^{T_j} \lambda\exp(-\lambda t)dt = \exp(-\lambda T_{j-1}) - \exp(-\lambda T_j) \quad (2.4.3)$$

TABLE 2.3 RELIEF TIMES OF ARTHRITIC PATIENTS RECEIVING AN ANALGESIC

INTERVAL (MINUTES)	FREQUENCY
$0 \leq t < 20$	19
$20 \leq t < 40$	19
$40 \leq t < 60$	21
$60 \leq t < 80$	10
$80 \leq t < 100$	13
$100 \leq t < 120$	6
$140 \leq t < 160$	7
$160 \leq t < 180$	5
$180 \leq t < 200$	4
$200 \leq t < 220$	2
$220 \leq t < 240$	3
$240 \leq t < 260$	1
$260 \leq t < 280$	2
$280 \leq t < 300$	1
$300 \leq t < 320$	1
$320 \leq t < 340$	2
$t \geq 340$	1
TOTAL	118

Clearly $\sum_{j=0}^{k} p_j(\lambda) = 1$. Thus to test Equation (2.4.3) for goodness of fit to the data we employ the χ^2 statistic

$$\chi^2 = \sum_{j=1}^{k} \frac{(f_j - np_j(\lambda))^2}{np_j(\lambda)} \qquad (2.4.4)$$

which is asymptotically distributed as a χ^2 variable with $k-1$ degrees of freedom when the hypothesis of exponentiality is true. If λ were known, the test described by Equation (2.4.4) would be appropriate. Under normal circumstances, λ is unknown and must be estimated from the data. Following Scheuer's development, we let χ^2 be defined as in Equation (2.4.4) and we try to find that value of λ, λ', (say) which minimizes χ^2 with respect to the data. This procedure we know as the χ^2-minimum procedure, and λ' is the minimum χ^2 estimator of λ. Formerly, then

$$\frac{\partial \chi^2}{\partial \lambda} = -\sum_{j=1}^{k} \left[\frac{2(f_j - np_j(\lambda))}{p_j(\lambda)} + \frac{(f_j - np_j(\lambda))^2}{np_j^2(\lambda)}\right] \frac{\partial p_j(\lambda)}{\partial \lambda}$$

where

$p_j(\lambda)$ is given by Equation (2.4.3), and

$$\frac{p_j(\lambda)}{\partial \lambda} = T_j \exp(-\lambda T_j) - T_{j-1}\exp(-\lambda T_{j-1}) \qquad j=1,2,\ldots,k$$

since $T_k = \infty$, $\exp(-\lambda T_k) = T_k \exp(-\lambda T_k) = 0$.

Thus λ', the minimum chi-square estimator of λ, is the solution to

$$\sum_{j=1}^{k} \left[\frac{(f_j - np_j(\lambda'))}{p_j(\lambda')} + \frac{(f_j - np_j(\lambda'))^2}{2np_j(\lambda')}\right]\frac{\partial p_j(\lambda')}{\partial \lambda'} = 0 \qquad (2.4.5)$$

which minimizes Equation (2.4.4). If we substitute λ' for λ in Equation (2.4.4), it can be shown that χ^2 will have an asymptotic χ^2 distribution with k-2 degrees of freedom, when the null hypothesis of exponentiality is true.

To find λ' from Equation (2.4.5), we employ the Newton-Raphson's method. Scheuer describes a computer program for the solution of Equation (2.4.5), in terms of λ'. The appropriate starting value λ'_o is

$$\lambda'_o = n\left[\frac{\sum_{j=1}^{k-1} f_j(T_{j-1}+T_j)}{2} + f_k T_{k-1}\right]^{-1}$$

This value is chosen because it is the analog to the maximum likelihood estimator $\hat{\lambda}$ for the grouped data situation. The values $\frac{(T_{j-1}+T_j)}{2}$ represent the midpoints of all the intervals except the last interval, for which we use the lower T_{k-1} because the upper limit of the last interval is infinite.

2.5 NON-PARAMETRIC ESTIMATION OF SURVIVAL CURVE

In this section, we assume that the theoretical form of the probability density function of survival time is unknown; but a sample of survival times is available and an estimate of the survivorship function is needed. Essentially, our task is to present methods appropriate for (i) small sample date, with each individual time of death recorded exactly, (ii) large sample data, with times of death usually grouped into fixed intervals. In this large sample data, we discuss problems associated with estimation of the survivorship function from a sample that represents the mortality experience of a cohort. The term "cohort" is used here in a rather special way. We define it as a sufficiently homogeneous group of individuals for which a certain initial event has already occurred. Furthermore, this group is under observation until the last number of the group fails - so we have complete records of each life subsequent to the initial event. We refer to such data as complete survival data. Thus we shall consider data which may be from a small or large and may or may not contain censored observations. A product-limit estimate of the survivor function derived by Kaplan and Meier (1958) is given when the sample size is small. A life table estimate is also given for large samples.

Both estimation procedures are valid regardless of whether censored data are present; and the life table procedure reduces to the product-limit estimate when the intervals are

made arbitrary small and include only one distinct time to failure. The usual life table estimate has been given by many others, including Berkson and Gage (1950) and Gehan (1969).

2.5.1 Estimation of Survivorship Function in Small Samples - The Product-Limit Approach

In the Kaplan and Meier (1958) approach, ordered observations are used instead of grouped data. This method has the advantage of yielding results that are not dependent on the choice of the time intervals. It has been used with small samples, where it is difficult to decide on an appropriate parametric distribution. This technique can be used when the data are progressively censored.

Kaplan and Meier (1958) introduce the product-limit $\hat{P}(t)$ or $\hat{S}(t)$ of $P(t)$ (the probability an individual survives beyond time t) in a life table. To define the product-limit estimate, $\hat{P}(t)$, let $t_i (i=1,2,\ldots,n)$ be a time to death or censoring. Relabel the n times in order of increasing magnitude and denote them by:

$$t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(n)}$$

If $t_{(i)}$ is a time to censoring, then write it as $t'_{(i)}$; if a time to death, omit the prime. Then, as described by Kaplan and Meier, an estimate of the survivorship function $\hat{S}(t)$ is defined by

$$\hat{S}(t) = \frac{(n-1)}{(n-1+1)} \cdot \frac{(n-2)}{(n-2+1)} \cdot \cdots \cdot \frac{(n-r)}{(n-r+1)} \qquad (2.5.1.1)$$

or

$$\hat{S}(t) = \prod_r \frac{(n-r)}{(n-r+1)}$$

where r runs through the positive integers for which $t_{(r)} \leq t$ and $t_{(r)}$ is a time to death. By convention, deaths are immediately before t, and losses or censoring are immediately after t. $\hat{S}(t)$ equals one until the first death occurs and decreases in steps at the times to death. If no individuals are censored, it reduces to the ordinary binomial estimate

$$\hat{S}(t_{(i)}) = 1 - \frac{i}{n} \qquad (2.5.1.2)$$

with the rule that the function is calculated at each distinct time to death, and when two or more observations are tied at the same time the largest (i) value is used.

Example 1: Maintenance of remissions of acute leukemia.

Freireich, Gehan et al (1963) conducted a clinical trial to compare 6-mercaptopurine (6-MP) to a placebo in the maintenance of remissions in patients with acute leukemia. A total of 92 patients were entered in the study and received prednisone therapy; 55 patients achieved complete and 7 achieved partial remission of their disease. On reaching remission, each patient was allocated randomly to receive 6-MP or placebo as

maintenance treatment. The patients were paired at each institution by status of complete remission; the trial was actually conducted sequentially, but the data are given as from a fixed sample size trial. One year after the start of study, the following lengths of remission were recorded (in weeks).

(1) Placebo (21) 1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12, 15,17,22,23

(2) 6-MP (21) 6,6,6,7,10,13,16,22,23,6+,9+,10+,11+, 17+,19+,20+,25+,32+,32+,34+,35+

A plus value indicates a censored observation; there were 12 such observations with 6-MP and none with placebo. The main interest in the study was to characterize the length of remission times between 6-MP and placebo.

The estimates of survivorship function for the placebo patients were calculated using Equation (2.5.1.1) as follows:

t	$\hat{S}(t)$
1,1	0.905
2,2	0.810
3	0.762
4,4	0.667
5,5	0.571
8,8,8,8	0.381
11,11	0.286
12,12	0.190

t	$\hat{S}(t)$
15	0.143
17	0.095
22	0.048
23	0.000

The data are plotted in Figure 2.5 and the estimated median remission time (50th percentile) is 8 weeks. Since there are no censored observations, the estimates of the survivorship function are binomial estimates. For example:

$$\hat{S}(8) = \frac{\text{(number surviving longer than 8 weeks)}}{21} = \frac{8}{21}$$

For the patients receiving 6-MP, the labels for the ordered survival times and times to censoring are shown in Table 2.4. As an example of the calculations, consider

$$\hat{S}(10) = \frac{20}{21}\cdot\frac{19}{20}\cdot\frac{18}{19}\cdot\frac{16}{17}\cdot\frac{14}{15} = 0.753$$

The estimate is obtained by successive multiplication of terms so that

$$\hat{S}(10) = \hat{S}(7)\cdot\frac{14}{15}$$

In general, the estimate $\hat{S}(t_{(i)})$ can be calculated from $\hat{S}(t_{(j)})$ when $t_{(j)}$ is the time to death immediately before $t_{(i)}$ (i.e., j=i-1) by

$$\hat{S}(t_{(i)}) = S(t_{(j)})\cdot\frac{(n-i)}{(n-i+1)} . \qquad (2.5.1.3)$$

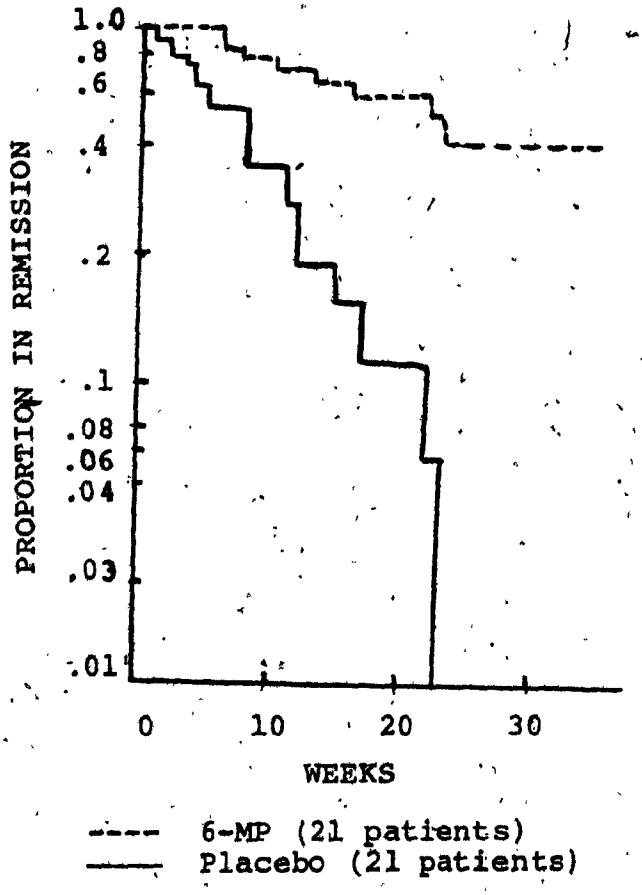

FIG. 2.5 SURVIVORSHIP FUNCTION FOR LENGTHS OF REMISSION IN ACUTE LEUKEMIA

TABLE 2.4 LABELS FOR ORDERED SURVIVAL TIMES AND TIMES TO CENSORING

LABELS	TIMES	$\hat{P}(t)$
$t_{(1)}$	6	
$t_{(2)}$	6	
$t_{(3)}$	6	0.857
$t'_{(4)}$	6+	
$t_{(5)}$	7	0.809
$t'_{(6)}$	9+	
$t'_{(7)}$	10	0.753
$t'_{(8)}$	10+	
$t'_{(9)}$	10+	
$t_{(10)}$	13	0.690
$t_{(11)}$	16	0.627
$t'_{(12)}$	17+	
$t'_{(13)}$	17+	
$t'_{(14)}$	20+	
$t_{(15)}$	22	0.538
$t_{(16)}$	23	0.448
$t'_{(17)}$	25+	
$t'_{(18)}$	32+	
$t'_{(19)}$	32+	
$t'_{(20)}$	34+	
$t'_{(21)}$	35+	

The values of the survivorship function for the 6-MP patients are plotted in Fig.2.5. The median length of remission is 23 weeks. The times beyond 23 weeks are all censored so that $\hat{S}(t)$ is undefined beyond that time. Upper and lower limits can be calculated for $\hat{S}(t)$ by making different assumptions about the censored observations beyond 23 weeks. The lower limit for $\hat{S}(t)$ is obtained by assuming that all individuals relapsed at the time of censoring, the upper limit by assuming that all individuals remain in remission at least up to 35 weeks. After 35 weeks there is no information about the survivorship function.

An approximate estimate of the variance of $\hat{P}(t)$ is

$$\text{Var}[\hat{P}(t)] \doteq [\hat{p}(t)]^2 \left(\sum_r \frac{1}{(n-r)(n-r-1)}\right)$$

where

r runs through positive integers for which $t_{(r)} \leq t$, and $t_{(r)}$ corresponds to a death.

In the example

$$\hat{\text{Var}}[\hat{P}(10)] = (0.753)^2 \left(\frac{1}{(20)(21)} + \frac{1}{(19)(20)} + \frac{1}{(18)(19)} + \frac{1}{(16)(17)} + \frac{1}{(14)(15)}\right) = 0.009284$$

2.5.2 Estimation of Survivorship Function in Large Samples - The Life Table Method

The life table is one of the basic tools in the description of the mortality experience of a population; it has been used extensively by biostatisticians and actuaries to portray the pattern of survival in populations. By survival we usually signify time to death; for example, we could be describing the length of stay in a mental hospital, wherein "birth"is entry into the hospital and "death" is the time at which he relapses.

There are three types of life table in common use - the population life table, the clinical life table and the cohort life table. We shall discuss only the clinical life table because it is useful in deciding what hazard rate to assume. Clinical life tables reflect the thinking of population life tables but use data from clinical studies of patients instead of census vital statistics data. To understand the procedure, let us examine a small set of data from a given clinic. In a study of a disease condition such as lung cancer, it is seldom possible to admit all the patients to the study sample at the same time because of the paucity of patients. We have to accept the patients as they enter for treatment. We then follow each patient to death. However, it is not uncommon to find a portion of the patients lost to follow-up because the researcher can neither find the patient nor determine that he

has died.

Figure 2.6 illustrates some typical cases. For example, patient 1 entered shortly after the start of the study and soon died. Patient 2 entered later and died after a long follow-up. Patient 3 was lost to follow-up during the study period. Patient 4 entered later but died during the study, whereas patient 5 entered later and was withdrawn alive because the study ended.

In working with these data, we assume that neither the treatment nor the characteristics of the patients change during the period of study; hence the starting point for each patient (see the illustration in Figure 2.7 for the same patients shown in Figure 2.6) can be moved to the beginning of the study.

Thus, in clinical trials when the length of time until death of some other event, such as the absence of clinical symptoms (remission), can be long, and when the disease is so rare that we enter patients one at a time over a long period, we obtain progressively censored samples.

The simplest technique for analyzing these data is the T-year survival rate discussed in Berkson and Gage (1950). Only patients who entered the study early (i.e., so that they have a known exposure to the risk of dying in T years) are used. Of this reduced number of patients, the T-year survival

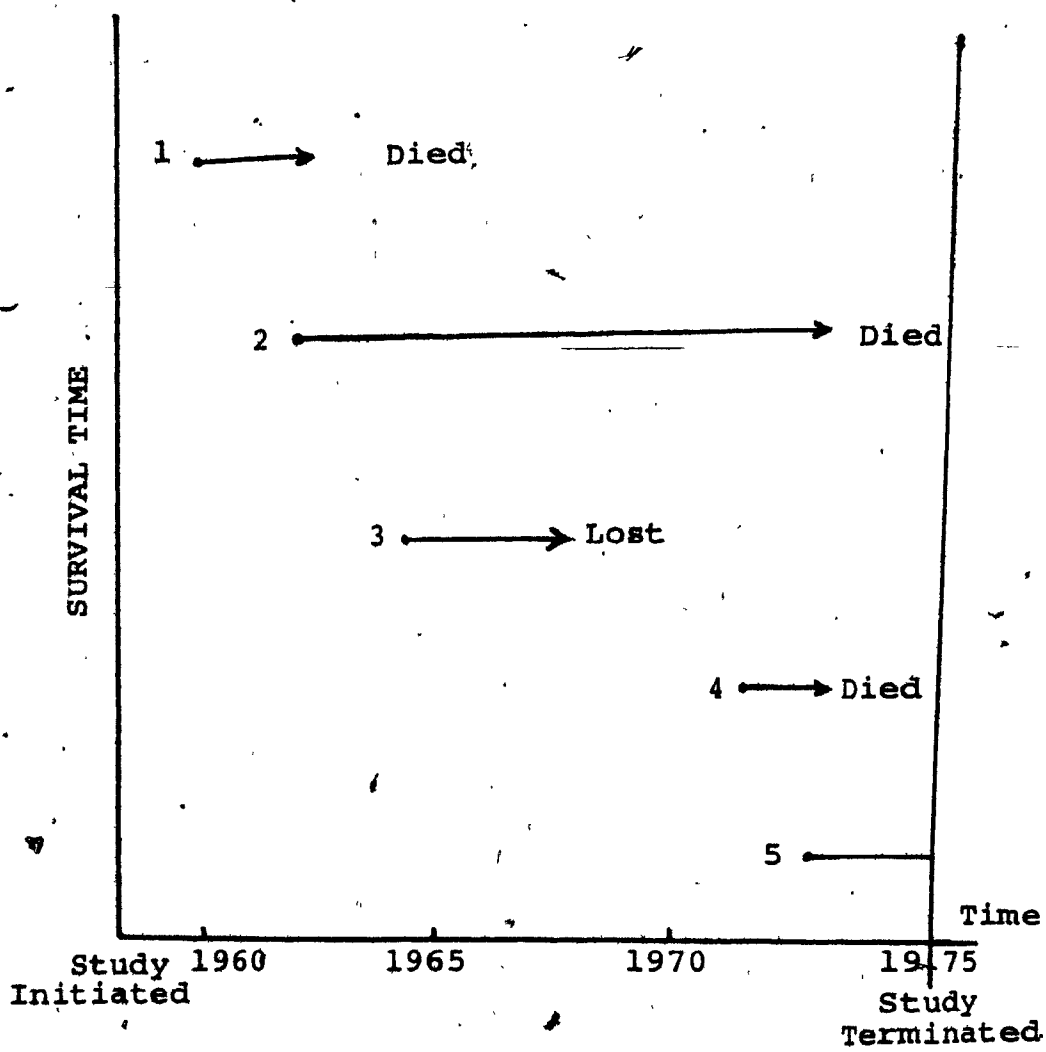

FIG. 2.6 PATIENTS ENTERING AND LEAVING STUDY

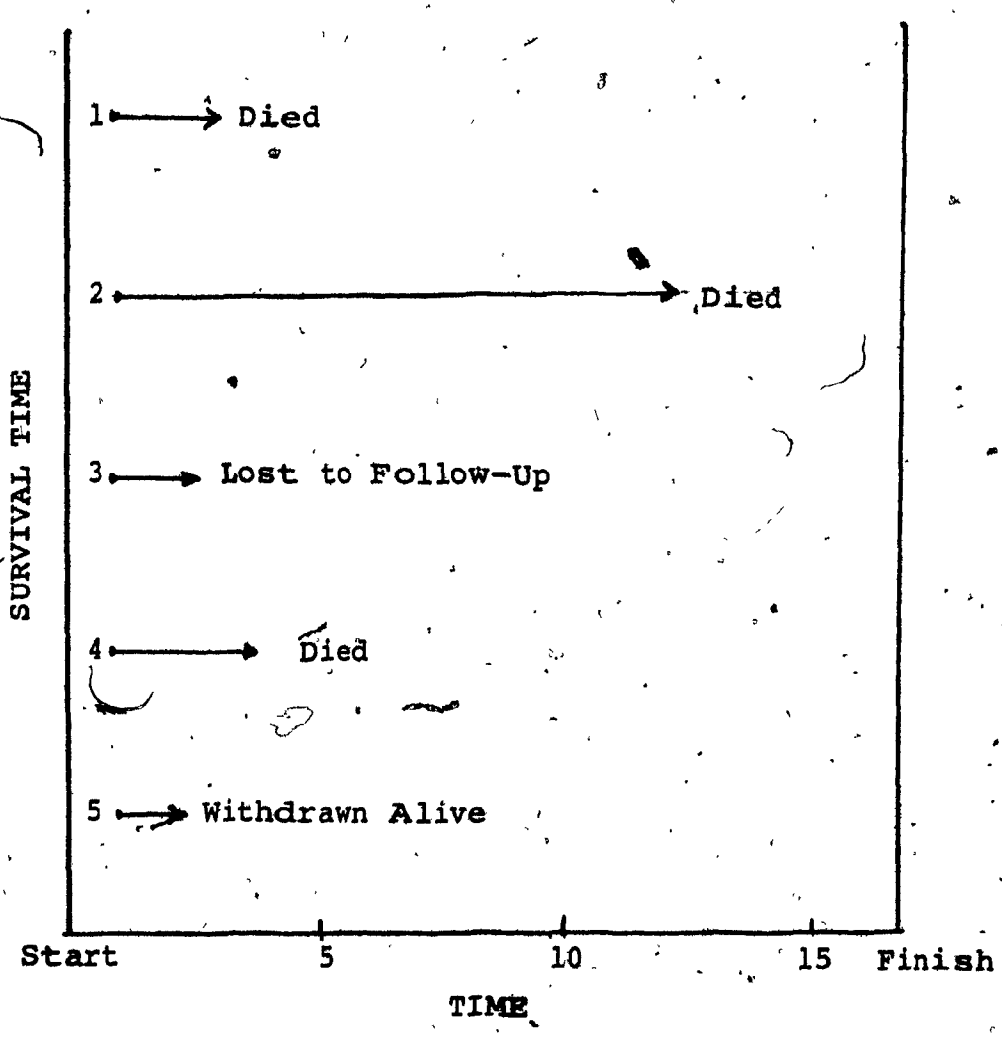

FIG. 2.7 ILLUSTRATION OF INTERVALS AS USED IN CLINICAL LIFE TABLES

rate is the proportion surviving T years. For example, suppose we wish to compute the 5-year survival rate for the patients in Figure 2.6. We would have

$$\text{5-year survival rate} = \frac{\text{Patient 2}}{\text{Patient 1 + Patient 2}} = \frac{1}{2} = 50\%$$

Patient 1 and patient 2 can be used because they entered the study early enough to be exposed to the risk of dying for more than 5 years and they were not lost before 5 years. Patients 3, 4 and 5 could not be used in calculating the rate, since they were not exposed to risk of dying for 5 years. (Patient 3 was lost before 5 years and patients 4 and 5 entered too late.) Thus, all the information available cannot be used, and larger sample sizes are needed to compare a new treatment to previous results or for the simultaneous comparison of two treatments.

Gehan (1967) showed that in conducting a clinical trial to compare the effects of two treatments, appreciably fewer patients would be needed if an exponential distribution were used instead of the T-year survival rate. Since obtaining a large sample is difficult in many medical studies, techniques that allow the investigator to reach conclusions with a smaller sample size (i.e., parametric techniques) have already been discussed.

For the patient who is neither lost nor withdrawn, two points determine the length of time a patient spends in the study - the starting point and the time of death. The time of

death is obvious, but the choice of a proper starting point can be quite difficult. Often comparisons between studies are complicated by different choices of starting points. In a population life table the starting point is birth at year 0. Analogously, to compute a clinical life table, we need to define the starting point from which we calculate survival. For example, consider using onset of the disease as a starting point. For some diseases where the onset is clear-cut and recent, such as skiing accident resulting in a broken leg, this may be adequate. But, for conditions in which the onset is insidious, it is difficult to establish an equivalent point of onset for patients. Other possibilities are first visit to a physician, diagnosis, admission to hospital, start of treatment (e.g., drugs or surgery), completion of treatment (e.g.,surgery), and discharge from hospital.

In addition to deciding on a starting point, the researcher also needs to determine the frequency and manner of follow-up. Each patient could be followed at regular intervals - say one year from his date of entry to the study. Another method is to attempt to trace all the subjects at a particular time, such as January of each year, regardless of when during the year the patient entered the study. In the latter case information often is discarded on the subject during the last period. Elveback (1958) discusses the two methods of follow-up and concludes that the loss due to the discarded information is slight. In either

case, we know at the end of each period whether a subject has lived through the period, has died, or has been lost to follow-up either by the study closing or because of failure to trace the subject. One of the advantages of life table analysis is that data on subjects lost to follow-up can be used up to the period in which they are lost.

To draw up a clinical life table, it is assumed that the number of observations is fairly large (i.e., 50 or more) so that it is sensible to group the survival times into intervals. The individuals whose survival experience is studied could be obtained from: a cohort study (i.e., a group of individuals studied from some zero point, for example the time of start of treatment to death); a series of cohorts analyzed at a particular date (for example, the cohorts could be cases of a disease diagnosed in 1965, 1966, ..., and the time of analysis could be 1974); or a clinical trial in which each patient is observed from the time treatment is started to the end of the study. In each case, the individual's survival time is measured from his own zero point.

To calculate the estimates of survival function, the survival times and times to censoring are grouped into intervals in a life-table format, as given in Table 2.5. Definitions and notations are as follows:

1) Interval $\{[t_{i-1},t_i),i=1,\ldots,s+1,t_{s+1} \equiv +\infty\}$:

This column gives the groupings into which the survival times and times to loss or withdrawal are distributed. The notation $[t_{i-1},t_i)$ represents the half-open time interval $t_{i-1} \leq t < t_i, i=1,\ldots,s+1$. The last interval extends theoretically to infinity. These intervals are assumed to be fixed; thus the number of individuals dying in each interval is a random variable which follows the multinomial distribution.

2) Midpoint (t_{m_i}):

t_{m_i} is the midpoint of the interval $[t_{i-1},t_i),i=1,\ldots,s$. The midpoint is included for convenience in plotting the hazard function and probability density function. Both functions are plotted at t_{m_i}.

3) Width (h_i):

$h_i = t_i - t_{i-1}$, is the width of the $i^{\underline{th}}$ interval. The width of each interval is needed for the calculation of $\hat{\lambda}(t)$, the hazard function and the death density function $\hat{f}(t)$. Since the width of the last interval is infinite, no estimate of either the hazard or death density function can be obtained for this interval.

4) Number lost to follow-up (ℓ_i):

This is the total number of individuals who are lost to observation for some reason and whose survival status thus became unknown in the i^{th} interval. (i.e., total number of individuals who are lost to follow-up during the i^{th} interval $[t_{i-1},t_i)$. Individuals may be lost to observation if they move or fail to return for treatment, etc. Every attempt should be made to trace these patients.

5) Number withdrawn alive (ω_i):

ω_i is the total number of individuals who are withdrawn from the study alive during the i^{th} interval $[t_{i-1},t_i)$. These individuals have not been lost, but since they started late in the study, we have incomplete information on them.

6) Number dying (d_i):

d_i is the number of individuals who die in the i^{th} interval. The time to death for each individual is measured from his own zero point.

7) Number entering the i^{th} interval (n_i'):

n_i' is the total number of individuals who enter the interval $[t_{i-1},t_i)$, $i=1,2,\ldots,s+1$. Thus the total sample size for the study is n_1', the number of individuals who enter the

study at t_o. Clearly $n_i' = n_{i-1}' - \ell_{i-1} - \omega_i - d_{i-1}$. That is, the number of individuals who remain on study at the beginning of the $i^{\underline{th}}$ interval equals the number of individuals who were on study at the beginning of the $(i-1)^{\underline{th}}$ interval minus the number who were lost, withdrawn alive, or who died during the $(i-1)^{\underline{th}}$ interval.

8) Number exposed to risk (n_i):

This number is defined as $n_i' - \frac{1}{2}(\ell_i + \omega_i)$. It is the number of individuals exposed to risk during the $i^{\underline{th}}$ interval. If there are no losses due to follow-up or withdrawal of patients, then clearly $n_i' = n_i$ in the $i^{\underline{th}}$ interval. It is assumed that times to loss or withdrawal are uniformly distributed. Thus on the average individuals who are lost to follow-up or who withdraw from the study are lost for half the total interval.

9) Conditional proportion dying ($\hat{q}_i$):

$\hat{q}_i = d_i/n_i, i=1,\ldots,s$ is the conditional proportion dying in the $i^{\underline{th}}$ interval. Clearly, $\hat{q}_{s+1} = 1$.

10) Conditional proportion surviving ($\hat{p}_i$):

$\hat{p}_i = (1-\hat{q}_i)$ is the conditional proportion surviving the $i^{\underline{th}}$ interval, $i=1,\ldots,s$.

11) Cumulative proportion surviving [$\hat{P}_i$ or $\hat{S}(t_i)$]:

This is an estimate of the survivorship function at time t_i; it is often referred to as the cumulative survival rate. The estimate is $P_i = \hat{p}_{i-1}\,\hat{P}_{i-1}$, where (i=1,...,s) and $\hat{P}_o = 1.00$. This is the usual life table estimate. The result is based on the fact that surviving to the start of the $i^{\underline{th}}$ interval means surviving to the start of the $(i-1)^{\underline{th}}$ interval and given survival to the start of the $(i-1)^{\underline{th}}$ interval, surviving the $(i-1)^{\underline{th}}$ interval. Frequently $\hat{P}_i$ is compared with the cumulative proportion surviving in a population life table starting at the average age of the patients in the study to determine roughly how seriously the disease affects life expectancy.

Table 2.5 summarizes the foregoing information in clinical life tables.

From the definition of $f(t)$, the death density function, the natural estimate is

$$\hat{f}(t_{m_i}) = \frac{\hat{P}_i - \hat{P}_{i+1}}{h_i} = \frac{\hat{P}_i\,\hat{q}_i}{h_i},\quad i=1,2,\ldots,s \qquad (2.5.2.1)$$

Thus $\hat{f}(t_{m_i})$ is the estimated probability of dying in the $i^{\underline{th}}$ interval per unit width; that is the definition of $f(t)$. The hazard rate $\hat{\lambda}(t_{m_i})$ is not directly obtained as $\hat{f}(t_{m_i})/\hat{P}_i$, since $\hat{P}_i$ is the probability of survival at t_i not t_{m_i}. We

TABLE 2.5 CLINICAL LIFE TABLE

INTERVAL	MID-POINT	WIDTH	NUMBER ENTERING INTERVAL	NUMBER LOST TO FOLLOW-UP	NUMBER WITHDRAWN ALIVE	NUMBER EXPOSED TO RISK	NUMBER DYING	CONDITIONAL PROPORTION DYING	CONDITIONAL PROPORTION SURVIVING	CUMULATIVE PROPORTION SURVIVING	$\hat{f}(t_{m_i})$	$\hat{\lambda}(t_{m_i})$
$[t_0,t_1)$	t_{m_1}	h_1	n'_1	ℓ_1	ω_1	n_1	d_1	$\hat{q}_1$	$\hat{p}_1$	$\hat{P}_1 = 1.00$	$\hat{f}(t_{m_1})$	$\hat{\lambda}(t_{m_1})$
$[t_1,t_2)$	t_{m_2}	h_2	n'_2	ℓ_2	ω_2	n_2	d_2	$\hat{q}_2$	$\hat{p}_2$	$\hat{P}_2$	$\hat{f}(t_{m_2})$	$\hat{\lambda}(t_{m_2})$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$[t_{s-1},t_s)$	t_{m_s}	h_s	n'_s	ℓ_s	ω_s	n_s	d_s	$\hat{q}_s$	$\hat{p}_s$	$\hat{P}_s$	$\hat{f}(t_{m_s})$	$\hat{\lambda}(t_{m_s})$
$[t_s,\infty)$	-	-	n'_{s+1}	-	-	n'_{s+1}	d_{s+1}	1.00	0.00	$\hat{P}_{s+1}$	-	-

$h_i = t_i - t_{i-1}, i=1,2,\ldots,s$

$n_i = n'_i - \frac{1}{2}(\ell_i + \omega_i), i=1,2,\ldots,s$

$\hat{q}_i = d_i/n_i, \; i=1,2,\ldots,s$

$\hat{q}_{s+1} = 1.00$

$\hat{p}_i = 1-\hat{q}_i, i=1,2,\ldots,s$

$\hat{S}(t_i) = P_i$

$\hat{P}_i = \hat{p}_{i-1}\,\hat{P}_{i-1}, i=2,\ldots,s$

$\hat{P}_1 = 1.00$

$$\hat{f}(t_{m_i}) = \frac{\hat{P}_i - \hat{P}_{i+1}}{h_i}, \; i=1,2,\ldots,s$$

$$\hat{\lambda}(t_{m_i}) = \frac{2(1-\hat{p}_i)}{h_i(1+\hat{p}_i)} = \frac{2\hat{q}_i}{h_i(1+\hat{p}_i)}, \; i=1,2,\ldots,s$$

thus must find $\hat{P}(t_{m_i})$, the probability of survival at t_{m_i}, the mid-point of the $i^{\underline{th}}$ interval. Clearly

$$\hat{P}(t_{m_i}) = \frac{\hat{P}_{i+1} + \hat{P}_i}{2} = \frac{\hat{P}_i(1+\hat{p}_i)}{2} \qquad (2.5.2.2)$$

Thus,

$$\hat{\lambda}(t_{m_i}) = \frac{1}{h_i} \frac{d_i}{(n_i - d_i/2)} = \frac{2\hat{q}_i}{h_i(1+\hat{p}_i)} \qquad (2.5.2.3)$$

This is the so-called actuarial estimate of hazard function and is described by Kimball (1960). In words, it is the number of deaths per unit time in the interval divided by the average number of survivors at the midpoint of the interval. The average number of individuals alive at t_{m_i} is approximated by $(n_i - d_i/2)$. Other estimates of hazard function are given by Kimball (1960) and Watson and Leadbetter (1964).

The variances of the estimates of the survival functions in the $i^{\underline{th}}$ interval are:

$$\mathrm{Var}[\hat{S}(t_i)] \cong \hat{P}_i^2 \sum_{j=1}^{i-1} \frac{\hat{q}_i}{n_j \hat{p}_j} \qquad (2.5.2.4)$$

$$\mathrm{Var}[\hat{f}(t_{m_i})] \cong \frac{(\hat{P}_i \hat{q}_i)^2}{h_i^2} \left(\sum_{j=1}^{i-1} \frac{\hat{q}_i}{n_j \hat{p}_j} + \frac{\hat{p}_i}{n_i \hat{q}_i} \right) \qquad (2.5.2.5)$$

TABLE 2.6 SURVIVAL OF MALIGNANT MELANOMA AT THE M.D. ANDERSON TUMOUR CLINIC, 1944-1960[a]

INTERVAL (YEARS)	MID-POINT	WIDTH	NUMBER ENTERING INTERVAL	NUMBER LOST TO FOLLOW-UP	NUMBER WITHDRAWN ALIVE	NUMBER EXPOSED TO RISK	NUMBER DYING	CONDITIONAL PROPORTION SURVIVING	CUMULATIVE PROPORTION SURVIVING	$\hat{f}(t_{m_i})$	$\hat{\lambda}(t_{m_i})$
[0,1)	0.5	1	913	19	77	865.0	312	0.639	1.000	0.361	0.441
[1,2)	1.5	1	505	3	71	468.0	96	0.795	0.639	0.131	0.228
[2,3)	2.5	1	335	4	58	304.0	45	0.852	0.508	0.075	0.160
[3,4)	3.5	1	228	3	27	213.0	29	0.864	0.433	0.059	0.146
[4,5)	4.5	1	169	5	35	149.0	7	0.953	0.374	0.018	0.048
[5,6)	5.5	1	122	1	36	103.5	9	0.913	0.356	0.031	0.091
[6,7)	6.5	1	76	0	17	67.5	3	0.956	0.325	0.014	0.045
[7,8)	7.5	1	56	2	10	50.0	1	0.980	0.311	0.006	0.020
[8,9)	8.5	1	43	0	8	39.0	3	0.923	0.305	0.024	0.080
$[9,\infty)$[b]	-	-	32	-	-	32.0	32	0.000	0.281	-	-

[a]Patient years are measured from time of initial diagnosis. It is also noted that these 913 patients did not all enter study at the same time.

[b]All 32 patients must die in the last interval.

$$\mathrm{Var}[\hat{\lambda}(t_{m_i})] \cong \frac{[\hat{\lambda}(t_{m_i})]^2}{n_i \hat{q}_i}\left(1 - \left[\frac{\hat{\lambda}(t_{m_i})h_i}{2}\right]^2\right) \qquad (2.5.2.6)$$

All the formulas are large-sample approximations. Equation (2.5.2.4) was first given by Greenwood (1926), and Equations (2.5.2.5) and (2.5.2.6) were derived by Gehan (1969). These may be used to obtain approximate confidence limits for the various survival functions.

To exhibit computations for the clinical life table we use the clinical data gathered on 913 male and female patients with malignant melanoma who were treated at the M.D. Anderson Tumour Clinic between 1944 and 1960. These data, given by MacDonald (1963), comprise Table 2.7. We note that the computed failure rate for these patients is decreasing. That is, if an individual survives 3 years with the disease it appears he has an excellent chance to survive 5 years or longer. These 913 patients include 256 males and 213 females who had metastases when they were first seen, plus an additional 444 patients in whom metastases had not occurred or could not be established.

CHAPTER III

SURVIVAL MODELS FOR MULTICOMPONENT SYSTEM

3.1 INTRODUCTION

We have so far presented methods for data analysis when there is a single, possibly censored, failure time on each study subject. Certain times there may be more than one failure time on each study subject. Such multivariate failure times correspond to repeated occurrences of events of entirely different natures. Consider, for instance, the two-organ system such as the kidneys or the lungs. Assume that if one organ fails, the surviving organ is subject to a different failure rate (usually higher); however, both failure rates are assumed constant over time. Such a survival distribution could be used to estimate the survival of individuals with lung or kidney disease. Persons who have a kidney removed due to some illness commonly exhibit a higher failure rate for the remaining kidney. On the other hand, if their kidney is removed because of an accident, the remaining kidney often does not show an increase in its failure rate. The derived survival distribution will account for both situations. In all these we assume that there are no survivors at the point the estimates are made. That is, we assume uncensored observations. In general, study subjects may experience a variable number of failures each with its own

type or cause.

There are several basic multivariate parameter families of distributions such as the multivariate gamma, Weibull, normal and exponential distributions and shock models that give rise to them which have been much investigated in reliability theory. Marshall and Olkin (1967) have discussed a multivariate exponential distribution, (i.e., multivariate distributions with exponential marginals) as applied to life-testing. Friday (1967) has also considered a general class of multivariate life distributions with some possible probalistic bases for them.

Although there exists an infinite number of bivariate distributions with given marginals, very few shed any light on their applicability. Freund (1961) derived a bivariate distribution starting with exponential marginals. However, the resulting distribution does not have exponential marginals. Marshall and Olkin (1967) have also considered another bivariate exponential distribution which has many interesting properties, but unfortunately, it has a singular component. Moreover, unlike in the univariate distribution, the conditional failure rate in this case is not constant everywhere. Since the exponential distribution is absolutely continuous and has constant failure rate everywhere in the univariate case, a natural bivariate exponential distribution is also expected to be absolutely continuous with constant bivariate failure rate.

Basu (1971) showed that no such bivariate (and therefore multivariate) distribution exists, except in the special case when the marginals are independently distributed.

In the next section we discuss some aspects of bivariate exponential model and develop a survival distribution for two organ systems such as the lungs or the kidneys and in Section 3.3, we consider the estimation procedures.

3.2 BIVARIATE FAILURE MODEL

Consider a system with two components. Let X and Y be the lives of the two components with failure laws governed by the respective marginal density functions

$$f(x) = \lambda_1 \exp(-\lambda_1 x), \quad f(y) = \lambda_2 \exp(-\lambda_2 y) \tag{3.2.1}$$

$$\lambda_1, \lambda_2 > 0, \ 0 < x, y < \infty$$

Let $F(x,y)$ be an absolutely continuous distribution function with density function $f(x,y)$ governing the life of the system as a whole. This implies

$$P(X = Y) = 0 \tag{3.2.2}$$

The physical interpretation of Equation (3.2.2) is that the model represents the life of the system under normal (preventive) maintenance policy so that the possibility of a catastrophe

resulting in simultaneous failure of both the components are removed and replaced by identical units and, so far as the system is concerned, failure can occur only in the intervals between successive inspections when the system is in operation. Thus after each replacement we may start with (for our purpose) essentially a new system. It should also be noted that, in any interval, to start with we may have $X > Y$ or $X < Y$ depending on which component was previously replaced. The preceding conditions are quite realistic and have important possible applications in a situation where we are concerned about having a failure while the system is carrying out some important mission, as for example, flight of a twin-engine plane, the space flight in progress, etc. For deriving the actual model we introduce the concept of bivariate failure rate as given in the following.

Definition:

Given an absolutely continuous bivariate distribution function $F(x,y)$ with density function $f(x,y)$, the bivariate failure rate at (x,y) is given by

$$\lambda(x,y) = \frac{f(x,y)}{P(X>x, Y>y)}$$

$$= \frac{f(x,y)}{1 + F(x,y) - F(x,\infty) - F(\infty,y)} \qquad (3.2.3)$$

Observe that in the case of independence we have (as it should be)

$$\lambda(x,y) = \frac{f(x)f(y)}{P(X>x)P(Y>y)} = \frac{f(x)}{1-F(x)} \cdot \frac{f(y)}{1-F(y)}$$

$$= \lambda(x)\lambda(y) \qquad (3.2.4)$$

where

$\lambda(x)$ and $\lambda(y)$ are the corresponding univariate failure rates.

Since we know that the failure rate of the univariate exponential distribution is a constant, for a natural bivariate exponential distribution $\lambda(x,y)$ is expected to be constant for all x and y. Basu (1971) showed that except in the case of independence there does not exist any absolutely continuous bivariate exponential distribution with constant failure rate (and marginal exponential distributions.)

Let $F(x,y)$ be the required distribution function with density function $f(x,y)$. Denoting the bivariate survival function $P(X>x,Y>y)$ by $S(x,y)$ we want to find $f(x,y)$ such that

$$\lambda(x,y) = \frac{f(x,y)}{S(x,y)} = \lambda \qquad (\lambda>0) \qquad (3.2.5)$$

$$1 - F(x,\infty) = S(x,0) = \exp(-\lambda_1 x) \tag{3.2.6}$$

$$1 - F(\infty,y) = S(0,y) = \exp(-\lambda_2 y) \tag{3.2.7}$$

$$F(\infty,\infty) = S(0,0) = 1 \tag{3.2.8}$$

That is, we have to solve the second-order partial differential equation

$$S_{xy}(x,y) - \lambda S(x,y) = 0 \tag{3.2.9}$$

with the boundary conditions (3.2.6), (3.2.7) and (3.2.8). To show that Equations (3.2.6) and (3.2.7) are necessary conditions, consider the survival function

$$S(x,y) = \sum_{i=1}^{\infty} \exp(-\lambda_i x - \mu_i y) p_i \tag{3.2.10}$$

where

$$\lambda_i \mu_i = \lambda, \quad i=1,2,\ldots \quad \text{and} \quad \sum_{i=1}^{\infty} p_i = 1.$$

Here $S(x,y)$ is a mixture of several survival functions corresponding to independent bivariate exponential distributions with different scale parameters. In the absence of Equations (3.2.6) and (3.2.7), $S(x,y)$ in Equation (3.2.10) will be a solution for the problem of a linear second-order partial differential equation, called the Goursat problem [see Garabedian (1964), p. 117].

Instead of solving for $S(x,y)$ let us first find out the Laplace-Stieltjes transform $\phi(s,t)$ corresponding to $S(x,y)$

$$\phi(s,t) = \int_0^\infty \int_0^\infty \exp(-sx-ty)S(x,y)\,dxdy \qquad (3.2.11)$$

Integrating by parts with respect to x first and using Equation (3.2.7), we have

$$\phi(s,t) = \frac{1}{s}\left\{\frac{1}{(t+\lambda_2)} + \int_0^\infty \int_0^\infty \exp(-sx-ty)S_x(x,y)\,dydx\right\} \qquad (3.2.12)$$

Integrating by parts with respect to y next and using Equations (3.2.9) and (3.2.6) we get

$$\phi(s,t) = \frac{1}{s(t+\lambda_2)} - \frac{\lambda_1}{st(s+\lambda_1)} + \frac{\lambda}{st}\phi(s,t) \qquad (3.2.13)$$

Hence,

$$\phi(s,t) = \frac{st - \lambda_1\lambda_2}{(t+\lambda_2)(s+\lambda_1)(st-\lambda)} \qquad (3.2.14)$$

Clearly, the Laplace-Stieltjes transform of $f(x,y) = \lambda S(x,y)$ is

$$\psi(s,t) = \lambda\phi(s,t) = \frac{\lambda(st-\lambda_1\lambda_2)}{(t+\lambda_2)(s+\lambda_1)(st-\lambda)} \qquad (3.2.15)$$

From Equation (3.2.14) it is clear that $\phi(s,t)$ blows up unless $\lambda = \lambda_1\lambda_2$, which yields

$$\psi(s,t) = \frac{\lambda_1\lambda_2}{(t+\lambda_2)(s+\lambda_1)} \qquad (3.2.16)$$

the Laplace-Stieltjes transform of the distribution with x and y following independent exponential distributions. Thus, we have shown that we cannot have an absolutely continuous bivariate exponential distribution with constant failure rate unless x and y are independently distributed.

Let us now develop a bivariate failure model for a two-organ system. Consider a two-organ system such that the failure rate of each organ is constant and equals λ_o if both organs are functioning. However, as soon as one organ fails, the failure rate of the remaining functioning organ is constant and equals $\lambda_1 \geq \lambda_o$. In such a two-organ system there are two states in which the individual may be and still is alive: S_o the state when both organs are functioning and S_1 the state when one organ has failed and the individual is left with only one functioning organ.

Let $p_o(t) = \Pr\{\text{System is in } S_o \text{ at time } t\}$

$p_1(t) = \Pr\{\text{System is in } S_1 \text{ at time } t\}$

We assume that at time $t = 0$, i.e., the time of onset of first symptoms, $p_o(0) = 1$, and $p_1(0) = 0$.

At present, we shall assume that $\lambda_1 \neq 2\lambda_o$ and without loss of generality with respect to this assumption we take

$\lambda_1 > 2\lambda_o$. Using the initial conditions $p_o(0) = 1$, and $p_1(0) = 0$, and the fact that λ_o and λ_1 are constant with respect to time, it is easily shown that

$$p_o(t) = \exp(-2\lambda_o t), \qquad t \geq 0 \tag{3.2.17}$$

$$p_1(t) = \frac{2\lambda_o}{\lambda_1 - 2\lambda_o} [\exp(-2\lambda_o t) - \exp(-\lambda_1 t)], \quad t \geq 0 \tag{3.2.18}$$

$S(t)$, the probability the individual survives to time t, is simply $p_o(t) + p_1(t)$. Thus for $t \geq 0$

$$S(t) = \exp(-2\lambda_o t) + \frac{2\lambda_o}{\lambda_1 - 2\lambda_o} [\exp(-2\lambda_o t) - \exp(-\lambda_1 t)] \tag{3.2.19}$$

$F(t)$, the probability the individual dies prior to t, is $1 - S(t)$ and so the requisite failure density $f(t)$ is $dF(t)/dt$.

Hence, for $\lambda_1 > 2\lambda_o$,

$$f(t) = \frac{2\lambda_o \lambda_1}{\lambda_1 - 2\lambda_o} [\exp(-2\lambda_o t) - \exp(-\lambda_1 t)] \quad t \geq 0 \tag{3.2.20}$$

The hazard rate, $h(t)$, corresponding to Equation (3.2.20) is

$$h(t) = \frac{2\lambda_o \lambda_1}{\{(\lambda_1 - 2\lambda_o)[1 - \exp(-\lambda_1 t + 2\lambda_o t)]^{-1} + 2\lambda_o\}} \tag{3.2.21}$$

From this it follows that $h(t)$ is an increasing function of t for $\lambda_1 \neq 2\lambda_o$. In fact, $h(0) = 0$ and $h(\infty) = 2\lambda_o$ or λ_1, accordingly as $\lambda_1 < 2\lambda_o$ or $\lambda_1 > 2\lambda_o$.

The moment generating function (mgf) corresponding to $f(t)$, $M(\theta)$, is given by

$$M(\theta) = [2\lambda_o\lambda_1/(\lambda_1-2\lambda_o)][(\theta-\lambda_1)^{-1} - (\theta-2\lambda_o)^{-1}] \quad (3.2.22)$$

where

$$\theta > \lambda_1 > 2\lambda_o$$

From Equation (3.2.22), it is easy to show that μ_r', the $r^{\underline{th}}$ moment about the origin of $f(t)$, is

$$\mu_r' = r!(2\lambda_o\lambda_1)^{-r}\left[\frac{(\lambda_1^{r+1}-(2\lambda_o)^{r+1}}{(\lambda_1-2\lambda_o)}\right] \quad (3.2.23)$$

Thus μ and σ^2, the mean and variance of t, are

$$\mu = \lambda_1^{-1} + (2\lambda_o)^{-2} \quad (3.2.24)$$

$$\sigma^2 = \lambda_1^{-2} + (2\lambda_o)^{-2} \quad (3.2.25)$$

Equations (3.2.24) and (3.2.25) are used (method of moments) to obtain initial estimates in solving iteratively for the maximum likelihood estimates.

The case for which $\lambda_1 = 2\lambda_o$ reduces in essence to the convolution of two exponential densities with the same parameter. The resulting density is thus a gamma density with shape parameter equal to 2.

Another special case occurs when the loss of one organ has no effect on the other organ. That is, $\lambda_o = \lambda_1 = \lambda$(say). For this situation,

$$S(t) = e^{-\lambda t}(2-e^{-\lambda t}), \quad t \geq 0, \ \lambda > 0$$

$$f(t) = 2\lambda e^{-\lambda t}(1-e^{-\lambda t}), \ t \geq 0, \ \lambda > 0$$

$$h(t) = 2\lambda(1-e^{-\lambda t})/(2-e^{-t}), \ t \geq 0, \ \lambda > 0$$

$$M(\theta) = 2\lambda^2[(\theta-\lambda)(\theta-2\lambda)]^{-1},$$

$$\mu_r' = r!(2\lambda)^{-r}(2^{r+1}-1),$$

$$\mu = \frac{3}{2\lambda}, \quad \sigma^2 = \frac{5}{4\lambda^2}$$

In order to apply the model to a clinical situation such as estimating the mean survival time for lung cancer patients or patients with severe kidney disease, it is necessary to know not only survival times for the patients, but one must be able to estimate the time the first of the two organ(lung or kidney) failures has occurred.

Recently Tosch and Holmes(1980) have proposed a new bivariate failure model in which the residual life time of one component is dependent on the working status of the other component. This is applicable when the failure of one component puts more (possibly less) strain on the remaining component. For example, consider a machine with two components, each performing the same job and contributing to the functioning of the machine. When one component fails, the other must bear the entire burden and thus suffers a shortened life-time. Another example is the function of the two kidneys in the human body. They derived properties of the life-times including their joint Laplace-Stieltjes transform and also considered the bivariate exponential life-time and the estimation of its parameters.

3.3 ESTIMATION OF PARAMETERS OF THE TWO-ORGAN SYSTEM

Suppose we have complete samples of failure times $t_1, t_2, \ldots, t_N$, on independent and identical organ systems whose failure density is given by Equation (3.2.20). With $\alpha = 2\lambda_o$ and $\beta = \lambda_1$ the likelihood function $L(\alpha, \beta)$ is

$$L = \prod_{i=1}^{n} \left[\frac{\alpha\beta}{(\beta-\alpha)}\right] (e^{-\alpha t_i} - e^{-\beta t_i}), \qquad (3.3.1)$$

which leads to the likelihood equations

$$\frac{N\hat{\beta}}{\hat{\alpha}(\hat{\beta}-\hat{\alpha})} - \sum_{i=1}^{N} \frac{t_i \exp(-\hat{\alpha}t_i)}{[\exp(-\hat{\alpha}t_i) - \exp(-\hat{\beta}t_i)]} = 0 \qquad (3.3.2)$$

$$\frac{-N\hat{\alpha}}{\hat{\alpha}(\hat{\beta}-\hat{\alpha})} + \sum_{i=1}^{N} \frac{t_i \exp(-\hat{\beta}t_i)}{[\exp(-\hat{\alpha}t_i) - \exp(-\hat{\beta}t_i)]} = 0 \qquad (3.3.3)$$

These are solved iteratively using the methods given in Section 2.2 to obtain $\hat{\alpha}$ and $\hat{\beta}$.

The large-sample variance covariance matrix for $\hat{\alpha}$ and $\hat{\beta}$ is required for the method of scoring, so we now proceed to obtain it. However, we first need to evaluate a non-standard integral.

Lemma 1:

Suppose $\beta > \alpha > 0$ are given. Then

$$\int_0^\infty \frac{t^2}{(e^{\beta t} - e^{\alpha t})}\, dt = \sum_{n=0}^{\infty} \frac{2}{[n(\beta-\alpha) + \beta]^3} \equiv \frac{2\zeta(3, \beta/(\beta-\alpha))}{(\beta-\alpha)^3}$$

where $\zeta(3, \beta/(\beta-\alpha))$ is the generalized Riemann zeta-function [see Whittaker and Watson (1927), p.45].

Proof:

Since $\beta > \alpha$, we can write

$$(e^{\beta t}-e^{\alpha t})^{-1} \equiv e^{-\beta t}(1-e^{-(\beta-\alpha)t})^{-1}$$

Now, by applying L'hôpital's rule

$$\lim_{t\to 0} \frac{t^2}{e^{\beta t}-e^{\alpha t}} = 0$$

Since for $t > 0$,

$$e^{-(\beta-\alpha)t} < 1$$

we can expand $[1 - e^{-(\beta-\alpha)t}]^{-1}$ in a power series and the result follows.

Gross and Clark (1971), for computational purposes derived the following bounds for the series in the above lemma.

$$2\{(\beta-\alpha)^{-3}[\zeta(3) - \sum_{n=1}^{m} \frac{1}{n^3}]\} < 2\sum_{n=0}^{\infty}[n(\beta-\alpha)+\beta]^{-3}$$

$$\leq 2\{(\beta-\alpha)^{-3}[\zeta(3) - \sum_{n=1}^{m-1} n^{-3}]\}$$

where $m = [\beta(\beta-\alpha)^{-1}]$ is the largest integer less than or equal to $\beta(\beta-\alpha)^{-1}$ and $\zeta(3) = 1.20257$.

Thus we have

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \frac{-N\beta(\beta-2\alpha)}{\alpha^2(\beta-\alpha)^2} - \sum_{i=1}^{N} \frac{t_i^2 e^{-(\alpha+\beta)t_i}}{(e^{-\alpha t_i}-e^{-\beta t_i})^2},$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = \frac{-N(\alpha-2\beta)}{\beta^2(\beta-\alpha)^2} - \sum_{i=1}^{N} \frac{t_i^2 e^{-(\alpha+\beta)t_i}}{(e^{-\alpha t_i}-e^{-\beta t_i})^2},$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial \beta} = \frac{-N}{(\beta-\alpha)^2} + \sum_{i=1}^{N} \frac{t_i^2 e^{-(\alpha+\beta)t_i}}{(e^{-\alpha t_i}-e^{-\beta t_i})^2}.$$

Using Lemma 1 with $\beta > \alpha$ we find the variance-covariance matrix $\underset{\sim}{V}$ for $\hat{\alpha}$ and $\hat{\beta}$ is

$$\underset{\sim}{V} = \begin{bmatrix} \frac{N\beta(\beta-2\alpha)}{\alpha^2(\beta-\alpha)^2} + N\alpha\beta\phi & \frac{N}{(\beta-\alpha)^2} - N\alpha\beta\phi \\ \frac{N}{(\beta-\alpha)^2} - N\alpha\beta\phi & \frac{N\alpha(\alpha-2\beta)}{\beta^2(\beta-\alpha)^2} + N\alpha\beta\phi \end{bmatrix} \qquad (3.3.4)$$

where

$$\phi = (\beta-\alpha)^{-4}\,\zeta(3,\beta/(\beta-\alpha)).$$

We now consider the problem of solving for the ML estimators $\hat{\alpha}$ and $\hat{\beta}$ using Equations (3.3.2) and (3.3.3). The corresponding iterative equations are

$$\underset{\sim}{\Delta\hat{\gamma}}^i = \underset{\sim}{V}^i \underset{\sim}{g}^i \qquad (3.3.5)$$

where

$$\underset{\sim}{V}^i \equiv \underset{\sim}{V}\Big|_{\alpha=\hat{\alpha}^i,\ \beta=\hat{\beta}^i} \tag{3.3.6}$$

$\underset{\sim}{V}$ is given by Equation (3.3.4) for the method of scoring, and the raw second derivatives would be used in place of the elements of $\underset{\sim}{V}$ for the Newton-Raphson method. Also

$$\underset{\sim}{\Delta\hat{\gamma}}^i = \begin{bmatrix} \Delta\hat{\alpha}^i \\ \Delta\hat{\beta}^i \end{bmatrix} \equiv \begin{bmatrix} \hat{\alpha}^{i+1} - \hat{\alpha}^i \\ \hat{\beta}^{i+1} \quad \hat{\beta}^i \end{bmatrix} \tag{3.3.7}$$

and

$$\underset{\sim}{g}^i = \begin{bmatrix} \partial \log L/\partial\alpha \\ \partial \log L/\partial\beta \end{bmatrix}\Bigg|_{\alpha=\hat{\alpha}^i,\ \beta=\hat{\beta}^i} \tag{3.3.8}$$

The subscript i in Equations (3.3.5) - (3.3.8) refers to the $i^{\underline{th}}$ iteration of the procedure. To obtain the starting values $\hat{\alpha}^o$ and $\hat{\beta}^o$, we use the method of moments employing Equations (3.2.24) and (3.2.25). Thus

$$\hat{\alpha}^o = 2[\bar{t} \pm \sqrt{(2s_t^2 - \bar{t}^2)}]^{-1} \tag{3.3.9}$$

and

$$\hat{\beta}^o = 2[\bar{t} \pm \sqrt{(2s_t^2 - \bar{t}^2)}]^{-1} \tag{3.3.10}$$

where

$$\bar{t} = \sum_{i=1}^{N} t_i/N \quad \text{and} \quad s_t^2 = \sum_{i=1}^{N} (t_i - \bar{t})^2/N-1)$$

If $\bar{t} > \sqrt{(2s_t^2 - \bar{t}^2)}$ we substitute each pair $\hat{\alpha}^o$ and $\hat{\beta}^o$ from Equations(3.3.9) and (3.3.10) into the likelihood function L and choose as our starting values that pair which gives the larger likelihood value. If $\bar{t} < 2s_t^2 - \bar{t}^2$, we use $\hat{\beta}^o = 2/\bar{t}$. The successive pairs $(\hat{\alpha}', \hat{\beta}'), \ldots, (\hat{\alpha}^n, \hat{\beta}^n)$ are then obtained iteratively by either the Newton-Raphson technique or the method of scoring. An appropriate elliptical confidence region for α and β and hence λ_o and λ_1 can be obtained following the procedure in Mood and Graybill ([1963],pp. 264-7).

Gross and Clark (1971) have given a numerical comparison of the Newton-Raphson and Scoring Methods, through Monte Carlo using survival data from a group of patients with intractable, bilateral renal calculi. These patients will lose, on the average, effective function of one kidney in 15 years after the onset of symptoms and have a mean time to death due to loss of the remaining kidney three years later. Using as population parameters $\lambda_o = \frac{1}{2}\alpha = \frac{1}{15} = 0.0667$ failures per year and $\lambda_1 = \beta = \frac{1}{3} = 0.333$ deaths per year. A Monte Carlo study of the two iterative procedures are performed. They obtained three different estimates of λ_o and λ_1: the method of moments estimates (ii) the method of scoring estimates (iii) the Newton-Raphson estimates.

The Monte Carlo study then proceeds as follows:

(i) Random numbers were drawn between 0 and 1. This was done for sample sizes of 200.

(ii) These random numbers were converted to data having survival distribution as given by Equation (3.2.19) by computing the two survival times separately for the two kidneys,

$$t_o = \log[1 - F_o(t)]/ - \alpha \qquad (3.3.11)$$

$$t_1 = t_o + \log[1 - F_1(t)]/ - \beta \qquad (3.3.12)$$

where a pair of random numbers was inserted in $F_o(t)$ and $F_1(t)$.

(iii) Initial values $\hat{\lambda}_o^o = \frac{1}{\hat{\alpha}^o}$ and $\hat{\lambda}_1 = \frac{1}{\hat{\beta}^o}$ were obtained using Equations (3.3.11), (3.3.12), and the method of moments (3.3.9) and (3.3.10).

(iv) These initial values were used in both the Newton-Raphson and the scoring iteration techniques to obtain the ML estimates $\hat{\lambda}_o$ and $\hat{\lambda}_1$. In addition, estimates of the variances, covariances, and correlations of $\hat{\lambda}_o$ and $\hat{\lambda}_1$ were obtained for both the Newton-Raphson and the scoring methods.

(v) This process was repeated 100 times. Averages and actual variances over the 100 samples were obtained for the estimated $\tilde{\lambda}_o$ and $\hat{\lambda}_1$.

The results are given on the following page in Table 3.1. The upper two sections of the table give the average results of (iv) for the 100 samples. The lower section gives the actual variances of the computed estimates.

The conclusions of this study may be given as follows:

1) Actually, the two methods are quite similar; however, the rate of convergence is slower in the scoring method as compared to the Newton-Raphson.

2) The method of moments estimates are not too unreasonable except for the bias in the estimates of λ_1.

3) The large variances obtained using the variance-covariance matrix or the inverse of the matrix of second derivatives in comparison to the variances of the actual 100 sample statistics are notable. However, the authors in checking the sample statistics, found that the variances obtained from the inverse were roughly five times as large in samples with large $\hat{\lambda}_o$ and $\hat{\lambda}_1$ as in samples with small values.

TABLE 3.1 MONTE CARLO ESTIMATION RESULTS OF THE PATIENTS WITH INTRACTABLE BILATERAL CALCULI: $\lambda_o = 0.0667$ and $\lambda_1 = 0.3333$

	AVERAGE SAMPLE STATISTICS		
	Method of Moments	Scoring	Newton-Raphson
$\hat{\lambda}_o$	0.06807	0.06874	0.06874
$\hat{\lambda}_1$	0.36506	0.34838	0.034842
Var $\hat{\lambda}_o$	---	0.000253	0.000344
Var $\hat{\lambda}_1$	---	0.013372	0.014989
Cov$(\hat{\lambda}_o, \hat{\lambda}_1)$	---	-0.001246	-0.001516
Corr$(\hat{\lambda}_o, \hat{\lambda}_1)$	---	-0.82	-0.82
	RESULTS FROM 100 SAMPLE STATISTICS		
Var $\hat{\lambda}_o$	0.000079	0.000080	0.000080
Var $\hat{\lambda}_1$	0.012317	0.009533	0.009533
Cov$(\hat{\lambda}_o, \hat{\lambda}_1)$	-0.000752	-0.000719	-0.000719
Corr$(\hat{\lambda}_o, \hat{\lambda}_1)$	-0.76	-0.82	-0.82

4) Histograms of $\hat{\lambda}_o$ and $\hat{\lambda}_1$ appeared normally distributed; however, the sample statistics are not independently distributed in this Monte Carlo study.

5) In one sample, the Newton-Raphson method resulted in a matrix which was not positive definite. That sample was removed from the comparisons. No similar problem existed with the scoring method.

CHAPTER IV

ANALYSIS OF FAILURE TIMES IN THE PRESENCE OF COMPETING RISKS

4.1 INTRODUCTION

The failure of an individual may be due to one of several distinct types or causes. In clinical trials the death of a patient may be due to a cause other than the disease for which he is under study. For example, a patient who is under study for prostate cancer may be involved in a fatal accident or he may succumb to a heart attack. These other risks, as well as the risk of death due to the disease under study are called competing risks; since it is assumed that the death of the patient comes from only a single underlying cause, these risks thus compete for the patient's life.

Distinct problems in the analysis of failure times with competing causes of failure include the estimation of treatment or exposure effects on specific failure types, the study of inter-relations among failure types, and the estimation of failure rates for some causes given the removal of certain other failure types. The usual formulation of these problems is in terms of conceptual or latent failure times for each failure type. Kalbfleisch and Prentice (1980) criticized this approach on the basis of unwarranted assumptions, lack of physical interpretation and identifiability problems. They

proposed an alternative approach utilizing cause-specific hazard functions for observable quantities, including time dependent covariates.

The term "competing risk problem" has come to encompass the study of any failure process in which there is more than one distinct cause or type of failure. Frequent reference is made to possible influences of competing failure types in studies reported in the clinical, epidemiologic, demographic, and industrial literature. Available data on each study subject typically include $T \geq 0$, the time of failure, which may be right censored and $J \in \{1,2,\ldots,k\}$, the type of failure, which will be unknown if T is censored. A regression vector $\underset{\sim}{Z} = (Z_1,\ldots,Z_p)$ may also be available to record characteristics of the study subject as well as treatment allocations or exposure levels. Some components of $\underset{\sim}{Z}$ may be time-dependent, that is, $\underset{\sim}{Z} = \underset{\sim}{Z}(t)$, as occurs when successive measurements are taken on study subjects as they are followed over time.

Three distinct problems arise in the analysis of failure times with competing risks. These are (1) inference on the effects of treatment, exposure or other regression variables on specific types of failure; (2) the study of interrelations among failure types under specified study conditions and (3) the estimation of failure rates for some causes given the removal of some or all other causes.

Problem (1) arose, for example, in the University Group Diabetes Project [see Kalbfleisch and Prentice (1980),p.164]. The fact that one treatment, Tolbutamide, appeared to give rise to greater cardiovascular mortality was the primary point of interest and controversy in the study. Problems (2) and (3) both involve inference on the relationship between failure types. Problem (2) is concerned with such relationships under actual study conditions, while to address problem (3) it is necessary to extrapolate to an altered situation in which some failure types are no longer operative.

Kalbfleisch and Prentice (1980) emphasize that problem (3) is the classical problem of competing risk analysis. They contend that to put forward solutions to problems of this type one must assume that data under one set of study conditions in which K failure types are operative are somehow relevant to different sets of study conditions in which only certain of the failure modes can occur. The failure rate function for a specific type may be affected in a variety of ways by "removal" of their failure types. "Removal", itself, may involve different mechanisms, with corresponding different structures for the remaining failure types. For this reason, the authors consider problem (3) to be nonstatistical. It is therefore unrealistic to think that general statistical methods can be developed to estimate failure rates under the removal of other causes. However, a good knowledge of the physical or biological mechanisms giving rise to the removal of certain failure types is

necessary before reasonable methods can be proposed in any given setting.

In Section 4.2, we consider the general model of competing risks when all lifetimes and associated causes of failure are known, as well as problems associated with it. We also discuss the additive, noninteractive model which is useful in analyzing the survival data of patients who have undergone surgery to remedy a serious problem. In both models, we assume independent causes of failure. We briefly discuss the cause-specific hazard functions proposed by Prentice et al (1978) in Section 4.3.

4.2 MODELS FOR THE ESTIMATION OF COMPETITING RISKS

4.2.1 General Model When All Lifetimes and Associated Causes of Failure Are Known

Underlying most of the applications of the concept of competing risk is the postulate of a population of objects, a random sample of which is observed over some specified period of time. Most models proposed for human or animal populations postulate the existence of a number of independently operating causes or risks of death. Each individual is exposed to the risk of death from any of the causes. The assumption that risks are mutually independent while perhaps less tenable than any others which are made, is an essential component of all generally useful procedures so far developed.

The assumption implies that the risk of death from one cause is independent of and unaffected by changes in the risk of death from other causes. The patent falsity of this assumption when incorporated in models for populations of living organisms is recognized by many authors, but in theoretical or mathematical developments the manifestations of the premise are often obscured and their importance is understressed. Chiang (1961(b)) has said, "The characteristic of a mortality study is that the basic event, the death of an individual, is not repetitive." An equally important, and not unrelated, characteristic is that measurements of times of death of individuals in a study are naturally ordered.

The most comprehensive mathematical treatment of mortality analysis is contained in a series of three papers by Chiang (1960(a),(b); 1961(a)), which provide basic models for many of the estimation problems encountered in studies of human populations. Within this broad category, another class of problems has been generated by the need of biomedical scientists for methods which facilitate the explanation of morbidity and mortality phenomena in terms of the actions of several disease processes operating simultaneously. The variety and complexity of problems in this area have given rise to a collection of literature including those of Chiang (1961(b)), Cornfield (1957), Kimball (1958), Berkson and Elvebach (1960) Moeschberger and David (1971) and Prentice et al. (1978). Although somewhat different in scope and purpose, this class of reports deals with

the fundamental problem of estimating the probability of death from a given cause or group of causes utilizing data obtained from individuals who are simultaneously exposed to other causes of death. In many cases, these other causes or risks are peripheral to the central focus of the study, but they may also include diseases of interest to the investigator. The Berkson and Elveback (1960) report illustrates both situations. In assessing the effect of smoking on the probability of death from lung cancer, adjustments were made for all other diseases including coronary disease. A similar calculation was made for all other diseases with adjustments for lung cancer and other diseases, because it was felt that it is also important to isolate the effect of smoking on coronary disease. The major purpose in studies of this kind is to describe the pattern of a disease, in relation to some environmental factor such as smoking, as it would be if that pattern were not dependent on other competing diseases. This problem has been discussed fully by Kimball (1969). In this section, we consider a general model when all lifetime and associated causes of failure are known.

Consider an individual who is exposed to several potential causes of failure during his or her life-time. Let there be a finite number of independent causes of failure, labelled $1,2,\ldots,J$. We associate with cause k a non-negative random variable $X^{(k)}$ with a continuous c.d.f., $F^{(k)}(x)$, $k = 1,\ldots,J$. Then, the observed failure time is given by the random variable

$$T = \min(X^{(1)},\ldots,X^{(J)}) \qquad (4.2.1.1)$$

Let N be a random index for which $T = X^{(N)}$. Then, we define

$$^{*}G^{(k)}(x) = \Pr\{\text{Failure is due to cause } k \text{ and occurs on or before } x\}$$
$$= \Pr\{T \leq x,\ N = k\}.$$

The derivative of $G^{(k)}x$ with respect to x is denoted by $g^{(k)}x$. Furthermore, let

$$F(x) = \Pr\{\text{failure occurs on or before time } x\}$$

and the corresponding p.d.f. is $f(x)$. The survivorship function of the random variable T is defined as

$$S(x) = \Pr\{T > x\} = \prod_{k=1}^{J} S^{(k)}(x) \qquad (4.2.1.2)$$

where

$$S^{(k)}(x) = 1 - F^{(k)}(x).$$

On using the relation

$$f(x) = \frac{d}{dx} F(x) = -\frac{d}{dx} S(x)$$

we get

$$f(x) = \sum_{k} f^{(k)}(x) \prod_{j \neq k} S^{(j)}(x) = \sum_{k} g^{(k)}(x) \qquad (4.2.1.3)$$

---

*Note that $G^{(k)}(x)$ is not a proper distribution function.

Finally, we can write, $f(x) = \sum_k g^{(k)}(x)$ by using the relation

$$\prod_{k=1}^{J} S^{(k)}(x) = 1 - \sum_k G^{(k)}(x) \qquad (4.2.1.4)$$

The hazard rate is given by

$$\phi^{(k)}(x) = -\frac{d}{dx} \log S^{(k)}(x) = \frac{f^{(k)}(x)}{S^{(k)}(x)} \qquad (4.2.1.5)$$

Hence, we have

$$f^{(k)}(x) = \phi^{(k)}(x)\ S^{(k)}(x)$$

That is

$$f^{(k)}(x) = \frac{g^{(k)}(x)}{S(x)} \exp\{-\int_0^x \frac{g^{(k)}(x)}{S(x)}\, dx\}$$

or

$$F^{(k)}(x) = 1 - \exp\{-\int_0^x \frac{g^{(k)}(x)}{1 - \sum_k G^{(k)}(x)}\, dx\} \qquad (4.2.1.6)$$

Thus the set of functions $\{G^{(k)}(x)\}$ is related to the set $\{F^{(k)}(x)\}$ by functional equations

$$G^{(k)}(x) = \int_0^x f^{(k)}(x) \prod_{j \neq k} S^{(j)}(x)\, dx \qquad (4.2.1.7)$$

The solution of this set of equations is

$$F^{(k)}(x) = 1 - \exp\{-\int_0^x \frac{g^{(k)}(x)dx}{1 - \sum_k G^{(k)}(x)}\}$$

which is a bona-fide c.d.f. provided the above-mentioned integral diverges. This requirement appears to have been overlooked by Cox (1959). Berman (1963) gives an elegant derivation of the above-mentioned result. If we impose some well-known parametric forms of the underlying distributions $F^{(k)}(x)$ we could develop maximum likelihood estimates for the parameters of the distributions employing the methods described in Section 2.2.

At this point, we should point out that there is a basic question raised on the assumption of "latent" or "potential" failure times. Cox (1959) and Moeschberger and David (1971), among others, define $X^{(k)}$ to be the time of failure from cause k that would be observed if the possibility of failure from causes other than k were removed. They further assume that the observed $T = \min(X^{(1)},\ldots,X^{(J)})$. While this point of view ascribes a physical meaning to the latent failure times, it involves the very strong assumption that the time of failure from cause k under one set of study conditions in which all J causes are operative is precisely the same as under an altered set of conditions in which all causes except the $k^{\underline{th}}$ have been removed. Such an assumption may be reasonable in very special situations, such as in clinical trials in which failure types occur in organs of an individual that are physically and functionally, as well as statistically, independent. More generally, however, the elimination of certain failure types may well alter the risks of other types of failures.

However, any assumption about the relationship between the observed T and the times to failure for specific causes, given the removal of other causes will require detailed knowledge of the system under study and of the mechanism for cause removal.

Another problem is the entire idea of limiting our attention to death only. In the study of morbidity and mortality pattern in a population we must recognize the fact that the death of an individual is usually preceded by an illness (condition, disorder). It is not realistic to speak of a person's chance of dying from tuberclosis when he is not even affected with the disease. Also, competition of risks of death depends on the health condition of an individual: a person affected with a disease (say, cardiovascular-renal (CVR) diseases) probably has a probability of dying of a second disease different from a person who is not affected with CVR. Therefore, a mortality study is incomplete unless illness is taken into consideration. Illness and death are distinct and different types of events. Illness are potentially concurrent, repetitive, and reversible, whereas death is an irreversible or absorbing state. The study of illness adds a new dimension and a new complexity to the general problem of mortality, but it makes the underlying assumption regarding mortality intensity functions more realistic and more reasonable.

4.2.2 The additive non-interactive model

The survival time for patients who have had surgery for invasive cancer of the bladder is their length of survival from surgery. This is a serious disease, and most patients will die from the cancer; however, some patients may die of other causes (e.g., heart failure or accidents). There are alternate ways to analyze such survival data. First, we can look at the patient survival times separately for each cause of death. Second, we can ignore the fact that several risks are competing for each patient's life. In this case, we analyze all survival times together.

The term "risk" refers to the probability of dying from a given cause prior to death. After death, the risk is the cause that was responsible for death. Suppose there are k independent competing risks such that the hazard rate of the individual due to the i^{th} risk at time t is $\lambda_i(t), i=1,2,\ldots,k$. If $S(t)$ is the overall survival probability of an individual at time t, the k risks are assumed to be independent. The survival probability at time t for the i^{th} risk is

$$S_i(t) = \exp\left(-\int_0^t \lambda_i(u)du\right) \qquad (4.2.2.1)$$

it then follows that

$$S(t) = \prod_{i=1}^{k} S_i(t) = \prod_{i=1}^{k} \exp\left(-\int_0^t \lambda_i(u)du\right) = \exp\left(-\sum_{i=1}^{k}\int_0^t \lambda_i(u)du\right) \qquad (4.2.2.2)$$

If we define $\lambda(t)$ as the overall hazard rate, that is, $\lambda(t)$ is defined by the integral equation

$$S(t) = \exp(-\int_0^t \lambda(u)du) \qquad (4.2.2.3)$$

then, from Equations (4.2.2.2) and (4.2.2.3) and certain elementary rules of calculus we obtain

$$\lambda(t) = \sum_{i=1}^{k} \lambda_i(t) \qquad (4.2.2.4)$$

The competing risk model defined by Equation (4.2.2.4) is called the additive, non-interactive model of cumulative risk and is discussed in detail by Berkson and Elvebach (1960), and Chiang (1968).

Since Equation (4.2.2.2) described the survival probability for the additive, non-interactive competing risk model, we can by elementary considerations, derive both $f(t)$ and $F(t)$, the death density and cumulative death distribution functions, respectively. Hence moments of $f(t)$ and the maximum likelihood estimators of the parameters can be obtained.

It may be important to isolate the probabilities an individual dies from the $i^{\underline{th}}$ risk, $i=1,2,\ldots,k$. We thus consider the following. The probability an individual dies in the interval $t < x < t + \Delta t$ of risk i is the probability he survives of all risks up to time t and dies of risk i in the interval. Mathematically, this probability is $S(t)\lambda_i(t)\Delta t$. If we add up these probabilities over the entire time axis and let

$\Delta t \to 0$, we have Q_i, the probability of death due to the i^{th} cause, given by

$$Q_i = \int_0^\infty S(t))\lambda_i(t)dt$$

$$= \int_0^\infty \exp(-\sum_{j=1}^{k} \int_0^t \lambda_j(u)du)\lambda_i(t)dt, \quad i=1,2,\ldots,k$$

(4.2.2.5)

The probability an individual dies in the interval $t < x < t + \Delta t$, given risk i, is the joint probability he lives to time t and dies in $t < x < t + \Delta t$ of risk i, divided by the probability he dies of risk i. Mathematically, this probability is

$$f(t|\text{risk } i)\Delta t = \frac{S(t)\ \lambda_i(t)\Delta t}{\int_0^\infty S(t)\lambda_i(t)dt} \qquad (4.2.2.6)$$

Again, letting $\Delta t \to 0$, the death density function of those persons dying, given risk i, is

$$f(t|\text{risk } i) = \frac{S(t)\lambda_i(t)}{\int_0^\infty S(t)\lambda_i(t)dt}, \quad i=1,2,\ldots,k \qquad (4.2.2.7)$$

Let us consider now a situation for which the hazard rates for the k risks are proportional. That is, $\lambda_i(t) = c_i\lambda(t)$, where $c_i > 0$ is independent of $t, i=1,2,\ldots,k$. It can be shown that Q_i, the probability of death due to risk i, is

$$Q_i = \frac{c_i}{\sum_{j=1}^{k} c_j} \qquad (4.2.2.8)$$

To show this, let

$$y = \sum_{j=1}^{k} \int_0^t \lambda_j(u)\,du = \sum_{j=1}^{k} c_j \int_0^t \lambda(u)\,du$$

Hence

$$dy = \sum_{j=1}^{k} c_j \lambda(t)\,dt$$

Thus

$$Q_i = \int_0^\infty \exp\left[\left(-\sum_{j=1}^{k} c_j\right) \int_0^t \lambda(u)\,du\right] c_i \lambda(t)\,dt$$

$$= \frac{c_i}{\sum_{j=1}^{k} c_j} \int_0^\infty e^{-y}\,dy = \frac{c_i}{\sum_{j=1}^{k} c_j} \qquad (4.2.2.9)$$

The death density function given the $i^{\underline{th}}$ risk is then

$$f(t|\text{risk } i) = c\, S(t)\, \lambda(t) \qquad (4.2.2.10)$$

where

$$c = \sum_{i=1}^{k} c_i$$

Hence, the death density function given the $i^{\underline{th}}$ risk, does not depend on the risk.

We shall now consider the problem of estimating c_i, $i=1,2,\ldots,k$, when all hazard rates are constant with respect to time. The assumption of constant hazard rates for all k risks implies that they are proportional (but not conversely). When the hazard rates are all constant, there is no loss of generality in assuming

$$\lambda_i(t) = c_i \qquad i=1,2,\ldots,k.$$

If t_{ij} is the survival time of the $j^{\underline{th}}$ individual who dies from the $i^{\underline{th}}$ risk, $j=1,2,\ldots,n_i$; $i=1,2,\ldots,k$, the likelihood of the sample is

$$L(c_1,\ldots,c_k) = \prod_{i=1}^{k}\prod_{j=1}^{n_i} S(t_{ij})c_i$$

$$= \exp\left[-\sum_{i=1}^{k}\sum_{j=1}^{n_i} c_i t_{ij}\right]\prod_{i=1}^{k} c_i^{n_i} \qquad (4.2.2.11)$$

Thus

$$\log L = -\sum_{i=1}^{k}\sum_{j=1}^{n_i} c_i\, t_{ij} + \sum_{i=1}^{k} n_i \log c_i \qquad (4.2.2.12)$$

It follows that $\hat{c}_i$, the ML estimator of c_i, for the $i^{\underline{th}}$ risk is

$$\hat{c}_i = \frac{n_i}{\sum_{j=1}^{n_i} t_{ij}} \qquad i=1,2,\ldots,k$$

For large samples, the vector $\hat{\underset{\sim}{c}} = (\hat{c}_1,\ldots,\hat{c}_k)$ has an approximate normal distribution with mean $\underset{\sim}{c} = (c_1,\ldots,c_k)$ and variance covariance matrix V, where

$$V^{-1} = \begin{bmatrix} \sigma_{11} & \cdots\cdots & \sigma_{1k} \\ \vdots & & \vdots \\ \sigma_{k1} & \cdots\cdots & \sigma_{kk} \end{bmatrix}$$

and

$$\sigma_{ij} = -E\left[\frac{\partial^2 \log L}{\partial c_i \partial c_j}\right], \quad i,j=1,2,\ldots,k$$

Differentiating Equation (4.2.2.12) with respect to i and then j, we find

$$-\frac{\partial^2 \log L}{\partial c_i^2} = \frac{n_i}{c_i^2}$$

and

$$-\frac{\partial^2 \log L}{\partial c_i \partial c_j} = 0, \quad i\neq j; i,j=1,2,\ldots,k$$

The $n_1,n_2,\ldots,n_k$ have a joint multinomial probability distribution, where

$$E(n_i) = NQ_i, N = \sum_{j=1}^{k} n_j, \quad Q_i = \frac{c_i}{\sum_{j=1}^{k} c_j}, \quad i=1,2,\ldots,k$$

Thus

$$\sigma_{ii} = C \frac{n_i}{c_i}$$

and

$$\sigma_{ij} = 0 \qquad i \neq j; i,j = 1,\ldots,k,$$

where

$$C = \sum_{i=1}^{k} c_i$$

Thus we have shown for large samples $\hat{c}_1,\ldots,\hat{c}_k$ are approximately independently normally distributed variables with means $c_1,c_2,\ldots,c_k$ and variances $c_1/Cn_1,\ldots,c_k/Cn_k$, respectively.

4.3 CAUSE-SPECIFIC HAZARD FUNCTIONS AND THE LIKELIHOOD FUNCTION

A statistical model for competiting risks data involves a specification of the distribution for the observable quantities $(T,J,\underset{\sim}{Z})$. As an example (Hoel, 1972), T may be the age of death in a mouse radiation study, J may describe the cause of death as thymic lymphoma, reticulum cell sarcoma, or other, and the regression variable may indicate whether or not each particular mouse was kept in germ-free isolation. Usually,it will be sufficient to specify a model for (T,J) given $\underset{\sim}{Z}$. We discuss such models in terms of cause-specific hazard functions.

Suppose failure time T is continuous. The overall failure rate or hazard function for an individual with regression vector $\underset{\sim}{Z}$ is given by

$$\lambda(t,\underset{\sim}{Z}) = \lim_{\Delta t \to 0} \frac{\Pr\{t \leq T < t + \Delta t | T \geq t;\ \underset{\sim}{Z}(t)\}}{\Delta t} \qquad (4.3.1)$$

where $\underset{\sim}{Z}(t)$ denotes the value of the regression vector at time t [Cox(1972)]. Cause-specific hazard functions [Chiang, 1968,p.244, and Prentice and Breslow(1978)] are defined by

$$\lambda_j(t;\underset{\sim}{Z}) = \lim_{\Delta t \to 0} \frac{\Pr\{t \leq T < t + \Delta t, J=j | T \geq t; \underset{\sim}{Z}(t)\}}{\Delta t} \qquad (4.3.2)$$

for $j=1,\ldots,m$. The function $\lambda_j(t;\underset{\sim}{Z})$ simply gives the instantaneous failure rate from cause j at time t, given the regression vector $\underset{\sim}{Z}(t)$, in the presence of the other failure types. Assuming distinct failure types, the overall hazard function can be expressed in terms of cause-specific hazard function as

$$\lambda(t;\underset{\sim}{Z}) = \sum_{1}^{m} \lambda_j(t;\underset{\sim}{Z}) \qquad (4.3.3)$$

The overall survivor function can be written as

$$F(t;\underset{\sim}{Z}^*) = \exp\{ - \int_0^t \lambda(u;\underset{\sim}{Z})du\} \quad , \qquad (4.3.4)$$

and the probability function for time to failure and cause of failure is

$$f_j(t;Z^*) = \lim_{\Delta t \to 0} \frac{\Pr\{t \leq T < t + \Delta t, J=j | \underset{\sim}{Z}\}}{\Delta t}$$

$$= \lambda_j(t;\underset{\sim}{Z})F(t;\underset{\sim}{Z}^*), j=1,\ldots,m \qquad (4.3.5)$$

where

$$\underset{\sim}{Z}^* = \underset{\sim}{Z}^*(t) \text{ denotes } \{\underset{\sim}{Z}(u); u \leq t\}.$$

We should note that Equations(4.3.3) to (4.3.5) indicate that the likelihood function can be written in terms of the cause-specific hazard functions.

Suppose now that n study subjects give rise to data $(t_i, j_i, \delta_i, \underset{\sim}{Z}_i^*), i=1,\ldots,n,$ where t_i is the failure time, j_i is the cause of failure, δ_i is a censoring indicator, and $\underset{\sim}{Z}_i^* = \underset{\sim}{Z}_i^*(t_i)$ is a vector-valued regression function for the $i^{\underline{th}}$ study subject. The censoring indicator takes value one if failure occurs and value zero otherwise. The cause of failure j_i may be specified arbitrarily if $\delta_i = 0$. As usual, an independent censoring mechanism will be assumed. This means that at any fixed $\{t, \underset{\sim}{Z}(t)\}$ individuals are not selectively censored on the basis of a relatively good or relatively poor prognosis. This condition is met by the usual censoring schemes such as fixed time censoring (Type I), independent random censoring, order statistic censoring (Type II), as well as by more general censoring schemes in which censorship at $\{t; \underset{\sim}{Z}(t)\}$ depends arbitrarily on the previous number of failures and censorings.

The likelihood function under an independent censoring mechanism is, up to proportionality

$$\prod_{i=1}^{n} \{[\lambda_{j_i}(t_i;\underset{\sim}{Z}_i)]^{\delta_i} F(t_i;\underset{\sim}{Z}_i^*)\} = \left(\prod_{i=1}^{n} [\lambda_{j_i}(t_i;\underset{\sim}{Z}_i)]^{\delta_i}\right) \prod_{j=1}^{m} \exp\{ - \int_0^{t_i} \lambda_j[u;\underset{\sim}{Z}(u)]du\} \qquad (4.3.6)$$

We note that the likelihood function is completely specified by the cause-specific hazard functions $\lambda_j(t,\underset{\sim}{Z}), j=1,\ldots,m$. We further note that upon rearrangement the likelihood factors into a component for each j. In fact, the likelihood factor for $\lambda_j(t;\underset{\sim}{Z})$ is precisely the same as would be obtained by regarding all failures from causes other than j as censored at their time of failure. This provides a formal justification, at least for the estimation of $\lambda_j(t;\underset{\sim}{Z})$, for the common procedure of regarding failures from other causes as censored when studying factors that effect a certain failure type. Also, the likelihood factorization along with standard survival data techniques make it clear that the $\lambda_j(t;\underset{\sim}{Z})$ functions are identifiable; that is, the cause-specific hazard functions have the potential to be directly estimated from data of the form $(t,j,\delta,\underset{\sim}{Z}^*)$.

The above likelihood development implicitly assumes that the covariate functions $\underset{\sim}{Z}^*(t)$ are deterministic or are generated by a stochastic mechanism external to the sample. In such circumstances, a survivor function for t given $\underset{\sim}{Z}^*(t)$, for example, has clear meaning. More generally, however, it is necessary to consider the likelihood based on the joint distri-

bution $\{T,J\}$ and $\underset{\sim}{Z}^*(T)$ which involves additional factors of type $P\{\underset{\sim}{Z}(t)|T \geq t;\underset{\sim}{Z}(u),u < t\}$. It is still appropriate to use Equation (4.3.6) for inference on the $\lambda_j(t,\underset{\sim}{Z})$ functions, though Equation (4.3.6) is properly referred to as partial rather than an ordinary likelihood [Cox (1975)]. Kalbfleisch and Prentice (1980) have discussed this matter in much detail.

CHAPTER V

REGRESSION MODELS

5.1 INTRODUCTION

In clinical, and other experimental trials, measurements of further characters beyond just time and mortality are often obtained. Some, if not all, of these may be expected to have some association with failure rates. We recall from Chapter I that the variables corresponding to such characters are called concomitant variables.

Consider a clinical trial of patients with an acute illness such as angina pectoris. Patients with this disease are frequently monitored to obtain readings on blood pressure, cardiogram, weight, age and other variables. Thus at a given point during the study we would have the following information for each patient: (1) his survival time, withdrawal or censored time (if he is still on trial); (2) his blood pressure reading; (3) his cardiogram results; (4) his current weight, and (5) his age. Within the clinical trial there may be two or more treatment groups - To test the null hypothesis that all treatment are equal in terms of survival times, it would be necessary to determine how the concomitant measurements (2) through (5) affect the survival time and to adjust these according to the null hypothesis. Because concomitant information on a patient's condition often accompanies survival time,

regression models have been introduced in recent years in analyzing survival data. Thus in this Chapter, survival models which incorporate a regression variable are considered with a view to estimating the regression parameters.

In Section 5.2, we shall consider a generalization of these models that take account of concomitant information on the individuals sampled. That is, we shall give a general parametric model of hazard function with observed covariates.

We shall discuss in Section 5.3, the general additive hazard rate as applied to the case when one concomitant variable is present.

A problem frequently encountered in survival studies in medical trials and elsewhere is that of incorporating explanatory variables into the model in order to make fair treatment comparisons or to describe survival behaviour. Recently Cox (1972) proposed a model for incorporating a vector of covariates $\underset{\sim}{Z}' = (Z_1,\ldots,Z_p)$ in which the survival time t has probability density function

$$\lambda_o(t)\, e^{\underset{\sim}{\beta}\underset{\sim}{Z}} \exp\{-\int_o^t \lambda_o(u)\,du\; e^{\underset{\sim}{\beta}\underset{\sim}{Z}}\} \qquad (t>0)$$

where

$\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_p)$ is a vector of unknown parameters.

The hazard at $\underset{\sim}{Z} = \underset{\sim}{0}$, $\lambda_0(t)$ is left arbitrary, so that this model is largely non-parametric. We shall consider this model in Section 5.4, in detail.

5.2 THE GENERAL MODEL

Let $\underset{\sim}{Z}' = (Z_1, \ldots, Z_p)$ denote a $1 \times p$ vector of covariates. Let us denote the hazard rate by $\lambda(t;\underset{\sim}{Z})$. However, this hazard rate may also depend on some parameters, $\underset{\sim}{\beta}' = (\beta_1, \ldots, \beta_s)$ so that, in effect we have

$$\lambda(t;\underset{\sim}{Z}) = \lambda(t;Z;\underset{\sim}{\beta}) \tag{5.2.1}$$

We shall write $\lambda(t;\underset{\sim}{Z})$ where no ambiguity arises. Observe that we must have $\lambda(t;\underset{\sim}{Z}) \geq 0$.

The cumulative hazard function (c.h.f.) is

$$\Lambda(t;\underset{\sim}{Z}) = \int_0^t \lambda(u;\underset{\sim}{Z})du \tag{5.2.2}$$

The survival function, $S(t;\underset{\sim}{Z})$ and the death density function are respectively given by

$$S(t;\underset{\sim}{Z}) = \exp[-\Lambda(t;\underset{\sim}{Z})] \tag{5.2.3}$$

and

$$f(t;\underset{\sim}{Z}) = \lambda(t;\underset{\sim}{Z})S(t;\underset{\sim}{Z}) \tag{5.2.4}$$

Let $\underset{\sim}{Z}_k' = (Z_{1k},\ldots,Z_{pk})$ denote the vector of observed concomitant variables for individual k. For example, the vector $\underset{\sim}{Z}_k'$ can represent a set of measurements taken at entry of individual k, or a set of average measurements over the course of study. However, once they have been observed, they represent a fixed set of values. Let us first assume that the Z's do not depend on time t. Thus for each individual k we define his own hazard rate by $\lambda(t;\underset{\sim}{Z}_k)$ and survival function by $S(t;\underset{\sim}{Z}_k)$.

We can find the likelihood function under this model if we consider a sample of n individuals taking part in a clinical trial at some time or other in the study.

Let τ_k be the time at which an individual k entered the study, and t_k the time at which he was last observed. Let D be the set of d individuals dying; thus we have (n-d) individuals who were alive when last observed; denote this set by $\bar{D}$. The likelihood function for this sample then is

$$L = \prod_{k\in D}\left[\frac{f(t_k;\underset{\sim}{Z}_k)}{S(\tau_k;\underset{\sim}{Z}_k)}\right] \times \prod_{\ell\in\bar{D}}\left[\frac{S(t_\ell;\underset{\sim}{Z}_\ell)}{S(\tau_\ell;\underset{\sim}{Z}_\ell)}\right]$$

$$= \prod_{k\in D}\left[\lambda(t_k;\underset{\sim}{Z}_k)\right] \times \prod_{i=1}^{n}\left[\frac{S(t_i;\underset{\sim}{Z}_i)}{S(t_i;\underset{\sim}{Z}_i)}\right] \qquad (5.2.5)$$

We can rewrite Equation (5.2.5) in a different form, if we let

$$\delta_k = \begin{cases} 1 & \text{if } k \text{ died at } t_k \\ 0 & \text{if } k \text{ was alive at } t_k \end{cases} \tag{5.2.6}$$

Then Equation (5.2.5) becomes

$$L = \prod_{k=1}^{n} [\lambda(t_k;\underset{\sim}{Z}_k)]^{\delta_k} \frac{S(t_k;\underset{\sim}{Z}_k)}{S(\tau_k;\underset{\sim}{Z}_k)} \tag{5.2.7}$$

and

$$\log L = \sum_{k=1}^{n} [\delta_k \log\lambda(t_k;\underset{\sim}{Z}_k) + \log S(t_k;\underset{\sim}{Z}_k) - \log S(\tau_k;\underset{\sim}{Z}_k)]$$

Note that when we have no new entries, for example, when n individuals are followed up from t=0, then $\tau_k=0$ and so $S(\tau_k;\underset{\sim}{Z}_k) = 1 \quad \forall \quad K$.

Such situations are very common in clinical trials and when they do occur, the likelihood, Equation (5.2.7) takes the form

$$L = \prod_{k=1}^{n} [\lambda(t_k;\underset{\sim}{Z}_k)]^{\delta_k} S(t_k;\underset{\sim}{Z}_k) \tag{5.2.8}$$

When we know the parametric form of the hazard function $\lambda(t;\underset{\sim}{Z})$, their maximum likelihood estimators can be obtained by maximizing Equation (5.2.8).

Example:

A common method of incorporating concomitant variables into the model, is to express the parameters of the model as a simple function of these variables. An example of this is given by Bailey et al (1977) where they have used a hazard rate model of the form

$$\lambda(t) = \alpha\exp(-\gamma t) + \delta \qquad (5.2.9)$$

where

$t > 0$ is the time elapsed since graft, and

$\alpha > 0$, $\gamma > 0$ and $\delta > 0$ are the parameters to analyze data on survival of 748 kidney transplant case.

The authors introduced two concomitant variables:

$$Z_1 = \begin{cases} 1 & \text{if male} \\ 2 & \text{if female} \end{cases}$$

Z_2 = age (in years) of the recipients.

The parameters α, γ, and δ in Equation (5.2.9) were replaced by

$$\alpha = \exp(\alpha_o+\alpha_1 Z_1+\alpha_2 Z_2)$$

$$\gamma = \exp(\gamma_o+\gamma_1 Z_1+\gamma_2 Z_2) \qquad (5.2.10)$$

$$\delta = \exp(\delta_o+\delta_1 Z_1+\delta_2 Z_2)$$

Estimates for the nine parameters were obtained using the maximum likelihood procedures.

There are different expressions of the general model but we shall not go into detail here. We give one form of it below:

Suppose instead of one group we have k groups (levels of a factor, treatments) in the experiment. Then, in the general case, the hazard rate function for each group might be of different form.

For the $j^{\underline{th}}$ group, we define

$$\lambda_j(t;\underset{\sim}{Z}) = \lambda_j(t;\underset{\sim}{Z};\underset{\sim}{\beta}_j) \quad j=1,2,..,k \qquad (5.2.11)$$

Let (jk) denote the $k^{\underline{th}}$ individual in the $j^{\underline{th}}$ group,

t_{jk} be the time at which (jk) is last observed, and

$\underset{\sim}{Z}_{jk} = (Z_{1jk},\ldots,Z_{pjk})$ be the $1 \times p$ vector of covariates.

Define

$$\delta_{jk} = \begin{cases} 1 & \text{if } (jk) \text{ died at } t_{jk} \\ 0 & \text{if } (jk) \text{ was alive at } t_{jk} \end{cases} \qquad (5.2.12)$$

Let n_j be the number of individuals followed-up from $t = 0$ in the $j^{\underline{th}}$ subgroup with $n_1 + n_2 + \ldots + n_g = n$. The likelihood function is

$$L = \prod_{j=1}^{g} \prod_{k=1}^{n_j} [\lambda_j(t_{jk}; \underset{\sim}{Z}_{jk})]^{\delta_{jk}} S_j(t_{jk}; \underset{\sim}{Z}_{jk}) \qquad (5.2.13)$$

where

$$S_j(t_{jk}; \underset{\sim}{Z}_{jk}) = \exp[-\int_0^{t_{jk}} \lambda_j(t; \underset{\sim}{Z}_{jk})\, dt]$$

Kay (1977) discusses this model.

5.3 HAZARD RATE FUNCTION (ADDITIVE MODEL)

For the last few years two types of hazard rate have attracted considerable attention. They are the Additive and the Multiplicative models. Probably their attractiveness may be due to their mathematical simplicity. One type of additive model is given by Fiegl and Zelen (1965). An example of the multiplicative model is the exponential-type hazard functions discussed by Cox (1972) which we consider in the next section.

Generally, the additive model is defined as

$$\lambda(t;\underset{\sim}{Z}) = \lambda(t) + \sum_{i=1}^{p} h_i(t) g_i(Z_i) \qquad (5.3.1)$$

where

$\lambda(t)$ is the underlying hazard rate.

we note that $h_i(t)$'s are functions of t only (in special cases they can be polynomials in t), while the functions $g_i(Z_i) = g_i(Z_i;\underset{\sim}{\beta}_i), i=1,\ldots,p)$ do not depend on t.

By appropriate definition of Z_i's (for example, $Z_i = \log x_i$, where x_i is the original measurement), we can represent Equation (5.3.1) as a linear function of Z_i's, that is

$$\lambda(t;\underset{\sim}{Z}) = \lambda(t) + \sum_{i=1}^{p} h_i(t) Z_i \qquad (5.3.2)$$

When the hazard rate does not vary with time then

$$\lambda(t;\underset{\sim}{Z}) = \beta_o + \sum_{i=1}^{p} \beta_i Z_i = \sum_{i=0}^{p} \beta_i Z_i = \underset{\sim}{\beta}'\underset{\sim}{Z} \qquad (5.3.3)$$

where

$Z_o = 1$ is a dummy variable.

One should note that the values of the β's must be such that the condition $\underset{\sim}{\beta}'\underset{\sim}{Z} \geq 0$ is satisfied.

For any given set of Z_i's, the linear function on the right-hand side of Equation (5.3.3) gives a constant value; thus as discussed earlier, the survival distribution is the exponential. Because of this, survival models with hazard rate defined by Equation (5.3.3) are referred to as linear-exponential.

For the k^{th} individual we have

$$\lambda(t; \underset{\sim}{Z}_k) = \sum_{i=0}^{p} \beta_i Z_{ik} \qquad (5.3.4)$$

The likelihood function for a random sample of size n is

$$L = \prod_{k=1}^{n} \{ (\sum_{i=0}^{p} \beta_i Z_{ik})^{\delta_k} \exp[-(\sum_{i=0}^{p} \beta_i Z_{ik}) t_k] \} \qquad (5.3.5)$$

where

$$\delta_k = \begin{cases} 1 & \text{if } k \text{ died at } t_k \\ 0 & \text{if } k \text{ was alive at } t_k \end{cases}$$

$$\log L = \sum_{k=1}^{n} [\delta_k \log(\sum_{i=0}^{p} \beta_i Z_{ik}) - (\sum_{i=0}^{p} \beta_i Z_{ik}) t_k] \qquad (5.3.6)$$

This leads to $(p+1)$ maximum likelihood equations

$$\frac{\partial \log L}{\partial \beta_\ell} = \sum_{k=1}^{n} \frac{\delta_k Z_{\ell k}}{\sum_{i=0}^{p} \beta_i Z_{ik}} - \sum_{k=1}^{n} Z_{ik}\, t_k = 0, \ell=0,1,\ldots,p \quad (5.3.7)$$

Byar et al (1974) have used this model in analyzing mortality of patients with cancer of the prostate.

5.3.1 A Case When the Mean Survival Time is A Linear Function of a Concomitant Variable Z

In clinical trials, survival time is typically measured as the time from diagnosis to death or the time from the initiation of a specific treatment to death. In some studies of chronic diseases, the survival time is the only quantitative response variable available for analysis. Often in addition to the survival time, there are objective measures available which indicate the severity of the disease at the time the patient first came under observation.

For example, in leukemia the white blood count at diagnosis is a useful concomitant variable for the analysis of survival. Leukemia is a cancer characterized by an over-proliferation of white blood cells; the higher the white blood count, the more severe the disease. When predicting a leukemia patient's survival time it is realistic to make the prediction dependent on white blood count and any other variables which are indicators of the progression of the disease.

Fiegl and Zelen (1965) developed a method where survival time for each patient is assumed to have a single exponential distribution in which the parameter depends on one concomitant variable noting that the extension to more concomitant variates is relatively straightforward.

Let $t_1, t_2, \ldots, t_n$ be a sample of n independent survival times where the survival time for the i^{th} patient has the probability density function

$$f_i(t_i) = \lambda_i \exp(-\lambda_i t_i) \qquad \lambda_i > 0, t_i \geq 0 \qquad i=1,2,\ldots,n \qquad (5.3.1.1)$$

Initially, we assume that all patients are followed until death. Furthermore, let $x_1, x_2, \ldots, x_n$ be observed values of a concomitant variable such that the expected value of the survival time for the i^{th} patient is

$$E(t_i) = \frac{1}{\lambda_i} = a + bx_i \qquad (5.3.1.2)$$

Thus the mean survival time of patients is assumed to be linearly related to the concomitant variable. For example, if the variates x represent cholesterol levels of patients with acute coronary disease, we might expect small values of the concomitant variate to correspond to relatively large values of the mean survival time. Note that x_i could refer to a transformed value (e.g., $x_i = z_i^{-1}$.) The problem is to estimate the parameters a and b.

Let the likelihood function of the n survival times be written as

$$L(a,b) = \prod_{i=1}^{n} f_i(t_i) = \prod_{i=1}^{n} \lambda_i \exp(-\lambda_i t_i)$$

$$= \left(\prod_{i=1}^{n} (a+bx_i)^{-1} \right) \left(\exp - \sum_{i=1}^{n} t_i (a+bx_i)^{-1} \right) \qquad (5.3.1.3)$$

having the log likelihood function

$$\log L = - \sum_{i=1}^{n} \log(a+bx_i) - \sum_{i=1}^{n} t_i (a+bx_i)^{-1} \qquad (5.3.1.4)$$

The maximum likelihood estimators may be found from Equation (5.3.1.4) by solving the equations

$$\frac{\partial \log L}{\partial a} = 0 = - \sum_{i=1}^{n} (\hat{a}+\hat{b}x_i)^{-1} + \sum_{i=1}^{n} t_i (\hat{a}+\hat{b}x_i)^{-2} \qquad (5.3.1.5)$$

$$\frac{\partial \log L}{\partial b} = 0 = - \sum_{i=1}^{n} x_i (\hat{a}+\hat{b}x_i)^{-1} + \sum_{i=1}^{n} x_i t_i (\hat{a}+\hat{b}x_i)^{-2} \qquad (5.3.1.6)$$

using iterative methods.

This can be done by expanding the two equations (5.3.1.5) and (5.3.1.6) in a Taylor series to first-order terms and solving successive sets of pairs of simultaneous equations.

If $(\hat{a}_k, \hat{b}_k)$ denote the estimates at the $k^{\underline{th}}$ iteration, the values for the $(k+1)^{\underline{th}}$ iteration are found by solving for $\delta\hat{a}_k, \delta\hat{b}_k$ in the two simultaneous equations below in the two unknowns $\delta\hat{a}_k, \delta\hat{b}_k$

$$A_k \delta\hat{a}_k + B_k \delta\hat{b}_k = D_k$$

$$B_k \delta\hat{a}_k + C_k \delta\hat{b}_k = E_k \qquad (5.3.1.7)$$

where

$$\delta\hat{a}_k = \hat{a}_{k+1} - \hat{a}_k; \quad \delta\hat{b}_k = \hat{b}_{k+1} - \hat{b}_k$$

$$A_k = \sum_{i=1}^{n} (\hat{a}_k + \hat{b}_k x_i)^{-2} - 2 \sum_{i=1}^{n} t_i (\hat{a}_k + \hat{b}_k x_i)^{-3}$$

$$B_k = \sum_{i=1}^{n} x_i (\hat{a}_k + \hat{b}_k x_i)^{-2} - 2 \sum_{i=1}^{n} t_i x_i (\hat{a}_k + \hat{b}_k x_i)^{-3}$$

$$C_k = \sum_{i=1}^{n} x_i^2 (\hat{a}_k + \hat{b}_k x_i)^{-2} - 2 \sum_{i=1}^{n} t_i x_i^2 (\hat{a}_k + \hat{b}_k x_i)^{-3}$$

$$D_k = \sum_{i=1}^{n} (\hat{a}_k + \hat{b}_k x_i)^{-1} - \sum_{i=1}^{n} t_i (\hat{a}_k + \hat{b}_k x_i)^{-2}$$

$$E_k = \sum_{i=1}^{n} x_i (\hat{a}_k + \hat{b}_k x_i)^{-1} - \sum_{i=1}^{n} x_i t_i (\hat{a}_k + \hat{b}_k x_i)^{-2} \qquad (5.3.1.8)$$

First of all, it is necessary to obtain initial estimates $\hat{a}_o$ and $\hat{b}_o$. These estimates $(\hat{a}_o, \hat{b}_o)$ can be obtained from a straight line eye fit to a scatter plot of the n observation

points (x_i,t_i) because t_i, the single observation from the $i^{\underline{th}}$ patient has the expected value

$$E(t_i) = \frac{1}{\lambda_i} = a + bx_i$$

That is, we can take $\hat{a}_o$ and $\hat{b}_o$ by plotting t_i and x_i and estimating them visually or computing the regression equation

$$E(t_i) = a_o + b_o x_i, \quad i=1,\ldots,n \qquad (5.3.1.9)$$

where the least squares estimates $\hat{a}_o$ and $\hat{b}_o$ obtained from Equation (5.3.1.9), become the initial estimates in the iterative procedure.

Fiegl and Zelen noted that the maximum likelihood estimates may easily be obtained by calculating the log likelihood function, equation (5.3.1.4) in the neighbourhood of an initial estimate for (a,b).

One should note that the solution $(\hat{a},\hat{b})$ must be in the set $\hat{a} + \hat{b}x_i > 0$ for all x_i.

The asymptotic variance-covariance matrix for $(\hat{a},\hat{b})$ is obtained from:

$$-\begin{bmatrix} E\left(\frac{\partial^2 \log L}{\partial a^2}\right) & E\left(\frac{\partial^2 \log L}{\partial a \partial b}\right) \\ E\left(\frac{\partial^2 \log L}{\partial a \partial b}\right) & E\left(\frac{\partial^2 \log L}{\partial b^2}\right) \end{bmatrix}^{-1} = \begin{bmatrix} \text{var } \hat{a} & \text{cov}(\hat{a},\hat{b}) \\ \text{cov}(\hat{a},\hat{b}) & \text{var } \hat{b} \end{bmatrix}.$$

Carrying out the necessary algebraic manipulations

$$\text{var } \hat{a} = \Delta^{-1} \sum_{i=1}^{n} x_i^2 (a+bx_i)^{-2} \qquad (5.3.1.10)$$

$$\text{var } \hat{b} = \Delta^{-1} \sum_{i=1}^{n} (a+bx_i)^{-2} \qquad (5.3.1.11)$$

$$\text{cov}(\hat{a},\hat{b}) = -\Delta^{-1} \sum_{i=1}^{n} x_i (a+bx_i)^{-2}$$

where

$$\Delta = [\sum_{i=1}^{n} x_i^2 (a+bx_i)^{-2}][\sum_{i=1}^{n} (a+bx_i)^{-2}] - [\sum_{i=1}^{n} x_i (a+bx_i)^{-2}]^2$$

and the parameters are evaluated at $(a,b) = (\hat{a},\hat{b})$.

5.4 THE HAZARD RATE FUNCTION (MULTIPLICATIVE MODEL)

The general multiplicative model is defined as

$$\lambda(t;\underset{\sim}{Z}) = \lambda(t)g(\underset{\sim}{Z}) \qquad (5.4.1)$$

where $\lambda(t)$ is an arbitrary unspecified base-line hazard function for continuous T, and $g(\underset{\sim}{Z}) = g(\underset{\sim}{Z};\underset{\sim}{\beta})$. Cox (1972) introduced models in which $g(\underset{\sim}{Z};\underset{\sim}{\beta}) = \exp(\underset{\sim}{\beta}'\underset{\sim}{Z})$, so that we have

$$\lambda(t;\underset{\sim}{Z}) = \lambda(t)\exp(\underset{\sim}{\beta}'\underset{\sim}{Z}) = \lambda(t)\exp(\sum_{i=1}^{p} \beta_i Z_i) \qquad (5.4.2)$$

or

$$\log \lambda(t;\underset{\sim}{Z}) = \log \lambda(t) + \sum_{i=1}^{p} \beta_i Z_i \qquad (5.4.3)$$

In this model, the covariates act multiplicatively on the hazard. An advantage of model (5.4.2) is that no restriction need to be imposed on the value of the expression $\underset{\sim}{\beta}'\underset{\sim}{Z}$, since $\exp(\underset{\sim}{\beta}'\underset{\sim}{Z}) > 0$ always. If $\lambda(t) = \lambda$, Equation (5.4.2) reduces to the exponential regression model $\lambda\exp(\underset{\sim}{\beta}'\underset{\sim}{Z})$.

The conditional density function of T given $\underset{\sim}{Z}$ corresponding to Equation (5.4.2) is

$$f(t;\underset{\sim}{Z}) = \lambda(t)e^{\underset{\sim}{\beta}'\underset{\sim}{Z}} \exp[e^{-\underset{\sim}{\beta}'\underset{\sim}{Z}} \int_0^t \lambda(u)du]$$

The conditional survivor function for T given $\underset{\sim}{Z}$ is

$$S(t;\underset{\sim}{Z}) = [S_o(t)]^{\exp(\underset{\sim}{\beta}'\underset{\sim}{Z})}$$

where

$$S_o(t) = \exp[-\int_0^t \lambda(u)du]$$

Let us consider the general multiplicative model (5.4.1). For the individual j this is

$$\lambda(t;\underset{\sim}{Z}_j) = \lambda(t)g(\underset{\sim}{Z}_j;\underset{\sim}{\beta}) \qquad (5.4.4)$$

where

$g(\underset{\sim}{Z}_j;\underset{\sim}{\beta})$ has a known form, while

$\lambda(t)$ is unspecified.

We assume that the Z's are independent of t; then for two individuals j and ℓ, the hazard rates are in constant ratio, $g(\underset{\sim}{Z}_j;\beta)/g(\underset{\sim}{Z}_\ell;\beta)$ whatever the time t. A model of this kind is called a proportional hazard model. We can obtain the ML-estimators of β's when the parametric form of $g(\underset{\sim}{Z};\beta)$ is specified.

To apply the ML-estimation, it would be convenient if $\lambda(t)$ could be expressed in terms of some (not too numerous) parameters and one way of going about this is to suppose it to be constant over a number of intervals of time. Divide the range of variation of t in w fixed (consecutive) intervals $I_k \equiv (t_{k-1}, t_k]$, $k=1,2,\ldots,w$. Suppose

$$\lambda(t) = \lambda_k \quad \text{for} \quad t_{k-1} < t \leq t_k \tag{5.4.5}$$

and so the hazard rate is

$$\lambda(t;\underset{\sim}{Z}) = \lambda_k \, g(\underset{\sim}{Z};\beta) \quad \text{for} \quad t_{k-1} < t \leq t_k \tag{5.4.6}$$

Let s_k be the set of all individuals who were in the study at any time in I_k; let $I_{k\ell}$ denote the part of I_k for which individual ℓ was in the study, and $h_{k\ell}$ is the length of $I_{k\ell}$. The length of I_k is $t_k - t_{k-1} = h_k$. If ℓ is in the study for the whole of I_k, then $h_{k\ell} = h_k$. If information on exact time of withdrawal is not available, we might use the approximation $h_{k\ell} \cong \frac{1}{2}(t_k - t_{k-1})$; if entry and withdrawal, or death, both occur in I_k, one might take $h_{k\ell} \cong \frac{1}{3}(t_k - t_{k-1})$.

For the period exposed to risk in I_k the cumulative hazard function for the individual ℓ is [using model (5.4.6)]

$$\Lambda_k(\underset{\sim}{Z}_\ell) = \lambda_k \int_{I_k} g(\underset{\sim}{Z}_\ell ; \underset{\sim}{\beta})\,dt = \lambda_k\, h_{k\ell}\, g(\underset{\sim}{Z}_\ell ; \underset{\sim}{\beta}) \qquad (5.4.7)$$

Let

$$\delta_{k\ell} = \begin{cases} 1 & \text{if } \ell \text{ dies in } I_k \\ 0 & \text{otherwise} \end{cases} \qquad (5.4.8)$$

The contribution to the likelihood for observations over the interval I_k is

$$L_k(\lambda_k ; \underset{\sim}{\beta}) = \prod_{\ell \in S_k} [\lambda_k\, g(\underset{\sim}{Z}_\ell ; \underset{\sim}{\beta})]^{\delta_{k\ell}} \exp[-\lambda_k\, h_{k\ell}\, g(\underset{\sim}{Z}_\ell ; \underset{\sim}{\beta})] \qquad (5.4.9)$$

The overall likelihood function is

$$L(\lambda_1, \ldots, \lambda_w ; \underset{\sim}{\beta}) = \prod_{k=1}^{w} L_k(\lambda_k ; \underset{\sim}{\beta}) \qquad (5.4.10)$$

If the parameters $\underset{\sim}{\beta}$ are fixed, the ML-estimator of λ_k is

$$\hat{\lambda}_k(\underset{\sim}{\beta}) = \frac{\sum_{\ell \in S_k} \delta_{k\ell}}{\sum_{\ell \in S_k} h_{k\ell}\, g(\underset{\sim}{Z}_\ell ; \underset{\sim}{\beta})} \qquad (5.4.11)$$

The numerator, $\sum_{\ell \in S_k} \delta_{k\ell}$, is the total number of deaths in I_k.

If no deaths are observed in I_k, then $\hat{\lambda}_k = 0$, whatever be $\underset{\sim}{\beta}$. We observe also that if there are no concomitant variables, [i.e., $g(Z_\ell;\underset{\sim}{\beta}) \equiv 1$], we obtain the formulas $\hat{\lambda}_k = \frac{\text{number of deaths}}{\text{exposed to risk}}$. If we substitute Equation (5.4.11) into Equation (5.4.9), we obtain the maximized value of the likelihood factor $L_k(\lambda_k;\underset{\sim}{\beta})$ given $\underset{\sim}{\beta}$

$$\hat{L}_k(\underset{\sim}{\beta}) = \left[e^{-1} \frac{\sum_{\ell \in S_k} \delta_{k\ell}}{\sum_{\ell \in S_k} h_{k\ell}\, g(\underset{\sim}{Z}_\ell;\underset{\sim}{\beta})} \right]^{\sum_\ell \delta_{k\ell}} \prod_{\ell \in S_k^*} g(\underset{\sim}{Z}_\ell;\underset{\sim}{\beta}) \qquad (5.4.12)$$

where

S_k^* is the set of individuals dying in I_k.

The maximum likelihood estimators of $\underset{\sim}{\beta}; \hat{\underset{\sim}{\beta}}$, maximize

$$L(\underset{\sim}{\beta}) = \prod_{k=1}^{w} \hat{L}_k(\underset{\sim}{\beta})$$

The estimates $\hat{\underset{\sim}{\beta}}$ are usually obtained by numerical analysis, employing the aid of computers.

5.4.1 Cox's "Conditional Likelihood" Approach

In an important paper, Cox (1972) proposed analyzing the model where the hazard rate

$$\lambda(t;\underset{\sim}{Z})dt = \frac{dF(t;\underset{\sim}{Z})}{\{1-F(t;\underset{\sim}{Z})\}}$$

for survival with independent variables $\underset{\sim}{Z} = (Z_1,\ldots,Z_p)$ is given by

$$\lambda(t;\underset{\sim}{Z}) = \lambda_o(t)\exp(\underset{\sim}{\beta}'\underset{\sim}{Z}) \qquad (5.4.1.1)$$

This is equivalent to assuming a Lehmann-alternative family of distributions

$$1 - F(t;\underset{\sim}{Z}) = [1-F(t)]^{\exp(\underset{\sim}{\beta}'\underset{\sim}{Z})} \qquad (5.4.1.2)$$

where

$$F(t) = 1 - \exp\{-\int_o^t \lambda(u)du\} \qquad (5.4.1.3)$$

The function $\lambda_o(t)$ is an unspecified function of time, and the multiplicative factor $\exp(\underset{\sim}{\beta}'\underset{\sim}{Z})$, where $\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_p)$ an unknown vector of parameters, gives the risk of failure at covariate $\underset{\sim}{Z}$ relative to that at $\underset{\sim}{Z} = \underset{\sim}{0}$. Kalbfleisch and McIntosh (1977) and others have noted that the model (5.4.1.1) permits $\underset{\sim}{Z}$ to vary with time.

Let us first deal with continuous time and assume that failures occur at distinct times $t_{(1)} < \ldots < t_{(m)}$. Assuming the model (5.4.1.1), Cox advocated a conditional likelihood approach to estimating $\underset{\sim}{\beta}$. Namely, let R_j denote the risk set at time $t_j - 0$, that is the set of individuals who have not

failed or been censored by that time. Further, let $\underset{\sim}{Z}_k$ denote the value of $\underset{\sim}{Z}$ for the $K^{\underline{th}}$ individual and $\underset{\sim}{Z}_{(j)}$ the value for the individual failing at time $t_{(j)}$ then Cox (1972) gave

$$\prod_j \frac{\exp(\underset{\sim}{\beta}'\underset{\sim}{Z})}{\sum_{k \in R_j} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_k)} \qquad (5.4.1.4)$$

as a likelihood for inference about $\underset{\sim}{\beta}$.

In other words, Cox's argument is that, condition on a death occurring at time t_j and the set of people known to be alive at t_j (called the risk set, $R(t_j)$ then the probability that the death happened to the $j^{\underline{th}}$ person is just

$$p_j = \frac{\lambda(t_j, \underset{\sim}{Z}_{(j)})}{\sum_{k \in R(t_j)} \lambda(t_j, \underset{\sim}{Z}_k)} = \frac{\exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_{(j)})}{\sum_{k \in R(t_j)} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_k)} \qquad (5.4.1.5)$$

Multiplying the p_j's for each observed death gives a likelihood depending only on $\underset{\sim}{\beta}$ as given by Equation (5.4.1.4), which we denote by $L(\underset{\sim}{\beta})$ [i.e., $L(\underset{\sim}{\beta})$ is defined to be the product of the terms (5.4.1.5) over all the uncensored points.] Various discussants (Lindley 1972; Kalbfleisch and Prentice 1972; Breslow 1972) noted that Equation (5.4.1.4) is not the claimed conditional likelihood and Cox (1975) admitted misleadingly calling it a conditional likelihood and later justified its use as a "partial" likelihood. Kalbfleisch and Prentice (1973) showed that Equation (5.4.1.4) is a marginal likelihood of ranks under the

very restrictive assumptions, that there is no censoring, and that $\underset{\sim}{Z}$ does not depend on time. Thus

$$\text{Log } L(\underset{\sim}{\beta}) = \underset{\sim}{\beta} \sum_{uc} \underset{\sim}{Z}_{(j)} - \sum_{uc} \log\Big(\sum_{R(t_j)} \exp(\underset{\sim}{\beta}' \underset{\sim}{Z}_k)\Big) \qquad (5.4.1.6)$$

Where the summation is over all the uncensored points in the sample. The derivative of Equation (5.4.1.6) is

$$U(\underset{\sim}{\beta}) = \frac{\log L(\underset{\sim}{\beta})}{\partial \underset{\sim}{\beta}} = \sum_{uc} \left(\underset{\sim}{Z}_{(j)} - \frac{\sum_{R(t_j)} \underset{\sim}{Z}_k \exp(\underset{\sim}{\beta}' \underset{\sim}{Z}_k)}{\sum_{R(t_j)} \exp(\underset{\sim}{\beta}' \underset{\sim}{Z}_k)} \right) \qquad (5.4.1.7)$$

and the value of $\underset{\sim}{\beta}$ for which this equals zero is usually found through iteration on a computer.

We should note that the hazard rate function $\lambda_o(t)$ does not enter the likelihood in Equation (5.4.1.6). Cox argued that since $\lambda_o(t)$ is unspecified in the model, it could be zero between uncensored t_j and therefore the sample contains no information about $\underset{\sim}{\beta}$ except at the uncensored points. Once $\hat{\underset{\sim}{\beta}}$ has been obtained Cox used a product-limit approach to derive an estimate of the distribution F.

In the discussion of Cox's paper and a subsequent paper (1974), Breslow proposed a slightly different likelihood function based on the assumption that the hazard rate function $\lambda_o(t)$ is constant between uncensored observations. The maximum likelihood estimate of $\underset{\sim}{\beta}$ is the same but the Breslow

estimate for F is simpler than Cox's estimate.

To accommodate discrete failure time $t \in (t_1, t_2, \ldots)$ Cox generalized Equation (5.4.1.1) to a logistic regression model in which the conditional failure probability at $t = t_1$ satisfies

$$\frac{\lambda(t_i, \underset{\sim}{Z})}{1 - \lambda(t_i, \underset{\sim}{Z})} = \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}) \frac{\lambda_o(t_i)}{1 - \lambda_o(t_i)} \qquad (5.4.1.8)$$

The multiplicative factor $\exp(\underset{\sim}{\beta}'\underset{\sim}{Z})$ now has an odds ratio, rather than a relative risk interpretation. The model (5.4.1.8) will approach Equation (5.4.1.1) as each $\lambda_o(t_i)$ for $i=1,2,\ldots$ becomes small. Note that the odds ratio can be allowed to vary with time by permitting components of $\underset{\sim}{Z}$ to depend on t. We note also that, upon setting $\alpha_i = \lambda_o(t_i)$, Equation (5.4.1.8) can be rewritten as

$$\hat{\lambda}(t_i, Z) = \frac{\exp(\alpha_i + \underset{\sim}{\beta}'\underset{\sim}{Z})}{(1 + \exp(\alpha_i + \underset{\sim}{\beta}'\underset{\sim}{Z}))} \qquad (5.4.1.9)$$

The models (5.4.1.1), (5.4.1.8) or (5.4.1.9) apply directly to prospective failure-time studies and in either case, yield a partial likelihood (Cox (1972,1975)) for β that is the product, over each distinct failure time, of terms

$$\sum_{i=1}^{r} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_i) \Big/ \sum_{\Psi} \prod_{i=1}^{r} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_{\Psi_i}) \qquad (5.4.1.10)$$

where

$\underset{\sim}{Z}_1,\ldots,\underset{\sim}{Z}_n$ denote the covariate vectors for the n individuals at risk at that failure time,

$\underset{\sim}{Z}_1,\underset{\sim}{Z}_2,\ldots\underset{\sim}{Z}_r$ correspond to the failures, and

Ψ denotes the set of all subsets $(\Psi_1,\Psi_2\ldots,\Psi_r)$ of size r from $\{1,\ldots,n\}$.

Farewell and Prentice (1980) noted that the denominator in Equation (5.4.1.10) is computationally impracticable for $\underset{\sim}{\beta}' \neq \underset{\sim}{0}'$ if both r and $n-r$ are at all large. Two approximations to Equation (5.4.1.10) may be considered. The "maximized" likelihood of Breslow (1974), which generalized a suggestion of Peto (1972), gives

$$\prod_{i=1}^{r} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_i)/[\binom{n}{r}\{n^{-1}\sum_{i=1}^{n} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_i)\}^{r}] \qquad (5.4.1.11)$$

as an approximation to Equation (5.4.1.10). Efron (1977) suggested that a more accurate approximation may be

$$\prod_{i=1}^{r} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_i)/\prod_{i=1}^{r-1}\{\sum_{i=1}^{n} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_i) - \frac{1}{r}\sum_{i=1}^{r} \exp(\underset{\sim}{\beta}'\underset{\sim}{Z}_i)\} \qquad (5.4.1.12)$$

Either approximation is likely to be adequate if r/n is small for most failures; in particular, Equations (5.4.1.11) and (5.4.1.12) are equal to Equation (5.4.1.10) if either $r = 1$ or $\underset{\sim}{\beta} = \underset{\sim}{0}$. Approximation (5.4.1.11) is used in a number of

computer programs that aim to fit Equation (5.4.1.1) to failure time data. Farewell and Prentice (1980) have examined the accuracy of the approximations (5.4.1.11) and (5.4.1.12) for situations in which r/n values may be large, for example, $r/n = \frac{1}{2}$, at most failure times. The authors found out that Equations(5.4.1.10) and (5.4.1.11) turn out to be generally very inadequate with large values of r/n.

Breslow (1972) points out that if $\lambda_o(t) = \lambda_i$ for $t_{(i-1)} < t \leq t_{(i)}$, $i=1,\ldots,k$ for distinct uncensored ordered failure times $t_{(1)} < t_{(2)} < \ldots < t_{(k)}$, $t_{(0)} = 0$ one obtains the Kaplan and Meier (1958) product limit estimate for probability of survival at the point $t_{(i)}$ when $\underset{\sim}{\beta}' = \underset{\sim}{0}'$, i.e.,

$$\hat{S}(t_{(i)}) = \prod_{j=1}^{i} \frac{n_j - 1}{(n_j)}$$

where

$\hat{S}(t_{(i)})$ is the estimated probability of survival at the point $t_{(i)}$, and

n_j is the number of individuals alive in the trial at t_j.

Breslow shows how the estimates in $\hat{S}(t_{(i)})$ can be extended to the case when $\underset{\sim}{\beta}' \neq \underset{\sim}{0}'$ and $\hat{\underset{\sim}{\beta}}'$ is the estimator.

REFERENCES

[1] AITKEN, M. and CLAYTON, D. (1968). The fitting of Exponential, Weibull and Extreme Value Distributions to Complex Censored Survival Data Using GLIM. Journal of Applied Statistics, 29, 156-163.

[2] ARMITAGE, P. (1959). The Comparison of Survival Curves. Journal of the Royal Statistical Society, A122, 279-300.

[3] BAILEY, R.C., HOMER, L.D., and SUMME, J.P. (1977). A Proposal for the Analysis of Kidney Graft Survival. Transplantation, 24, 309-315.

[4] BAIN, L.J. and ANTLE, C.E. (1967). Estimation of Parameters in the Weibull Distribution. Technometrics, 9, 629-628.

[5] BANCROFT, G.A. and DUNSMORE, I.R. (1976). Predictive Distributions in Life Tests Under Competing Causes of Failure. Biometrika, 63, 195-217.

[6] BARLOW, R.E. and HUNTER, L.C. (1960). System Efficiency and Reliability. Technometrics, 2, 43-53.

[7] BARLOW, R.E. and PROSCHAN, F. (1965). Mathematical Theory of Reliability, John Wiley and Sons, New York.

[8] BARTHOLOMEW, D.J. (1957). A Problem in Life Testing. Journal of the American Statistical Association. 52, 350-355.

[9] BASU, A.P. (1964). Estimates of Reliability for Some Distributions Useful in Life Testing. Technometrics. 6, 215-219.

[10] BASU, A.P. (1971). Bivariate Failure Rate. Journal of the American Statistical Association. 66, 103-104.

[11] BERGER, A. and GOLD, R.Z. (1961). On Comparing Survival Times. Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, 67-76.

[12] BERKSON, J. and ELVEBACK, L. (1960). Competing Exponential Risks With Particular Reference to the Study of Smoking and Lung Cancer. Journal of the American Statistical Association. 55, 415-428.

[13] BERKSON, JOSEPH and GAGE, R. (1952). Survival Curve for Patients Following Treatment. Journal of the American Statistical Association, 47,501-515.

[14] BERMAN, S.M. (1963). Notes on Extreme Values, Competing Risks and Semi-Markov Processes. Annals of Mathematical Statistics, 34, 1104-1106.

[15] BHATTACHARYA, S.R. (1967). Bayesian Approach to Life Testing and Reliability Estimation. Journal of the American Statistical Assoc., 62,48-62.

[16] BOAG, J.W. (1949). Maximum Likelihood Estimates of the Proportion Cured by Cancer Therapy. Journal of the Royal Statistical Society. B11,15-53.

[17] BOARDMAN, T.J. and KENDELL, P.J. (1970). Estimation in Compound Exponential Failure Models. Technometrics. 12, 891-900.

[18] BRESLOW,N.E. (1972). Discussion of Cox (1972). Journal of the Royal Statistical Society. B34,187-220.

[19] BRESLOW,N. and CROWLEY, J. (1974). A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship. Annals of Statistics.2,437-453.

[20] BROADBENT, S. (1958). A Simple Mortality Rate. Journal of Applied Statistics.7,86-95.

[21] BYAR, D.P., HUSE, R., and BAILAR, J.C. (1974). An Exponential Model Relating Censored Survival Data and Concomitant Information for Prostatic Cancer Patients. Journal of National Cancer Institute. 52, 321-326.

[22] CHIANG, C.L. (1960a). A Stochastic Study of the Life Table and Its Applications: I. Probability Distribution of the Biometric Functions. Biometrics. 16,618-635.

[23] CHIANG, C.L. (1960b). A Stochastic Study of the Life Table and Its Applications:II. Sample Variance of the Observed Expectations of Life and Other Biometric Functions. Human Biology. 32,221-238.

[24] CHIANG, C.L. (1961a). A Stochastic Study of the Life Table and Its Applications: III. The Follow-Up Study With the Consideration of Competing Risks. Biometrics. 17,57-58.

[25] CHIANG, C.L. (1961b). On the Probability of Death From Specific Causes in the Presence of Competing Risks. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. 169-180.

[26] CHIANG, C.L. (1967). Variance and Covariance of Life Table Functions Estimated From a Sample of Deaths. Vital Health Statistics. Series 2, No.20.

[27] CHIANG, C.L. (1968). Introduction to Stochastic Processes in Biostatistics. John Wiley & Sons, New York.

[28] CHIANG, C.L. (1970). Competing Risks and Conditional Probabilities. Biometrics. 26, 767-776.

[29] COHEN, A.C. (1951). Estimating Parameters of the Logarithmic Normal Distribution by Maximum Likelihood. Journal of the American Stats. Association. 46,206-212.

[30] COHEN, A.C. (1959). Simplified Estimators for the Normal Distribution When Samples are Singly Censored or Truncated. Technometrics. 1,217-237.

[31] COHEN, A.C. (1963). Progressively Censored Samples in Life Testing. Technometrics. 5,327-339.

[32] COHEN, A.C. (1965). Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and Censored Samples. Technometrics. 7, 579-588.

[33] COHEN, A.C. (1966). Life Testing and Early Failure. Technometrics. 8,539-545.

[34] CORNFIELD, J. (1957). The Estimation of the Probability of Developing a Disease in the Presence of Competing Risks. American Journal of Public Health. 47,601-607.

[35] COX, D.F. and KEMPTHORNE, O. (1963). Randomization Tests for Comparing Survival Curves. Biometrics. 19,307-317.

[36] COX, D.R. (1959). The Analysis of Exponentially Distributed Life Times With Two Types of Failure. Journal of the Royal Statistical Society. B21,411-421.

[37] COX, D.R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society. B34,187-220.

[38] COX, D.R. (1975). Partial Likelihood. Biometrika.62,269-276.

[39] CRAMÉR,H.(1946). Mathematical Methods of Statistics. Princeton University Press, Princeton,New Jersey.

[40] CROWLEY, John (1974). Asymptotic Normality of a New Non-Parametric Statistic for Use in Organ Transplant Studies. Journal of the American Statistical Association. 64,1006-1011.

[41] CROWLEY, J. and HU, MARIE (1977). Covariance Analysis of Heart Transplant Survival Data. Journal of the American Statistical Association. 72, 27-36.

[42] CUTLER, S.J. and EDERER, F. (1968). Maximum Utilization of the Life Table Method in Analyzing Survival. Journal of Chronic Diseases. 8,699-713.

[43] DAVID, H.A. (1970). On Chiang's Proportionality Assumption in the Theory of Competing Risks. Biometrics.26, 336-337.

[44] DAVIS, H.T. and FELDSTEIN, M. (1979). The Generalized Pareto Law as a Model for Progressively Censored Survival Data. Biometrika.66,299-306.

[45] DIXON, W.J. and MASSEY, F.J. (1969). Introduction to Statistical Analysis, McGraw-Hill Book Company, New York.

[46] DUBEY, SATYA D. (1963). On Some Statistical Inferences for Weibull Laws. Journal of the American Statistical Association.58, 549-551.

[47] EFFRON, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. Journal of the American Statistical Association. 72,557-565.

[48] EHRENTHIEL, O.F. and MUENCH, H. (1959). Survival Rate of Patients Hospitalized With the Diagnosis of General Paresis. Journal of Chronic Diseases.9,41-54.

[49] ELANDT-JOHNSON, R., and JOHNSON, N.L. (1980). Survival Models and Data Analysis. John Wiley & Sons, Toronto.

[50] ELVEBACK, L. (1958). Estimation of Survivorship in Chronic Disease: The "Actuarial" Method. Journal of the American Statistical Association. 53,420-440.

[51] EPSTEIN, B. and SOBEL, M. (1953). Life Testing. Journal of the American Statistical Association. 48, 486-502.

[52] ESARY, J.D. and PROSCHAN, F. (1963). Relationship Between System Failure Rate and Component Failure Rates. Technometrics. 5,183-189.

[53] FAREWELL, V.T., and PRENTICE, R.L. (1980). The Approximation of Partial Likelihood With Emphasis on Case-Control Studies. Biometrika. 67,273-278.

[54] FIEGL, P., and ZELEN, M. (1965). Estimation of Exponential Survival Probabilities With Concomitant Information. Biometrics. 21,826-838.

[55] FEINLEIB, M. (1960). A Method of Analyzing Log-Normally Distributed Survival Data With Incomplete Follow-Up. Journal of the American Statistical Association. 55,534-545.

[56] FEINLEIB, MANNING and MacMAHON, B. (1960). Variation in the Duration of Survival of Patients With Chronic Leukemia. Blood. 15,220-235.

[57] FELLER, W. (1968). An Introduction to Probability Theory and Its Applications. Volumes I and II. John Wiley & Sons, New York.

[58] FIX, E., and NEYMAN, J. (1951). A Simple Stochastic Model of Recovery, Relapse, Death and Loss of Patients. Human Biology. 24,205-241.

[59] FREI, E., et al. (1963). Studies of Sequential and Combination Antimetabolite Therapy in Acute Leukemia: 6-Mercaptopurine and Methotrexate. Blood. 18,431-454.

[60] FREIREICH, E.J., GEHAN, E., et al (1963). The Effect of 6-Mercaptopurine on the Duration of Steroid-Induced Remissions in Acute Leukemia: A Model for Evaluation of Other Potentially Useful Therapy. Blood. 21,699-716.

[61] FREUND, J.E. (1961). A Bivariate Extension for the Exponential Distribution. Journal of the American Statistical Association. 56, 971-977.

[62] FRIDAY, D.S. (1976). A New Multivariate Life Distribution. Unpublished Thesis. Pennsylvania State University, Department of Statistics.

[63] FROME, E.L. and BEAUCHAMP,J.J. (1968). Maximum Likelihood Estimation of Survival Curve Parameters. Biometrics. 24,595-606.

[64] GADDUM, J.H. (1945). Lognormal Distributions. Nature. 156,463-466.

[65] GARABEDIAN, P.R. (1964). Partial Differential Equations. John Wiley & Sons, New York.

[66] GEHAN, E.A. (1967). Some Considerations in the Planning of a Clinical Trial. Talk at the 6th International Biometric Conference, Sydney, Australia.

[67] GEHAN, E.A. (1969). Estimating Survival Functions From the Life Table. Journal of Chronic Diseases. 21, 629-644.

[68] GEHAN, E.A. (1970). Unpublished Notes on Survivability Theory. University of Texas, M.D. Anderson Hospital and Tumour Institute, Houston.

[69] GEHAN, E.A. and SIDDIQUI,M.M. (1973). Simple Regression Methods for Survival Time Studies. Journal of the American Statistical Association. 68, 848-856.

[70] GLASSER,M. (1967). Exponential Survival With Covariance. Journal of the American Statistical Association. 62,561-568.

[71] GREENWOOD, M. (1926). The Natural Duration of Cancer. Reports on Public Health and Medical Subjects. (Her Majesty's Stationery Office,London). Vol.33, 1-26.

[72] GREENWOOD, J.A. and DURAND, D. (1960). Aids for Fitting the Gamma Distribution by Maximum Likelihood. Technometrics.2, 55-65.

[73] GROSS, A., CLARK, V., and LIU, V. (1971). Estimation of Survival Parameters When One of Two Organs Must Function for Survival. Biometrics. 27, 369-377.

[74] GROSS, A., and CLARK, VIRGINIA (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. John Wiley & Sons, New York.

[75] GUMBEL, E.J. (1958). Statistics of Extremes. Columbia University Press, New York.

[76] HALL, W.J. and WELLNER, J.A. (1980). Confidence Bands for Survival Curve From Censored Data. Biometrika. 67,133-143.

[77] HALPERIN, M. (1952). Maximum Likelihood Estimation in Truncated Samples. Annals of Mathematical Statistics. 23, 226-238.

[78] HANKEY, B.F. and MYERS, M.H. (1971). Evaluating Differences in Survival Between Two Groups of Patients. Journal of Chronic Diseases. 24, 523-531.

[79] HARTER, H.L., and MOORE, A.H. (1965). Maximum Likelihood Estimation of Parameters of Gamma and Weibull Populations From Complete and From Censored Samples. Technometrics. 7,639-643.

[80] HARTLEY, H.O. (1958). Maximum Likelihood Estimation From Incomplete Data. Biometrics. 14,174-194.

[81] HERMAN, R.J. and PATELL, R.K.N. (1971). Maximum Likelihood Estimation for Multi-Risk Model. Technometrics. 13,385-397.

[82] HOEL, D.G. (1972). A Representation of Mortality Data by Competing Risks. Biometrics. 28, 475-488.

[83] HOLT, J.D. and PRENTICE, R.L. (1974). Survival Analysis in Twin Studies and Matched Pair Experiments. Biometrika. 61,17-30.

[84] IRWIN,J.O. (1942). The Distribution of the Logarithm of Survival Times When the True Law is Exponential. Journal of Hygiene. 42,328-333.

[85] KALBFLEISCH, JOHN D. (1974). Some Efficiency Calculations for Survival Distributions. Biometrika,61,31-38.

[86] KALBFLEISCH, J.D., and McINTOSH, A.A. (1977). Efficiency in Survival Distributions With Time-Dependent Covariables. Biometrika, 64,47-50.

[87] KALBFLEISCH,J., and PRENTICE, R.L. (1972). Discussion of Cox (1972). Journal of the Royal Statistical Society, B34,187-220.

[88] KALBFLEISCH, J.D. and PRENTICE, R.L. (1973). Marginal Likelihoods Based on Cox's Regression and Life Model. Biometrika, 60,267-278.

[89] KALBFLEISCH,J.D. and PRENTICE, R.L. (1980). The Statistical Analysis of Failure Time Data. John Wiley & Sons, Toronto.

[90] KALBFLEISCH,J.D. and SPROTT, D.A. (1970). Application of Likelihood Methods to Models Involving a Large Number of Parameters. Journal of the Royal Statistical Society, B32,175-208.

[91] KALE, B.K. (1961). On the Solution of the Likelihood Equation by Iteration Process. Biometrika,48,452-456.

[92] KALE, B.K. (1962). On the Solution of the Likelihood Equation by Iteration Processes. The Multiparametric Case. Biometrika,49,479-486.

[93] KAPLAN, E.L., and MEIER, P. (1958). Nonparametric Estimation From Incomplete Observations. Journal of the American Statistical Association,53,457-481.

[94] KAO, J.H.K. (1959). A Graphical Estimation of Mixed Weibull. Parameters in Life-Testing of Electron Tubes. Technometrics,1,389-407.

[95] KAY, R. (1977). Proportional Hazard Regression Models and the Analysis of Censored Survival Data. Journal of the Royal Statistical Society, C26,227-237.

[96] KENDALL, M.G. and STUART, A. (1973). The Advanced Theory of Statistics,Vol.II, Hafner Publishing Company,New York.

[97] KENNEDY, W.J. and GENTLE, J.E. (1980). Statistical Computing, Marcel Dekker Inc., New York.

[98] KIMBALL, A.W. (1958). Disease Incidence Estimation in Populations Subject to Multiple Causes of Death. Bull. Inst.Int.Statistics,36,193-204.

[99] KIMBALL, A.W. (1969). Models for the Estimation of Competing Risks From Grouped Data. Biometrics,25,329-337.

[100] KRANE, S.A. (1963). Analysis of Survival Data by Regression Techniques. Technometrics,5,161-174.

[101] LAIRD,N. and OLIVIER, D. (1981). Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. Journal of the American Statistical Association, 76,231-240.

[102] LEA, D.E. (1945). Biological Assay of Carcinogens. Cancer Research,5,633-640.

[103] LEHMAN, Eugene H. (1963). Shapes, Moments and Estimation of the Weibull Distribution. I.E.E.E. Transactions on Reliability, R-12,3,32-38.

[104] LINDLEY, D.V. (1972). Discussion of Cox (1972). Journal of the Royal Statistical Society, B34,187-220.

[105] LINGAPPAIAH, G.S. (1979). Bayesian Approach to the Prediction of the Restricted Range in the Censored Samples From the Exponential Population. Biometrics, 21,361-366.

[106] LITTEL, A.S. (1952). Estimation of the T-Year Survival Rate From Follow-Up Studies Over a Limited Period of Time. Human Biology,24,87-116.

[107] LUSTBRADDER, Edward D. (1980). Time Dependent Covariates in Survival Analysis. Biometrika,67,697-698.

[108] MacDONALD, E.J. (1963). The Epidemiology of Melanoma. Annals of the New York Academy of Sciences,100,4-15.

[109] MANTEL, N. and BYAR, D.P. (1974). Evaluation of Response - Time Data Involving Transient States: An Illustration Using Heart-Transplant Data. Journal of the American Statistical Association,69,81-86.

[110] MANTEL,N. and MYERS, M. (1971). Problems of Convergence of Maximum Likelihood Iterative Procedures in Multiparameter Situations. Journal of the American Statistical Association,66,484-491.

[111] MARSHALL, A. and OLKIN, I. (1967). A Multivariate Exponential Distribution. Journal of the American Statistical Association,62,30-44.

[112] MENDENHALL, W. and HADER, R.J. (1958). Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions From Censored Life Test Data. Biometrika,45,504-520.

[113] MENON, M.V. (1963). Estimation of the Shape and Scale Parameters of the Weibull Distribution. Technometrics,5, 175-182.

[114] MILLER, I. and FREUND, J.E. (1965). Probability and Statistics for Engineers. Prentice-Hall,Inc.,Englewood Cliffs,New Jersey.

[115] MILLER, R. (1976). Least Squares Regression With Censored Data. Biometrika,63,449-464.

[116] MOESCHBERGER,M.L. (1974). Life Tests Under Dependent Competing Causes of Failure. Technometrics,16,39-47.

[117] MOESCHBERGER,M.L. and DAVID, H.A. (1971). Life Tests Under Competing Causes of Failure and the Theory of Competing Risks. Biometrics,27,909-933.

[118] MOOD, A.M. and GRAYBILL, F.A. (1963). Introduction to the Theory of Statistics. McGraw-Hill Book Company,New York.

[119] MURTHY,V.K. and SWARTZ,G.B. (1975). Estimation of Weibull Parameters From Two-Order Statistics. Technometrics,17,96-102.

[120] NAGARSENKER, P.B. (1980). On Test of Equality of Several Exponential Survival Distributions. Biometrika, 67,475-478.

[121] NELSON, Wayne (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. Technometrics,14, 945-965.

[122] OSGOOD, E.E. (1958). Methods for Analyzing Survival Data. Illustrated by Hodgkins Disease. American Journal of Medicine,24,40-47.

[123] OSGOOD, E.E. and SEAMAN, A.J. (1952). Treatment of Chronic Leukemia. Journal of the American Medical Association,150,1372-1379.

[124] PETO, RICHARD (1972). Discussion of Cox (1972). Journal of the Royal Statistical Society,B34,187-220.

[125] PETO, R. and PETO, JULIAN (1972). Asymptotically Efficient Rank Invariant Test Procedures.Journal of the Royal Statistical Society,A135,185-207.

[126] PRENTICE, R.L. (1973). Exponential Survivals With Censoring and Explanatory Variables. Biometrika,60,279-288.

[127] PRENTICE, R.L. et al (1978). The Analysis of Failure Times in the Presence of Competing Risks. Biometrics, 34,541-554.

[128] PROSCHAN, F. (1963). Theoretical Explanation of Observed Decreasing Failure Rate. Technometrics,5,375-383.

[129] RAO, C.R. (1952). Advanced Statistical Methods in Biometric Research. John Wiley & Sons, New York.

[130] RAO, C.R. (1965). Linear Statistical Inference and Its Applications. John Wiley & Sons, New York.

[131] SCHEUER, E.M. (1968). Testing Grouped Data For Exponentiality. Rand Memorandum,RM-5692-PR, Santa Monica, California.

[132] SCHOENFELD, DAVID (1981). The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions. Biometrika,68,316-319.

[133] SPROTT, D.A. and KALBFLEISCH, J.D. (1969). Examples of Likelihoods and Comparison With Point Estimates and Large Sample Approximations. Journal of the American Statistical Association,64,468-484.

[134] STABLEIN, D.M. et al. (1980). Survival Analysis of Drug Combinations Using a Hazard Model With Time-Dependent Covariates. Biometrics,36,537-546.

[135] SUSARLA, V. and RYZIN, J.V. (1976). Nonparametric Bayesian Estimation of Survival Curves From Incomplete Observations. Journal of the American Statistical Association,71,897-902.

[136] TOSCH, J.T. and HOLMES, P.T. (1980). A Bivariate Failure Model. Journal of the American Statistical Association, 75,415-417.

[137] TSIATIS, ANASTASIOS A. (1981). The Asymptotic Joint Distribution of the Efficient Scores Test for the Proportional Hazards Model Calculated Over Time. Biometrika, 68,311-315.

[138] TURNBULL,B.W. (1974). Nonparametric Estimation of a Survivorship Function With Doubly Censored Data. Journal of the American Statistical Association,69,169-173.

[139] TURNBULL,B.W. et al (1974). Survivorship Analysis of Heart Transplant Data. Journal of the American Statistical Association,69,74-80.

[140] WATSON, G.S. and LEADBETTER, M.R. (1964). Hazard Analysis I. Biometrika,51,175-184.

[141] WEDDERBURN, R.W.M. (1976). On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models. Biometrika,63,27-32.

[142] WEINMAN, D.G. et al. (1973). On a Test For the Equality of Two Exponential Distributions. Technometrics,15,177-182.

[143] WILK, M.B., GNANADESIKAN,R. and HUYYETT, M.J. (1962). Estimation of Parameters of the Gamma Distribution Using Order Statistics. Biometrika,49,525-545.

[144] WHITAKER, E.T. and WATSON, G.N. (1927). A Course of Modern Analysis, Cambridge University Press, London.

[145] ZELEN, M. (1969). Play the Winner Rule and the Controlled Clinical Trial. Journal of the American Statistical Association,56,971-977.