ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

SURVIVAL ANALYSIS TECNIQUES

By

Marios T. Kondakis

A THESIS

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
2003

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

SURVIVAL ANALYSIS TECNIQUES

B y

Marios T. Kondakis

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece
October 2003

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΤΕΧΝΙΚΕΣ ΜΕΘΟΔΟΙ ΣΤΗΝ ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ

Μάριος Θ. Κονδάκης

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Οκτώβριος 2003

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Science

Survival Analysis Techniques

Marios T. Kondakis

Approved by the Graduate Committee

A. Dimaki	A. Kostaki	V. Vasdekis
Assistant Professor	Assistant Professor	Assistant Professor
Thesis Supervisor	Members of the Commitee	

Athens, October 2003

Michael Zazanis, Associate Professor
Director of the Graduate Program

DEDICATION

To: My parents

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Aikaterini Dimaki for the trust that she showed by assigning me the specific project, her valuable help and advice during the development of this project. In addition I want to express my appreciation to L.J. Wei and T. Cai for their useful advises on technical matters. I am also grateful the committee, which accepted my application and gave me the opportunity to participate in this program.

II

VITA

I was born in Piraeus in October 1976. In 1994 I entered the Department of Mathematics in the University of Crete and in July 1998 I received my degree. In October 1998 I was accepted from Department of Statistics of the Athens University of Economics and Business, to follow the M. Sc. Program in Statistics. During the period of March to June 1999 I attended courses for the M. Sc. Program as an Erasmus student in the Katholieke University of Leuven, Belgium. My research interests are in Survival data analysis and its application in statistical packages

ABSTRACT

Marios T. Kondakis

SURVIVAL ANALYSIS TECNIQUES

October 2003

The human life span, has always been attraction subject for the scientists. Survival techniques have appeared not only in the Statistics' literature but also in actuarial, medical and social studies as well. The effect of explanatory variables, which stems mainly from medical statistics but also from industrial life testing, plays a crucial role in understanding and analysing the quality of our life and thus it braces our effort of improving, or sometimes even extending it. In addition, the application ranges of the Survival Analysis extend much more widely, from physics to econometrics.

The object of the present report is to stage popular and recent techniques for tackling survival data queries and for making the adequate statistical inference about the population of these data. In this vicinity, the paper actually consists of seven parts. The first is a historical review on the survival analysis. Following the second part describes the main aspects of the subject and introduces the notation used in the thesis. Moreover, examples of survival random variable patterns are discussed in the third section. The common used Cox procedure and Classical models and procedures used in survival analysis can be found in the forth and fifth section respectively while the sixth part includes more recent **generalized** techniques used for the same kind of data, applying the **failures** of 23 **AML** leukemia patients divided into two treatment groups. The final conclusions are congregated in the last part.

ΠΕΡΙΛΗΨΗ

Μάριος Θ. Κονδάκης

ΤΕΧΝΙΚΕΣ ΜΕΘΟΔΟΙ ΣΤΗΝ ΑΝΑΛΥΣΗ ΕΠΙΒΙΩΣΗΣ

Οκτώβριος 2003

Η διάρκεια ζωής αποτελούσε ανέκαθεν πόλος έλξης για τους επιστήμονες. Μέθοδοι στην ανάλυση επιβίωσης εμφανίστηκαν όχι μόνο στη στατιστική βιβλιογραφία αλλά και στην ιατρική καθώς και στις κοινωνικές επιστήμες. Η επίπτωση των επεξηγηματικών μεταβλητών, οι οποίες προέρχονται και χρησιμοποιούνται κυρίως στην ιατρική στατιστική καθώς επίσης και στο κλάδο του ελέγχου της βιομηχανικής ζωής, παίζει καθοριστικό ρόλο στην κατανόηση και ανάλυση της ποιότητας της ζωής μας και έτσι ενδυναμώνει την προσπάθεια μας για βελτίωση ή ακόμη για επέκτασή της. Επιπλέον, η έκταση των εφαρμογών της ανάλυσης επιβίωσης εκτείνεται πολλή ευρύτερα, από το χώρο της φυσικής έως το χώρο της οικονομετρίας.

Ο σκοπός της παρούσας αναφοράς είναι να παρουσιάσει δημοφιλής και πρόσφατες μεθόδους για την αντιμετώπιση προβλημάτων δεδομένων επιβίωσης και για το σχηματισμό κατάλληλης στατιστικής συμπερασματολογίας σχετικά με τον πληθυσμό αυτών των δεδομένων. Ειδικότερα, η διατριβή αποτελείται από επτά μέρη. Το πρώτο είναι μια ιστορική αναδρομή στην ανάλυση επιβίωσης. Στη συνέχεια το δεύτερο περιγράφει τα κύρια σημεία του αντικειμένου και εισάγει την ορολογία που χρησιμοποιείται στην εργασία. Επιπλέον, παραδείγματα κατανομών επιβίωσης περιγράφονται στο τρίτο τμήμα. Η διαδεδομένη διαδικασία του S. Cox και άλλα κλασσικά μοντέλα και διαδικασίες που χρησιμοποιούνται στη ανάλυση επιβίωσης, βρίσκονται στο τέταρτο και πέμπτο μέρος ενώ το έκτο μέρος περιλαμβάνει πιο πρόσφατες **γενικευμένες** μεθόδους που χρησιμοποιούνται για την ίδια φύση δεδομένων, εφαρμόζοντας τις σε δεδομένα **αποτυχίας** 23 ασθενών **οξείας μυελογενούς λευχαιμίας**, οι οποίοι ομαδοποιούνται σε δύο κατηγορίες.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

On Survival Analysis

1.1 Introduction

According to Cox and Oakes (1984), the use of survival techniques have appeared not only in the Statistics' literature but in actuarial, medical and social studies as well. For instance life tables have been used for demographers and actuaries for many years to describe and compare the so-called expectation of life. The product limit estimator appears first to have been proposed by Bohmer but the actuarial estimator itself is much older. Greenwood and Kaplan & Meier (1958) derived the product limit estimator from maximum likelihood arguments. A key reference is Efron (1967). Mann et al. (1974), Gross and Clark (1975) and Lawless (1982) concentrate largely on fully parametric methods for survival distributions. Elandt-Johnson and Johnson (1980) describe the applications in actuarial science and demography. Miller (1981) describes nonparametric and semiparametric methods. For applications in industrial reliability see DePriest and Launer (1983). Miké and Stanley (1982) have edited a collection of papers on medical statistics including discussion of survival data. Also, Pett (1997) presents nonparametric statistical techniques for health care research settings. One recent paper, for the estimation of the survival distribution with right-censored data and covariates when collection of data is delayed, is by Van Der Laan and Hubbard (1998).

Chapter 2

Introduction to Survival Analysis

2.1 Introduction

Survival analysis is a number of statistical procedures in which interest centers on the time length until an event occurs at most once, called the **failure**. In particular, survival time corresponds to the time from the beginning of follow-up period of an individual until the failure. For instance we may interest in the time (in weeks) in remission for leukemia patients, or in the years until death for elderly people (60+).

In all the cases time is determined by a time origin, which is apparently defined, a scale for measuring the passage of time and the meaning of failure. The time origin should be precisely defined for each individual. Also, all individuals should be as comparable as possible at their time origin. In other words the time origin definition should not differ for each individual. The selection of the first instant at which the patient's symptoms meet certain criteria of severity as a time origin, may be more biologically meaningful, nevertheless such a value for the time origin is not only difficult to be determined but also subject to certain kind of bias. Such information might be useful as an explanatory variable, though.

In addition, the time origin need not be (as usual) at the same calendar time for each individual. Indeed, the time origin in most clinic trials has staggered entry over a

substantial time period. Also there are cases that the subject enters the study in a time point after the real time origin. For instance in a case of observing the failure time of an already used machine component, the time origin does not coincide with the time the component enters the study. The last kind of data are called **left-truncated** as there is an unknown time length at the left of the time the subject enters the study in the real line, which should have included in the failure time.

Referring to the scale parameter, the clock time is usually utilized. For instance hours, days, months, years, etc. are counted until the failure. Other possibilities are the age of a patient or the operating time of a system respectively. Also in geometrical probability applications, the length of a line segment contained in a convex body is the time to failure.

Finally, a clear definition of the failure event is essential. This event is typically called failure as the kind of event of interest usually is death, disease incidence, or some other negative individual experience. However, failure may be defined arbitrary like in some industrial contexts, where failure can be the first instance at which the performance, measured in some quantitative way, falls below an acceptable level, defined perhaps by a specification. In case where more than one event is considered in the same analysis, as death from any of several causes, the statistical problem is generally characterized by as a **competing risk** problem.

2.2 Censored Data

A familiar difficulty in the analysis of survival data is when we have some information about individual failure time, but **we do not know the real time to failure.** General reasons of censoring are during to Lee (1992) :

Firstly that the subject does not experience the failure event before the end of the study or the subject is lost to follow-up during the study period, or the subject withdraws from the study because of some reason (e.g. death if not the event of interest, adverse

drug reaction, etc.). This kind of censorship is called **Type I**. Secondly, another option is the study to end until a fixed portion of the subjects to fail. The latter is also known as **Type II** censoring Thirdly, in many clinical studies the period of study is fixed and patients enter the study at different times during that period. For subjects that are lost to follow up or do not fail until the end of the study, **Type III** censoring occurs and their survival times begin from their entrance until the censor time point. The first two types of observations are also called **singly censored data** while the third type is called **progressively censored data** or **random censoring**.

For instance, consider leukemia patients followed until they go out of remission (Here the survival time is the time in remission). If a given patient dies from a heart disease, then that patient's failure time is considered censored. We then know that, for this person, the survival time is at least as long as the period that the person has been followed, but we cannot know in any case the full failure time.

A Type II example is in animal studies. Particularly, in this kind of cohorts there is an initial number of animals, say 100 and the study ceases when a portion of the original animals, say 80 dies, and the rest animals are sacrificed. In this case, if there are no accidental losses, the censored observations equal the twenty largest uncensored observations.

To consider the third Type of censoring, we assume, six patients with acute leukemia enter a clinical study during a total study period of one year. The remission times of the patients varies according to each organism and leukemia type. If a patient get into remission in the beginning of the fifth month and he is still in remission at the end of the study, then the observed survival or censor time for the particular patient is seven months.

2.3 Notation and terminology

In the current section the basic mathematical terminology and notation that is used in the sequel, is introduced.

Let us assume the observed failure time for the *ith* individual as X_i, and the real one as T_i. We suppose also that there is a period of observation c_i such that the observation on that individual ceases after time c_i if failure has not occurred by then. As a consequence $X_i = min(T_i, c_i)$. To complete the notation of the observation we need also an indicator variable $V_i = 1$ if $T_i \leq c_i$ i.e. in uncensored case, and $V_i = 0$ if $T_i > c_i$ i.e. in censored case. Considering T as the nonnegative random variable of the failure times, we then write $\bar{F}_T = pr\,(T > t)$ for the survivor function of T. The failure density function $f(t)$ and the failure cumulative probability $F(t)$ are defined in the continuous case as:

$$f_T\,(t) = -\bar{F}'_T\,(t) = \lim_{h \to 0+} \frac{pr\,(t \leq T < t+h)}{h} \tag{2.1}$$

and so

$$\bar{F}_T\,(t) = \int_0^{\infty} f_T\,(u)\,du \tag{2.2}$$

While in the discrete case just, $pr\,(T \leq t) = F\,(t) = 1 - \bar{F}\,(t)$.

Another function of great importance, for the survival analysis is the hazard function h(t), or the age-specific failure rate. It is defined as the probability of the subject to fail within time t given that it had already survived until the time point t. Specifically,

$$h\,(t) = \lim_{h \to 0} \frac{pr\,(t < T \leq t+h | T > t)}{h} = \lim_{h \to 0} \frac{pr\,(t < T \leq t+h)}{h\bar{F}_T\,(t)} \tag{2.3}$$

In other words the hazard rate function can be written in the continuous case, as:

$$h\,(t) = \frac{f_T\,(t)}{\bar{F}_T\,(t)}, \qquad \text{where } t \geq 0 \tag{2.4}$$

Whereas in the discrete case,

$$h(t) = \frac{pr(T = t)}{pr(T \geq t)} \tag{2.5}$$

The hazard rate function h(t) is anyhow, a specialized characteristic of the data. However, is very useful for the study of the survival time and thus for the failure distribution, if we consider also that usually the information available is about the diachronic evolution of h(t). In this sense, we can choose the functional expression of the hazard rate function for the specific system. For that reason, we end up with a differential equation, or an equation of differences, depending on the type of the random variable. From the univocal relation between the hazard rate function and the failure cumulative function given below, the last can be calculated, in different forms for the continuous and the discrete case. See Dimaki (1995).

2.3.1 Continuous case

From the equation in (2.4), we get that:

$$\frac{d \log \bar{F}_T(t)}{dt} = -h(t) \Rightarrow \log \bar{F}_T(t) - \log \bar{F}_T(0) = -\int_0^t h(p)\, dp \tag{2.6}$$

But from the definition of the survivor function,

$\bar{F}_T(0) = 1 - pr(T \leq 0) = 1 - F(0) = 1 - 0 = 1$ and so $\log \bar{F}_T(0) = 0$

As a sequence the failure density function is given by using the first equation of (2.1) and is:

$$f_T(t) = h(t) \exp(-H(t)) \tag{2.7}$$

or

$$\bar{F}(t) = \exp(-H(t)) \tag{2.8}$$

Where H(t) is the integrated hazard function $\left(H(t) = \int_0^t h(p)\, dp\right)$.

2.3.2 Discrete case

Using equation (2.5) for $t = r$ and $t = r + 1$ respectively, and taking the difference, we get that:

$$pr(T \geq r) - pr(T \geq r+1) = pr(T = r) = \frac{p(T = r)}{h(r)} - \frac{pr(T = r+1)}{h(r+1)} \Leftrightarrow$$

$$pr(T = r+1) - \frac{[1 - h(r)]\, h(r+1)}{h(r)} pr(T = r) = 0 \tag{2.9}$$

Using equation (2.9) recursively, the probability of the failure time to be equal to the integer r is:

$$pr(T = r) = pr(T = 0) \prod_{i=0}^{r-1} \frac{[1 - h(i)]\, h(i+1)}{h(i)}, \qquad r = 0, 1, 2, \ldots \tag{2.10}$$

Assuming, also that $H(t) = \sum_{t_{(j)} < t} (\log(1 - h(j)))$, equation (2.8) is still valid in the discrete case.

Where $t_{(j)}$ is the j^{th} ordered failure time.

Another useful tool, used especially in the likelihood calculation is the risk set, introduced by Cox (1972). Principally, we assume n subjects, observed to failure and k failures. In other words, ties are assumed concerning the failure times, unless k is equal to n. Also, let $m_{(i)}$ be the multiplicity of the failure $t_{(i)}$. Then $\sum m_{(i)} = n$ and $m_{(i)} = 1$, $k = n$ in the continuous case. Assuming the order failure times as $\tau_1, \tau_2, \ldots, \tau_k$ $\left(or\ t_{(1)}, t_{(2)}, \ldots, t_{(k)}\right)$, the risk set is defined as all those subjects whose failure or censoring is at least equal to a specific time point. The definition of the time risk at time $t = \tau_i$ is:

$$\Re(i) = \{j,\ t_j \geq \tau_i, \quad \text{Where } t_j \text{ is the failure time of the random subject j}\}$$

The total number of subjects that belong to risk set $\Re(i)$ is denoted by $r_{(i)}$. Moreover,

we denote by I_j the label of the subjects who fails at τ_j. As a result I_j is equal to the integer i if and only if $\tau_j = t_i$

Chapter 3

Specific Distributions

3.1 Introduction

In this chapter we consider some specific distributions that are useful for survival data. Even though any non negative random variable is a possible candidate for the distribution of T, or even further any distribution including also negative real values is a likely alternative for $\log(T)$ distribution. For analysis purposes the most used are included in the current thesis and presented in the Tables (3.1), (3.2), below. The Survivor function, as well with the Hazard rate function are displayed (whenever an explicit form exists) in both continuous and the discrete case. Especially, in the continuous case the various distributions can be classified according to whether they are over or underdispersed relative to the exponential distribution, which has a constant hazard rate. Greek letters are used for the parameters of the distributions. Thus ρ will denote the reciprocal of time and can be interpreted as a rate. Other letters like κ and τ will be treated as dimensionless parameters.

3.2 Continuous Distributions

3.2.1 Exponential Distribution

Definition

A continuous random variable T with Survivor function

$$\bar{F}_T(t) = e^{-\rho t}, t \geq 0, \rho > 0 \tag{3.1}$$

is **exponentially distributed** with parameter the Greek letter ρ. Using mathematical notation, we write that $T \backsim \exp(\rho)$.

The cumulative function is given by

$$F_T(t) = 1 - \bar{F}_T(t) = 1 - e^{-\rho t}\ ,\ t \geq 0,\ \rho > 0 \tag{3.2}$$

and the known density failure function is given by the minus derivative of the survivor function (3.1). That is

$$f_T(t) = -\bar{F}'_T(t) = \rho e^{-\rho t}, t \geq 0, \rho > 0 \tag{3.3}$$

The Hazard ratio is given by the formula (2.34) illustrated in the previous chapter. Otherwise

$$h(t) = \frac{f_T(t)}{\bar{F}_T(t)} = \rho \tag{3.4}$$

Properties

- The last formula (3.4) just illustrates that the hazard rate of an exponential failure distribution is a constant. This condition is necessary and capable to ensure that any non negative random variable T is exponentially distributed. Both the necessity and capability arise from the univocal relation between the hazard rate function

and the failure cumulative function.

Distribution Name	Survivor Function	Density Function	Hazard
Exponential	$e^{-\rho t}$	$\rho e^{-\rho t}$	ρ
Gamma	$\int_t^\infty \frac{\rho(\kappa\rho)^{y-1}e^{-\rho y}}{\Gamma(\kappa)}dy$	$\rho(\rho t)^{\kappa-1}e^{(-(\rho t)^\kappa)}$	$\frac{\rho(\rho t)^{\kappa-1}e^{(-\rho t)}}{\Gamma(\kappa)[1-I_\kappa(\rho t)]}$
Weibull	$e^{-(\rho t)^\kappa}$	$\kappa\rho(\rho t)^{\kappa-1}e^{-(\rho t)^\kappa}$	$\kappa\rho(\rho t)^{\kappa-1}$
Compertz-Makeham	$e^{\left(\frac{\rho_1}{\rho_2}-\rho_0 t-\frac{\rho_1}{\rho_2}e^{\rho_2 t}\right)}$	$(\rho_1 e^{\rho_2 t}+\rho_0)\bar{F}(t)$	$\rho_0+\rho_1 e^{\rho_2 t}$
Compound exp/tial	$\frac{(\kappa/\rho_0)^\kappa}{(t+\kappa/\rho_0)^\kappa}$	$\frac{\kappa(\kappa/\rho_0)^\kappa}{(t+\kappa/\rho_0)^{\kappa+1}}$	$\frac{\kappa}{t+\kappa/\rho_0}$
Log normal	—	$\frac{1}{\sqrt{2\pi\sigma^2 t^2}}e^{\left(-\frac{(\log t-\mu)^2}{2\sigma^2}\right)}$	nonmonotonic
Log logistic	$(1+(\rho t)^\kappa)^{-1}$	$\kappa\rho^\kappa t^{\kappa-1}(1+(\rho t)^\kappa)^{-2}$	$\frac{\kappa\rho^\kappa t^{\kappa-1}}{(1+(\rho t)^\kappa)}$
Inverse Gussian	—	$\sqrt{\frac{\kappa/\rho}{2\pi t^3}}e^{\left(-\frac{\kappa\rho(1-1/\rho)^2}{2t}\right)}$	—
Scale family	$\ddot{F}(\rho t)$	$\rho\ddot{g}(\rho t)$	$\rho\ddot{h}(\rho t)$
Proportional hazard	$\left(\ddot{F}(t)\right)^\rho$	$\rho\left(\ddot{F}(t)\right)^{\rho-1}\ddot{g}(t)$	$\rho\ddot{h}(t)$

Table 3.1: Survivor Density and Hazard functions in the Continuous case

Distribution Name	Survivor Function	Density Function	Hazard
Geometric distribution	$(1-p)^{t+1}$	$p(1-p)^t$	ρ
Yule distribution	$\frac{t+1}{\rho}pr(T=t)$	$\frac{\rho t!}{(\rho+1)(\rho+2)\cdots(\rho+t+1)}$	$\frac{\rho}{\rho+t+1}$

Table 3.2: Survivor Density and Hazard functions in the Discrete case

- In case the survival or else failure time follows the exponential distribution with parameter $\rho > 0$, it can be shown that the conditional probability of the failure time T exceeds the $\tau + x$, given that $T > \tau$, is equal to the unconditional probability of $T > x$. The last property is called **lack of memory** and it defines univocal the distribution of the failures to be the exponential. The property of lack of fit allows the use of the exponential distribution for the description of the life time of a system when there is no actual loss in the system due to the passage of time. Nevertheless, in the framework of survival analysis this situation is unreal as it actual accepts that the working time does not result to damage in the survival time. There are cases although where this situation is found, like in reliability theory when analysis focuses in the life time control with replacement.

Applications

The exponential distribution was widely used in early work on reliability of, for example, electronic components and to a more limited extent in medical studies. Only one adjustable parameter is essential for the exponential distribution. Thus the methods based on it are not robust, even in modest variations from the real failures, for instance departure from the tail of the distribution. For that reasons the methods used are based on less stringent assumptions about the distributional form. Various idealized models lead to the exponential distribution.

Genesis Schemes

- Assume $\{X_t,\ t \geq 0\}$ be a Poisson stochastic process with rate ρ and $T_0 = 0$, T_1, $T_2, \ldots$ be the sequence of the time points in which failures occur. Then for every integer $n \geq 0$, the sequence of the intermediate time between two events, $T_1 - T_0$, $T_2 - T_1, \ldots$ is a sequence of independent and identical exponential that do not depend on the specific time points T_1, $T_2, \ldots$ in which the events occur.

- When the hazard rate function is a constant $h(t) = r, \quad r > 0$ then the distribution of the random variable T which describes the failure time of a subject has survivor function equal to the (3.2) equation by replacing the parameter ρ with the constant r. In other words $T \sim \exp(r)$.

- The next two parameter families of distributions can be reduced into the exponential distribution, by letting only one parameter to vary.

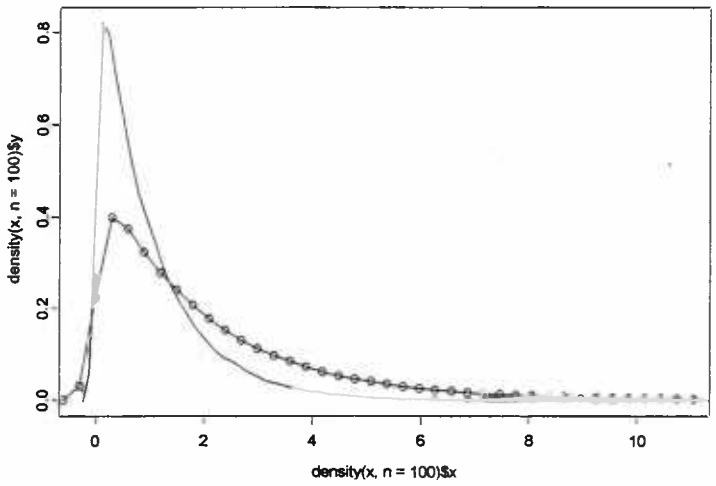

Figure 3-1: Exponential density functions. Solid line: Mean=1 and Line with circles: Mean=2

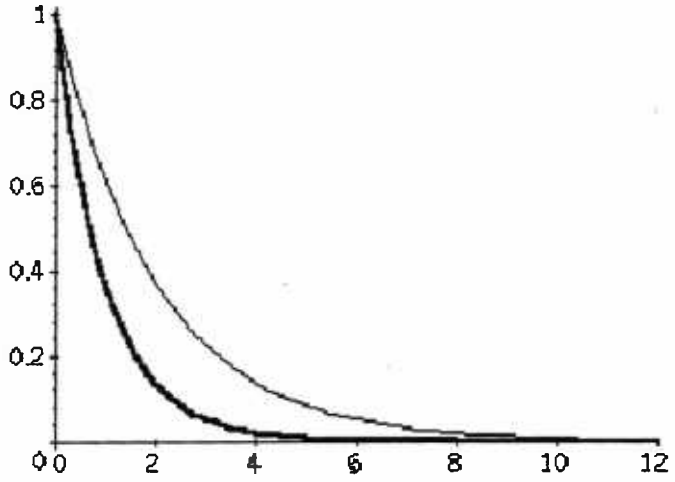

Figure 3-2: Exponential survival time. Bold line: Mean=1 and Normal line: Mean=2

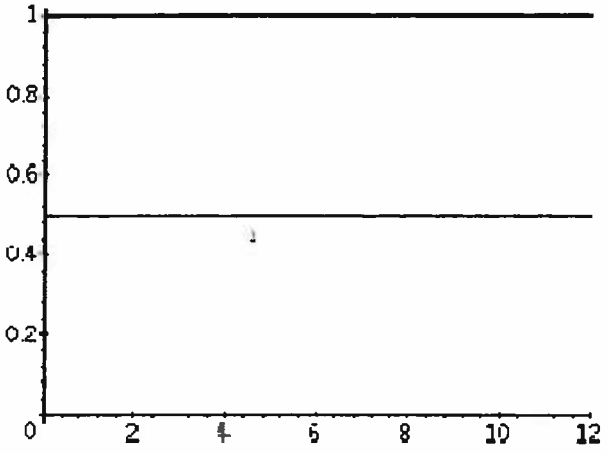

Figure 3-3: Exponential hazards. Bold line: Mean=1 and Normal line: Mean=2

3.2.2 Gamma Distribution

Definition

The continuous random variable T follows the **Gamma distribution** with parameters ρ and an extra one $\kappa > 0$ if the survivor function is given by

$$\bar{F}_T(t) = \int_t^{\infty} \frac{\rho(\kappa\rho)^{y-1}e^{-\rho y}}{\Gamma(\kappa)} dy \tag{3.5}$$

where $\Gamma(\kappa)$ is the Gamma function defined by the equation

$$\Gamma(x) = \int_0^{\infty} e^{-y}y^{x-1}dy, \quad t > 0 \tag{3.6}$$

The density function of T is given by

$$f(t) = \frac{e^{-t\rho}\rho(\rho t)^{\kappa-1}}{\Gamma(\kappa)} \tag{3.7}$$

By putting κ equal to unity, the density function of (3.7) reduces to the exponential density. In addition when $\rho = 1$ the distribution derived from (3.7) is called the standardized Gamma. In this spot, the cumulative function of the standardized Gamma is given by

$$I_\alpha(t) = F_T(t) = \int_0^t \frac{1}{\Gamma(\alpha)} e^{-y}y^{\alpha-1}dy, \quad \text{t,y>0} \tag{3.8}$$

Properties

The hazard rate is given by the use of the last equation (3.8) and equals to

$$h(t) = \frac{\rho(\rho t)^{\kappa-1}e^{(-\rho t)}}{\Gamma(\kappa)[1 - I_\kappa(\rho t)]} \tag{3.9}$$

- The function $I_\kappa(t)$ is called the defective Gamma function and its values are usually contained in tables. These values can be used along with the following first remark to evaluate the probabilities of unstandardized Gamma distributions.

Remark 1 *If* $T \sim Gamma(\kappa, \rho)$ *then* $p(T \leq t) = I_{\kappa}(\rho t)$

- The hazard function of the Gamma distribution either decreases for $\rho < 1$ (see the thin curve in Figure 3-6), or increases for $\rho > 1$. In case $\rho = 1$ the exponential distribution arises, which was explored previously (Section 3-2).

Genesis Schemes

- Assume that T is the waiting time until an event happens, in a Poisson process with parameter ρ. Then T$\sim Gamma(\kappa = n, \rho)$.

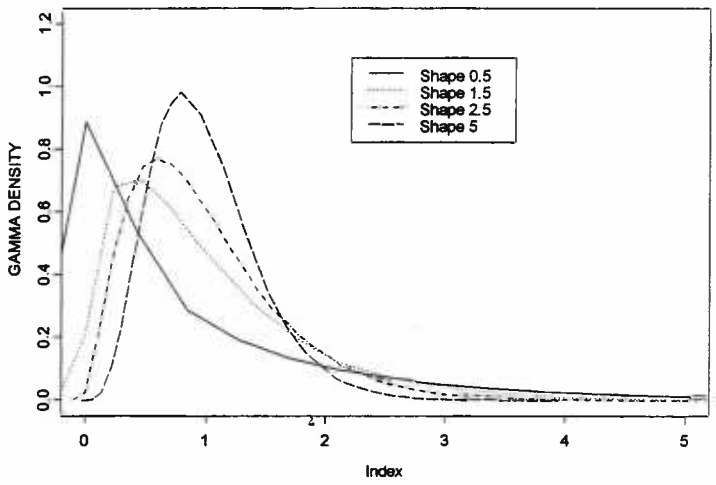

Figure 3-4: Gamma density functions all with mean equals 1 and different shapes

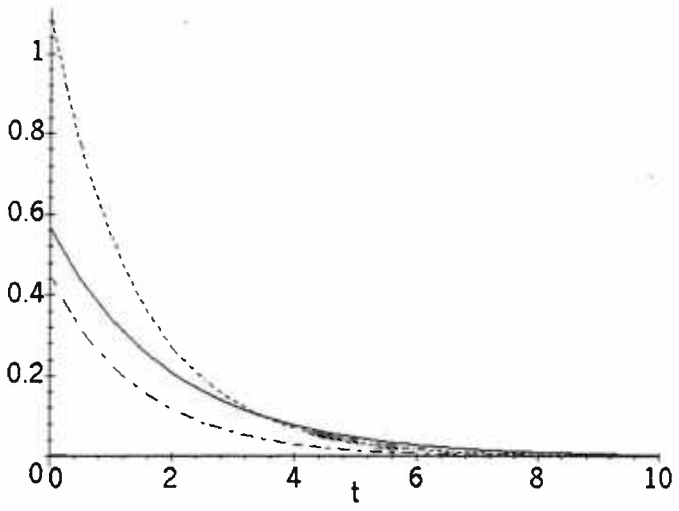

Figure 3-5: Gamma survival time all with mean equals 1 and different shapes

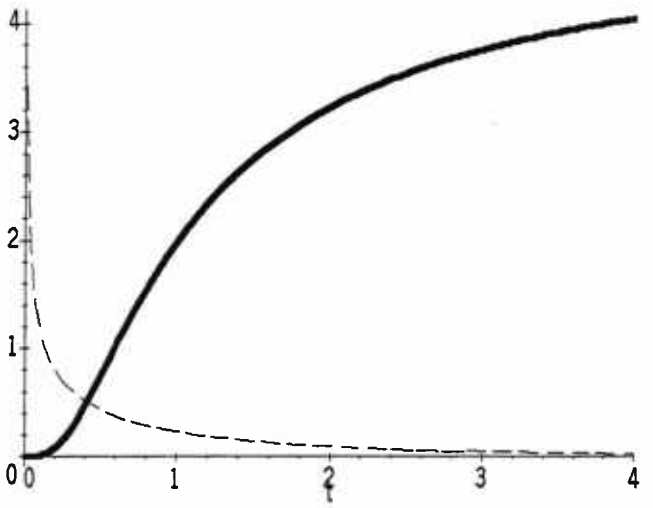

Figure 3-6: Gamma hazards for $\rho = 0.5$ (thin line) & $\rho = 5$

3.2.3 Weibull Distribution

Definition

The continuous random variable T with survivor function

$$\bar{F}_T(t) = \exp[-(\rho t)^\kappa], \quad t \geq 0, \ \kappa, \rho > 0 \tag{3.10}$$

follows the **Weibull distribution** with parameters $\rho, \kappa > 0$. The density function of the failure T is given by

$$f_T(t) = \kappa\rho(\rho t)^{\kappa-1} e^{-(\rho t)^\kappa} \tag{3.11}$$

There is a univocal relation between the Weibull and the Exponential distribution given in the second Remark,

Remark 2 *If* $T \sim \exp(\rho)$ *then the random variable* $Y = T^{\frac{1}{\kappa}} \sim Weibull(\rho^{\frac{1}{\kappa}}, \kappa)$

The last remark make Weibull a useful tool not only in Survival analysis but in the reliability theory, as well.

Properties

The hazard rate is then equal to,

$$h_T(t) = \kappa\rho(\rho t)^{\kappa-1} \tag{3.12}$$

- The non negative random variable T follows the Weibull(ρ, κ) if and only if the age-specific or else hazard rate is given by the equation (3.12), where ρ, κ are positive constants.
- Similar to the Gamma distribution, the hazard rate of the Weibull decreases for $\kappa < 1$ and increases for k>1. For k=1 we have the exponential case. Also, for k>2 we have that the hazard increases faster than linearity (Figure 3-9).

Genesis Schemes

- If the hazard rate evaluated in the time point t $h_T(t)$ is a power function of time then the distribution of the random variable T, which describes the life time of a

subject, is Weibull, defined as in (3.11).

- The asymptotic distribution of the smallest ordered statistical function from a predetermined r.v. is proved to be the Weibull distribution.

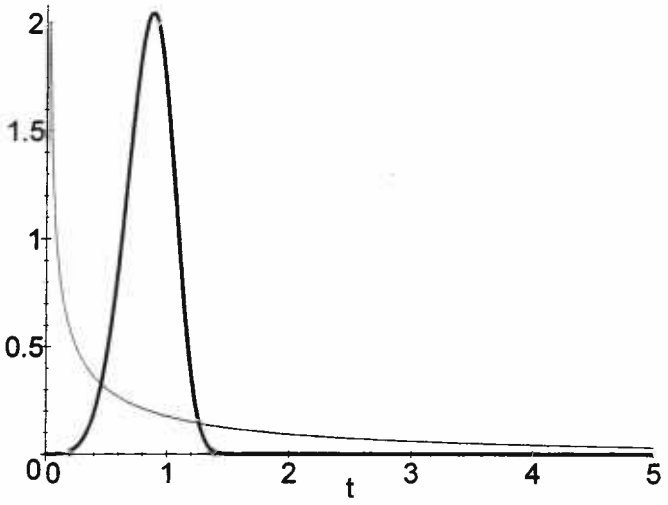

Figure 3-7: Weibull density curves with k =0.5 (thin line) and k=5 (dark line)

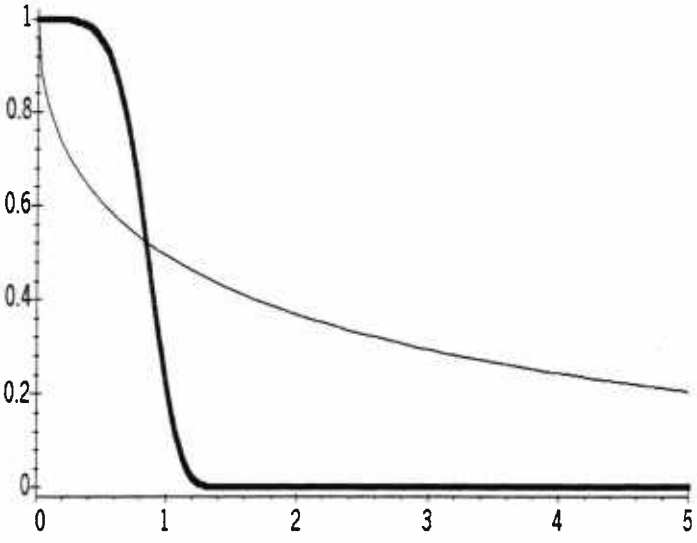

Figure 3-8: Weibull survival time with k =0.5 (thin line) and k=5 (dark line)

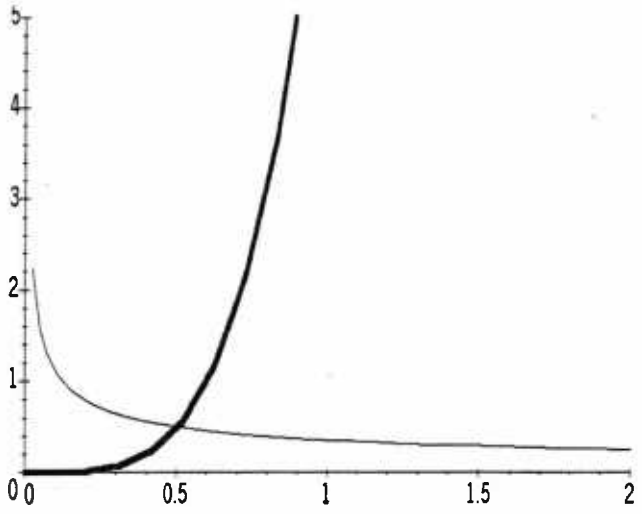

Figure 3-9: Weibull hazards with k =0.5 (thin line) and k=5 (dark line)

3.2.4 Gompertz-Makeham Distribution

Definition

The hazard function using **Gompertz-Makeham distribution** as the distribution of the failures T is

$$h_T(t) = \rho_0 + \rho_1 e^{\rho_2 t} \tag{3.13}$$

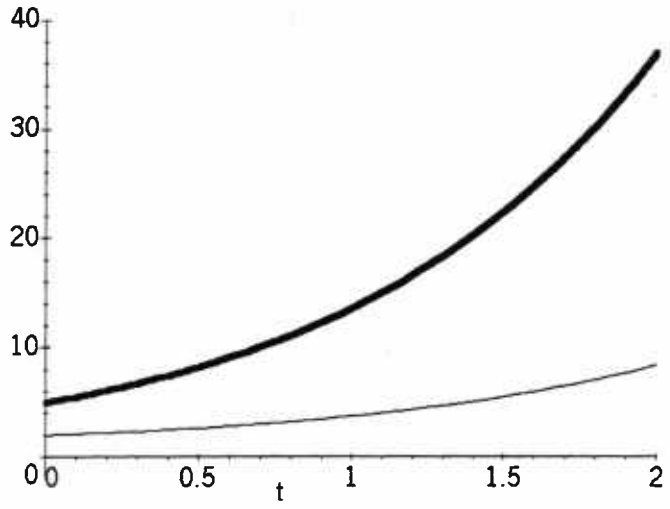

Figure 3-10: Gompertz-Makeham hazards with k= 0.5 (thin line) and k=5 (dark line)

Putting $\rho_0 = 0$, we have the Gompertz form of the distribution.

The survivor function $\bar{F}$ is given by using (2.8) and equals to

$$\bar{F}(t) = \exp\left(\frac{\rho_1}{\rho_2} - \rho_0 t - \frac{\rho_1}{\rho_2} e^{\rho_2 t}\right) \tag{3.14}$$

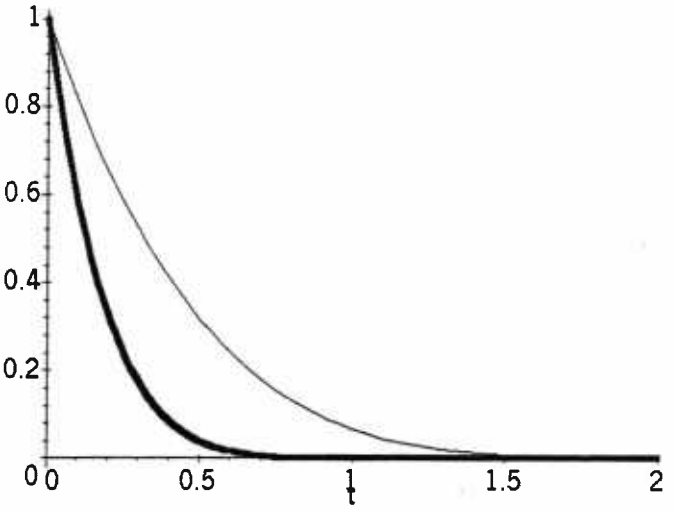

Figure 3-11: Gompertz-Makeham survival time with k= 0.5 (thin line) and k=5 (dark line)

Finally the density function of the failures is

$$f(t) = (\rho_1 e^{\rho_2 t} + \rho_0) \exp\left(\frac{\rho_1}{\rho_2} - \rho_0 t - \frac{\rho_1}{\rho_2} e^{\rho_2 t}\right) \tag{3.15}$$

3.2.5 Compound Exponential or Pareto Distribution

Definition

The continuous random variable T with survivor function

$$\bar{F}_T(t) = \frac{(\kappa/\rho_0)^\kappa}{(t + \kappa/\rho_0)^\kappa} \tag{3.16}$$

follows a **Compound exponential distribution or Pareto** with parameters ρ_0 and κ The density of the T random variable is now given by

$$f_T(t) = \frac{\kappa(\kappa/\rho_0)^\kappa}{(t + \kappa/\rho_0)^{\kappa+1}} \tag{3.17}$$

Properties

- The hazard rate function which arises from the (2.1) (3.17) formulas, is

$$h_T(t) = \frac{\kappa}{t + \kappa/\rho_0} \tag{3.18}$$

Some Pareto hazards are plotted in the Fig 3-14.

- As follows directly from the way of its construction, the current distribution is overdispersed relative to the exponential distribution, to which it tends as $\kappa \longrightarrow \infty$. When κ is small (3.17) has a very long tail (see Fig 3.12).

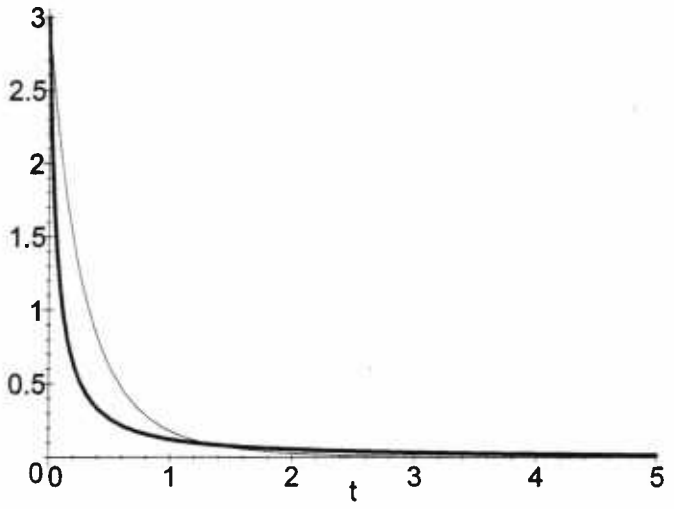

Figure 3-12: Pareto density curves for ρ_0=3 and κ equals to 0.5 (dark line) and 5 (thin line)

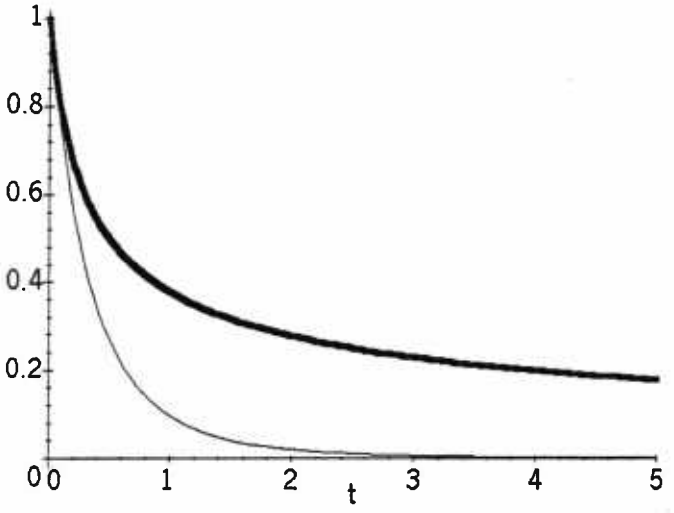

Figure 3-13: Pareto survival time for ρ_0=3 and κ equals to 0.5 (dark line) and 5 (thin line)

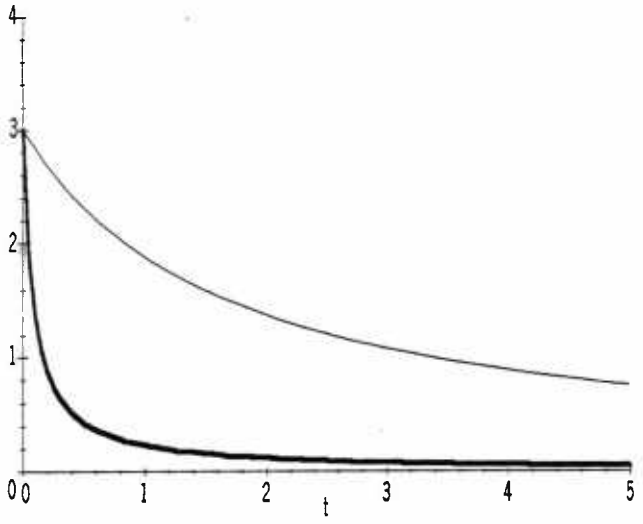

Figure 3-14: Pareto hazards for ρ_0=3 and κ equals to 0.5 (dark line) and 5 (thin line)

Genesis Schemes

- Let us assume that for each individual survival time is exponentially distributed as in section 3.1. In addition, considering this time that the ratio varies randomly between individuals, the conditional distribution of T given $P = \rho$ is

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

Then the unconditional density of failures T is

$$f_T(t) = \int_0^\infty \rho e^{-\rho t} f_p(\rho) d\rho \tag{3.19}$$

Assuming that $f_p(\rho)$ is coming from the Gamma distribution with mean ρ_0 and index κ, the density in (3.19) is reduced to the density of a Pareto distribution in (3.17).

3.2.6 Log-Normal Distribution

Definition

One of the most commonly used distribution is the Normal distribution with density function

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{3.20}$$

Since the Normal distribution allows negative value (3.20), a plausible way of using it in Survival analysis is to take $logT$ normally distributed. This is equivalent to assuming a **lognormal distribution** for the failure times. The corresponding density is defined in the positive domain of Real numbers and given by

$$f_T(t) = (2\pi\sigma^2 t^2)^{-1/2} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right) \tag{3.21}$$

In both (3.20) and (3.21), μ and σ are the mean and variance of the Normal distribution, i.e. $\log(T) \sim N(\mu, \sigma^2)$. The mean and variance of the lognormal distribution are $\exp(\mu + \frac{1}{2}\sigma^2)$ and $\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$ respectively. The survivor and hazard functions of the lognormal distribution cannot be written explicitly, but only in terms of integrals. For small values of σ, the lognormal density looks very like a Normal one. We can notice this in the Fig 3-15, where the solid thin line, which is closer to a Normal density, has the smaller σ. Also, as the standard deviation increases, the density curve has greater tails and departure from the that of a Normal distribution.

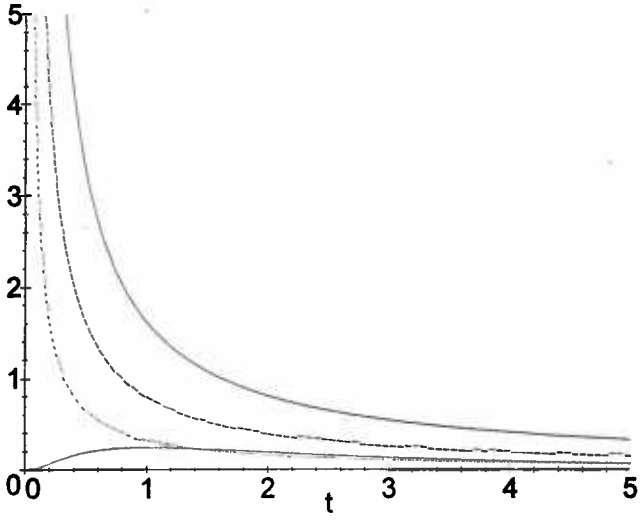

Figure 3-15: Log Normal densities with mean 1 and σ =0.25,0.5,1 and 1.25

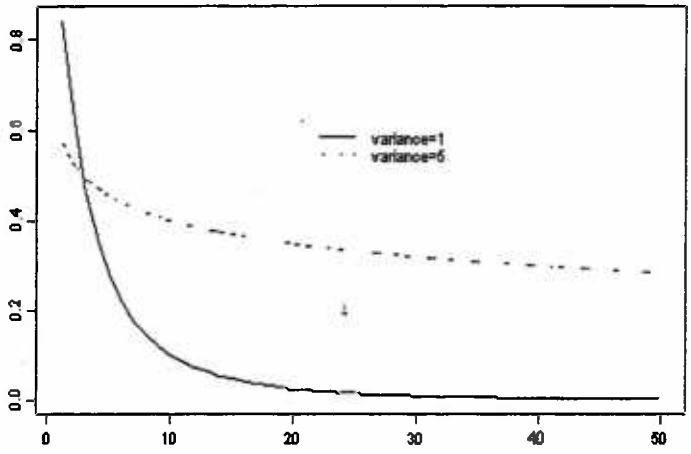

Figure 3-16: Log Normal survival time with mean equals 1 and σ^2 =1 (solid line) and σ^2 =5 (dashed line)

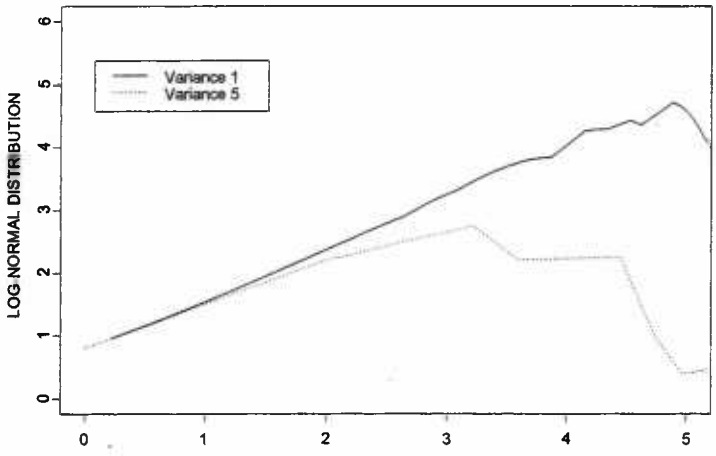

Figure 3-17: Log Normal hazards with mean equals 1 and $\sigma^2 = 1$ (solid line) and $\sigma^2 = 5$ (dashed line)

Properties

- The hazard associated with (3.21) is non monotonic. Particularly, considering the hazard functions of some lognormal distributions, they are initially increasing, but eventually decreasing, tending to zero. Fig 3-17 includes a couple of simulated hazard curves, where the solid one has variance 1, whereas the dashed one has a greater variance equals to 5. This behavior is somewhat counter to what is usually expected of lifetimes in practice. However, the lognormal failure is sensitive when predictions are based on small values.

- Even if, the exponential distribution is not a special case of the lognormal one, a substantial amount of data is necessary to discriminate empirically between them, especially in the case when we have that $T \sim \log N(\exp(\mu + \frac{1}{2}\sigma^2), 0.64)$.

3.2.7 Log Logistic Distribution

Definition

As given in Cox & Oakes (1984), the continuous logistic density $logit(\nu, \tau)$ is very similar to a normal distribution. As before an appropriate distribution for the failure T random variable is the **log logistic family**, obtained as previously in the log normal case. Hence, the survivor function, the density and the hazard function become respectively

$$\bar{F}_T(t) = (1 + (\rho t)^{\kappa})^{-1} \tag{3.22}$$

$$f_T(t) = \kappa \rho^{\kappa} t^{\kappa - 1} (1 + (\rho t)^{\kappa})^{-2} \tag{3.23}$$

$$h(t) = \frac{\kappa \rho^{\kappa} t^{\kappa - 1}}{(1 + (\rho t)^{\kappa})} \tag{3.24}$$

Where ρ and κ are related with ν and τ respectively by the equations: $\exp(\nu) = \rho^{-1}$ and $\kappa = \frac{1}{\tau}$

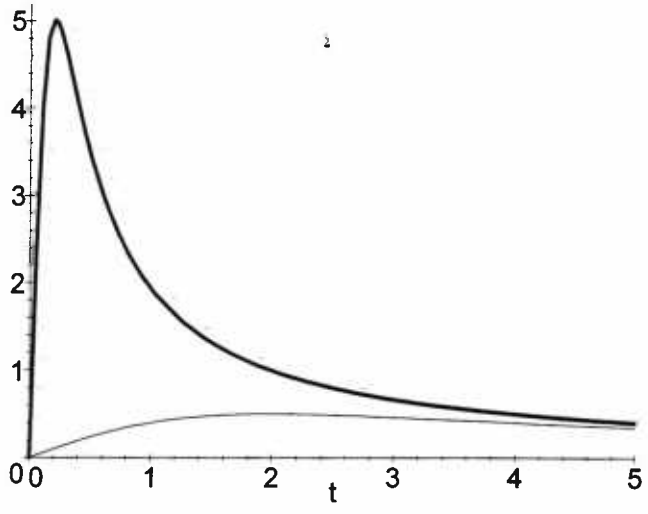

Figure 3-18: Log logistic densities for k=2 and ρ equals 0.5 (thin line) and 5 (thin line)

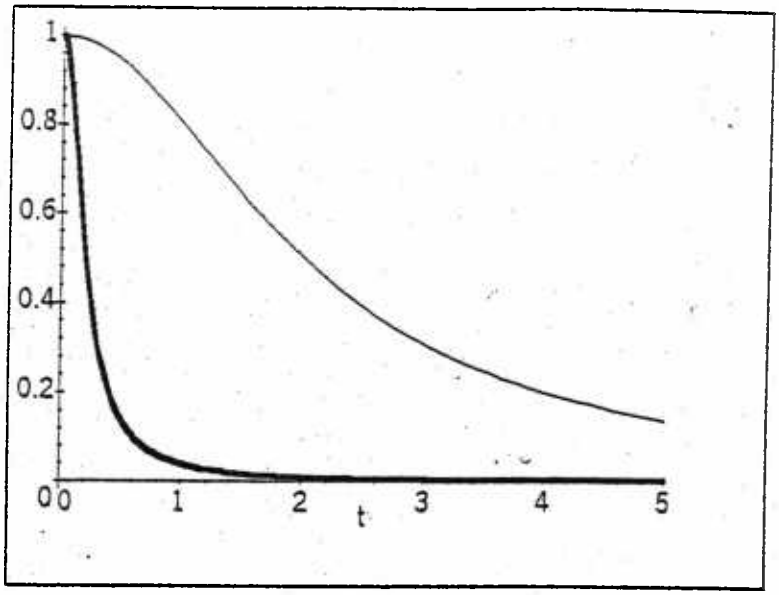

Figure 3-19: Log logistic survival time for k=2 and ρ equals 0.5 (thin line) and 5 (dark line)

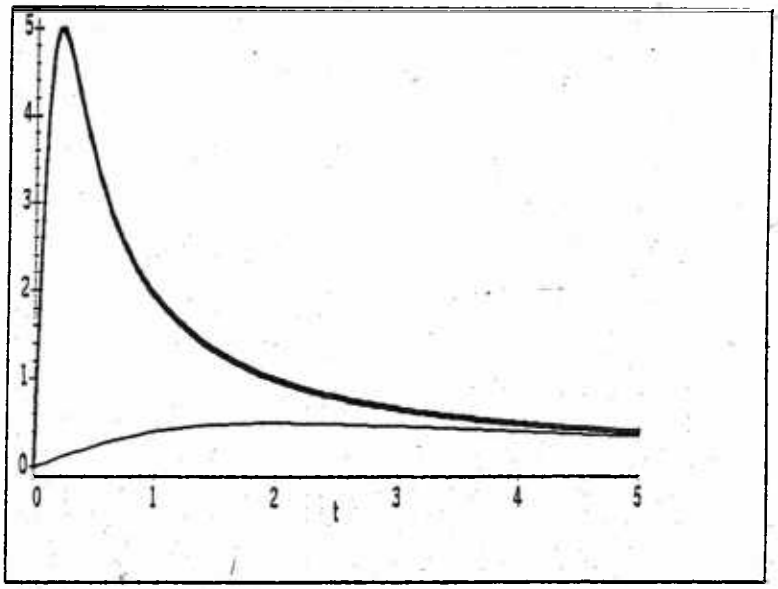

Figure 3-20: Log logistic hazards for k=2 and ρ equals 0.5 (thin line) and 5 (dark line)

Properties

- Comparing the last two models, the log logistic has relatively simpler forms achieved for $\bar{F}_T(t)$, $f_T(t)$, $h(t)$. If $\kappa > 1$ the hazard has a single minimum. On the other hand

when $\kappa < 1$ the hazard is decreasing.

- A condition for the existence of the **rth** moment is the parameter κ to be greater than r.

3.2.8 Generalized F Distribution

Definition

This particular distribution is obtained by taking T to be a multiple of the κ_1th power of a random variable $F_{(\kappa_2,\kappa_3)}$ having the standard (central) variance ratio distribution with (κ_2, κ_3) degrees of freedom. In other words

$$T = \rho^{-1} F^{\kappa_1}_{(\kappa_2,\kappa_3)} \tag{3.25}$$

The three-parameter generalized gamma family actually involves when $\kappa_3 \to \infty$. Many of the above distributions can be derived from the generalized F by adjusting the dimensionless parameters $(\kappa_1, \kappa_2, \kappa_3)$.

For an example we simulated times from the following expressions of the F distribution

$$T = \begin{cases} \frac{1}{2} F^2_{(3,2)} \\ F^2_{(3,2)} \end{cases}$$

The following survival curves $\bar{F}_T(t)$ are obtained. Both curves decrease and tend to zero in an exponential rate.

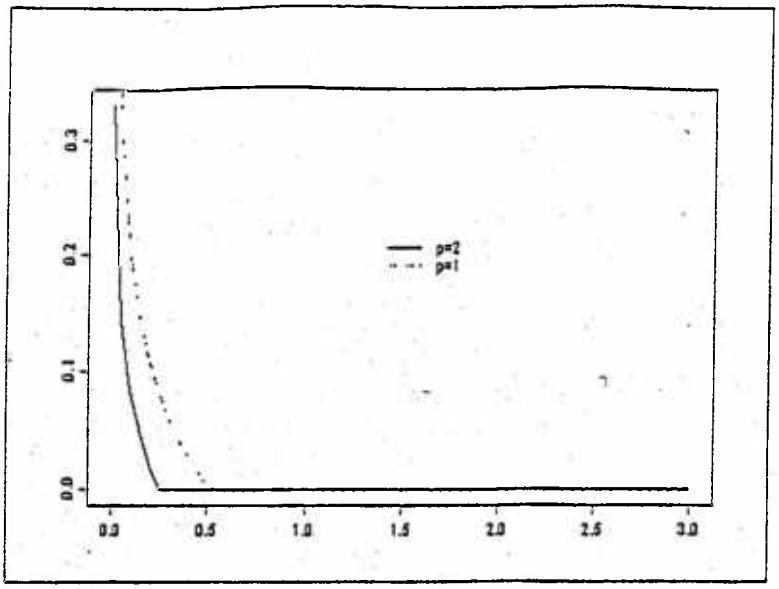

Figure 3-21: Generalized F survival time for k_1, k_3=2 k_2=3 and ρ=2 (solid line) and 1 (dashed line)

3.2.9 Inverse Gaussian Distribution

Definition

Considering the **Inverse Gaussian distribution**, the failure time random variable T will have density function

$$\left(\frac{\kappa/\rho}{2\pi t^3}\right)^{1/2} \exp\left(-\frac{\kappa\rho(1-1/\rho)^2}{2t}\right) \tag{3.26}$$

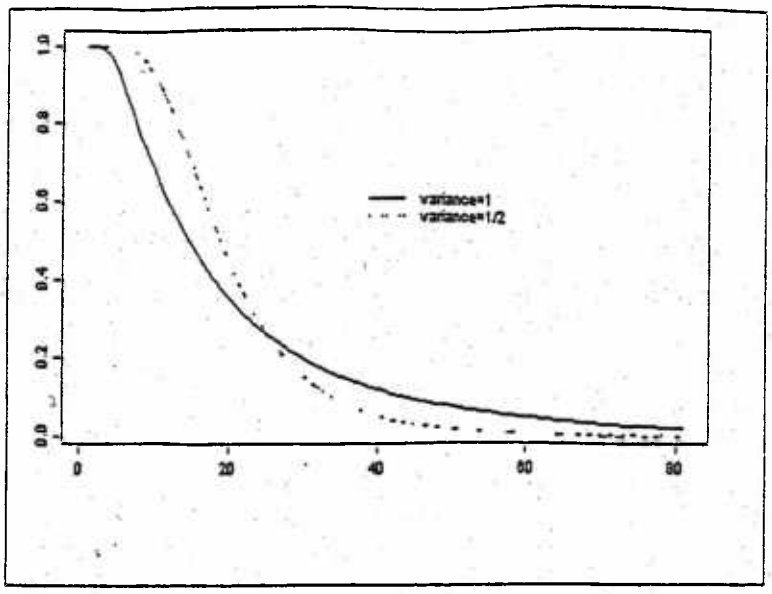

Figure 3-22: Inverse Gaussian survival time for ρ=1 and κ=1 (solid line) and 4 (dashed line)

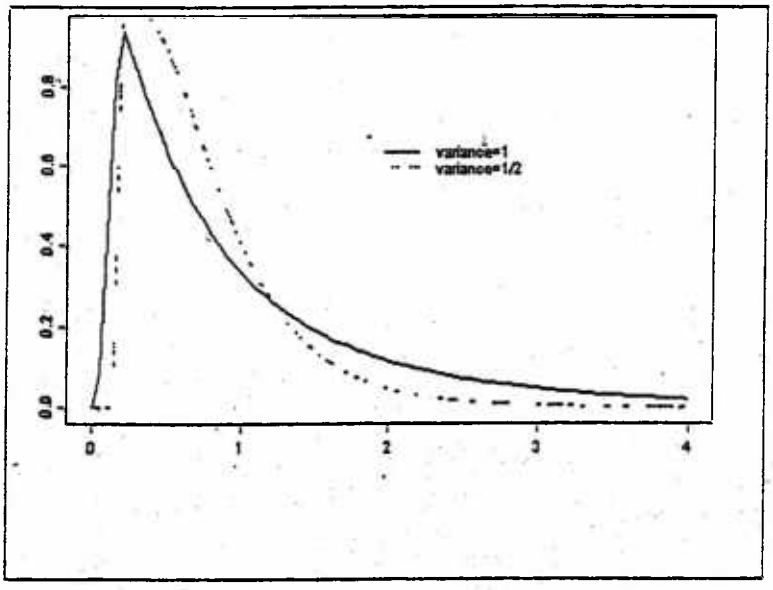

Figure 3-23: Inverse Gaussian hazards for ρ=1 and κ=1 (solid line) and 4 (dashed line)

with mean $1/\rho$ and variance $1/\sqrt{\kappa}$. The survivor function has the complicated form

$$1-\Phi\left[\left(\frac{\kappa}{\rho t}\right)^{1/2}(-1+\rho t)\right]-e^{2\kappa}\Phi\left[-\left(\frac{\kappa}{\rho t}\right)^{1/2}(1+\rho t)\right] \tag{3.27}$$

Where $\Phi(.)$ is the standardized normal cumulative function.

Applying (3.27) with mean equals to unity (or else ρ =1) and κ=$\{^{1}_{4}$ the curves in Fig. 3-22 are obtained. As well, the hazard curves Fig. 3-23 follow by using the relation between the survival function and the hazard ratio in (2.8).

Genesis Schemes

- Considering the stochastic process of the Brownian motion, the first passage time to a barrier has the inverse Gaussian distribution.

3.2.10 Scale Family

Definition

Suppose that $\ddot{F}$, $\ddot{g}$ and $\ddot{h}(t)$ denote respectively a survival function, density and hazard over non-negative values, the corresponding functions

$$\left.\begin{aligned} \ddot{F}(t;\rho) &= \ddot{F}(\rho t) \\ \ddot{g}(t;\rho) &= \rho\ddot{g}(\rho t) \\ \ddot{h}(t;\rho) &= \rho\ddot{h}(\rho t) \end{aligned}\right\} \tag{3.28}$$

define the **scale family** generated by the $\ddot{F}$(t)

3.2.11 Proportional Hazard Family

Definition

Finally, another useful general family is generated from the survivor function, density and hazard $\ddot{F}$, $\ddot{g}$ and $\ddot{h}(t)$, as before. Specifically, the relations are used

$$\left.\begin{aligned} \ddot{F}(t;\rho) &= \left(\ddot{F}(t)\right)^{\rho} \\ \ddot{g}(t;\rho) &= \rho\left(\ddot{F}(t)\right)^{\rho-1}\ddot{g}(t) \\ \ddot{h}(t;\rho) &= \rho\ddot{h}(t) \end{aligned}\right\} \tag{3.29}$$

This is called the **Lehmann or proportional hazards family** based on $\ddot{F}(t)$. The families (3.28) and (3.29) are equivalent if and only if $\ddot{h}(t) \propto t^q$ for some q, so that both represent Weibull distributions.

3.3 Discrete Failure Distributions

3.3.1 Geometric Distribution

Definition

A discrete random variable T with survivor function

$$\bar{F}_T(t) = (1-p)^{t+1} \tag{3.30}$$

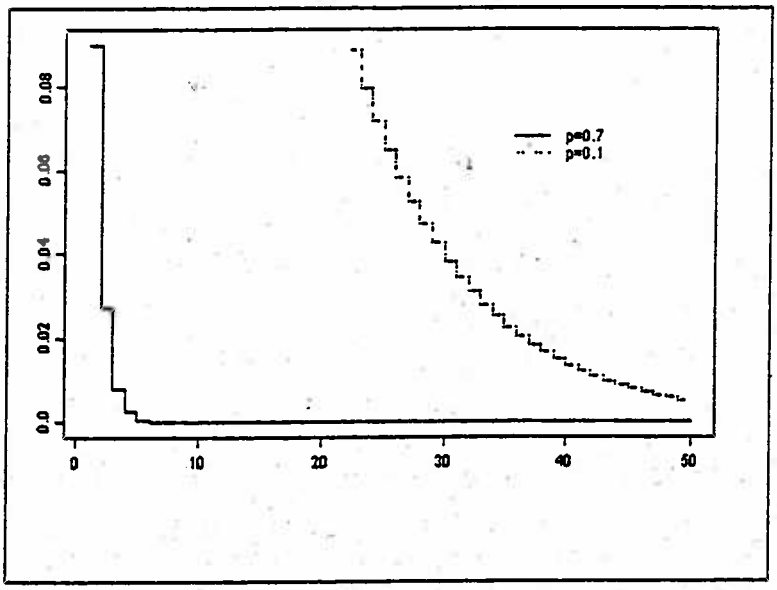

Figure 3-24: Geometric Survival times for ρ=0.7 (solid line) and 0.1 (dashed line)

is said to follow the **geometric distribution** with parameter p. The function in (3.30) is defined on the positive integers, while p$\in$ $(0,1)$. By notation T$\sim$ $G(p)$. The probability function is given by

$$p(T = t) = p(1-p)^t \tag{3.31}$$

given the previous restrictions for p and t.

The example step survivors Fig. 3-24 are derived by applying (3.30) with probability 0.7 and 0.1 respectively. The corresponding hazards are constant and equals the probability values as we will see.

Properties

- The hazard rate of a discrete failure distribution T, with non-negative values, is constant equals to p if and only if T follows the geometric distribution. This condition arrives from the univocal relation between the hazard rate function and the failure cumulative function, given in (3.10).

- The relation $h_T(t)\mu^T(t) = 1$ defines univocal the distribution of T to be geometric with parameter p>0. The function $\mu^T(t)$ is called the mean residual life at time t and defined as the expected value $E(T - t|T > t)$, where t=0,1,2,.... See Dimaki (1995).

- An important property of the geometric distribution is the **lack of the cumulative memory**, which is analogous to the property of the exponential one. In particular, the conditional probability $pr(T > t + y|T > t)$ is equal to the unconditional probability $pr(T > y)$.

Genesis Schemes

- We assume a sequence of Bernoulli trials. Let us also denote as T the number of failures until the first success. Then T follows a geometric distribution with the survivor function given in (3.30).

- When the hazard function $h(t) = p, \quad p > 0$ then the distribution of the discrete random variable which describes the lifetime of a subject has the survivor function given in (3.30).

3.3.2 Yule Distribution

Definition

A discrete random variable T with survivor function

$$\bar{F}_T(t) = \frac{t+1}{p} pr(T = t) \tag{3.32}$$

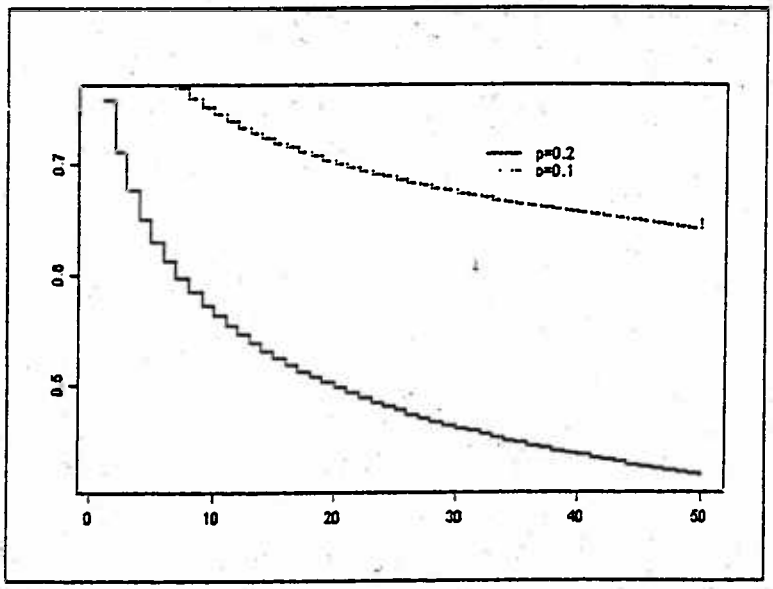

Figure 3-25: Yule survival times for ρ equals 0.2 (solid line) and 0.1 (dashed line)

is said to follow the **Yule distribution** with parameter p (see Xekalaki (1983a), (1983b) and Dimaki (1995)). The function in (3.32) is defined on the positive integers, i.e. t=0,1,2,... $p > 0$. By notation T$\sim Yule(p)$. The probability function is given by

$$p(T = t) = \frac{\rho t!}{(\rho+1)(\rho+2)\cdots(\rho+t+1)} \tag{3.33}$$

given the previous restrictions t.

The hazard manipulation is as follows

From the equation (2.5) we have that $h(t) = \frac{pr(T=t)}{pr(T \geq t)} = \frac{pr(T=t)}{pr(T>t)+pr(T=t)}$ and using the survivor function (3.32), the last equation becomes $h(t) = \frac{pr(T=t)}{\frac{t+1}{\rho} pr(T=t)+pr(T=t)} = \frac{\rho}{\rho+t+1}$

$$h(t) = \frac{\rho}{\rho + t + 1} \tag{3.34}$$

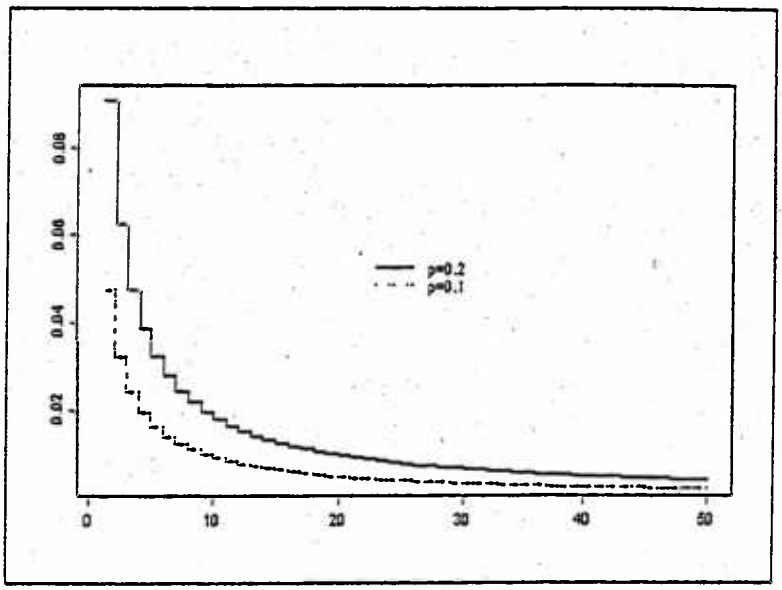

Figure 3-26: Yule hazards for ρ equals 0.2 (solid line) and 0.1 (dashed line)

Properties

- The hazard rate of a discrete failure distribution T, at time t, is inversely proportional to the time t if and only if T follows the Yule distribution.
- The relation $h_T(t)\mu^T(t) = c$ defines univocal the distribution of T to be the Yule with parameter p>0.

Chapter 4

Popular and Recent Approaches

4.1 Introduction

In the current chapter, we present some survival analysis methods that are used commonly in the literature. Especially, we firstly consider the Kaplan-Meier estimator for the survival function. Secondly, the Accelerated Life Model and the Proportional Odds Model are given along with some numerical examples. Finally, Additive risk Models are included in the last part of the chapter, as presented by several authors.

4.2 The Kaplan-Meier or Product-limit estimate

A non-parametric estimate of the survivor function in the case of any right-censored sample, is the product-limit (PL) or Kaplan-Meier estimate (Kaplan Meier (1958)).

Assuming once again n subjects under study and κ failures to occur. Also let $m_{(i)}$ be the number of failures at time τ_i, where $\tau_1 < \tau_2 < \ldots < \tau_k$ are the ordered failure times. We recall endmost the number r_j as the total number of individuals in risk at τ_i. The Kaplan-Meier estimator of the survivor function, say $S(t)$, is

$$\hat{S}(t) = \prod_{j \in \{j:\ \tau_j < t\}} \left(1 - \frac{m_{(j)}}{r_j}\right) \tag{4.1}$$

In case of uncensored observations, equation (4.1) reduces to the simple expression

$$\hat{S}(t) = \frac{\#\text{ surviving past t}}{n} \tag{4.2}$$

An estimate, based on asymptotic theory for the standard error of $\hat{S}(\tau)$ at a fixed value τ is

$$st.er\left(\hat{S}(\tau)\right) = \hat{S}(\tau)\left\{\sum_{j\in\{j:\ \tau_j<t\}} \frac{m_{(i)}}{r_j\left(r_j - m_{(i)}\right)}\right\}^{1/2} \tag{4.3}$$

An apparent estimate of the cumulative hazard H(t) is given by

$$\hat{H}(t) = -\log\hat{S}(t) = -\log\left[\prod_{j\in\{j:\ \tau_j<t\}}\left(1 - \frac{m_{(j)}}{r_j}\right)\right] \tag{4.4}$$

A slightly simpler estimate of H(t) is

$$\bar{H}(t) = \sum_{j\in\{j:\ \tau_j<t\}} \frac{m_{(j)}}{r_j} \tag{4.5}$$

The last estimate of the cumulative hazard H, can be found as the Nelson's estimate.

The standard error of $\bar{H}(.)$ is the same as that of $\hat{H}(.)$ at a specific time τ and is given by the Greenwood formula (Klein (1991))

$$st.er\left(\hat{H}(\tau)\right) = st.er\left(\bar{H}(\tau)\right) = \left\{\sum_{j\in\{j:\ \tau_j<t\}} \frac{m_{(i)}}{r_j\left(r_j - m_{(i)}\right)}\right\}^{1/2} \tag{4.6}$$

The product limit theory can also be found in Crowder et al (1995) and in Lee (1992).

The Nelson's estimate is a step function. It starts at zero and has a step of size $\frac{m_{(j)}}{r_j}$ at each failure. One disadvantage with this estimation is that it is susceptible to ties in the data. For that reason, a modified Nelson estimate is suggested by Nelson and Fleming and Harrington (1984), denoted by $\tilde{H}(.)$. The relationship $H(t) = -\log\bar{F}(t)$, which holds for any continuous distribution, leads to the Fleming-Harrington

(FH) (Fleming and Harrington (1984)) estimate of the survival, which is

$$\hat{S}_{FH}(t_j) = \exp\left(-\tilde{H}(t_j)\right) \tag{4.7}$$

A step function is also derived after plotting $\hat{S}(t)$ versus t, for all t values. In the special case of Weibull distributed data, a plot of the points $(\log \tau_j, \log\{-\log(1-p_j)\})$ for j=1,2,...,k should be approximately linear if the Weibull model is suitable. Where p_j is equal to the quantity, $1 - \frac{1}{2}\left(\hat{S}(a_j) + \hat{S}(a_{j+1})\right)$, calculated in the time points a_i from i=1...k. Similarly a plot of the points $(\log \tau_j, \Phi^{-1}(p_j))$ should be approximately linear if lognormal model is appropriate.

4.2.1 Example Study

Maintained		Non-Maintained	
Time	*Status*	*Time*	*Status*
9	1	5	1
13	1	5	1
13+	0	8	1
18	1	8	1
23	1	12	1
28+	0	16+	0
31	1	23	1
34	1	27	1
45+	0	30	1
48	1	33	1
161+	0	43	1
–	–	45	1

Table 4.1: Data set for AML maintenance study. The + indicates a censored value.

The Data in Table (4.1) are those from Embury et al. (1977) on trial to evaluate efficacy of maintenance chemotherapy for acute myelogenous leukemia. Acute myelogenous leukemia is a fatal type of leukemia that arises within some weeks, and the remission period (period that the process of leukemia reduces or even stops) is smaller than that of the other types. The patients are assigned into two groups. In the first group, patients received **maintenance** chemotherapy are included, whereas, in the second group they did not. The objective of the trial was to see if maintenance chemotherapy prolonged the time until relapse. Indeed, the columns time on Table (4.1) refer to the time of remission, and the status columns indicates whether the observation is censored (status equals to 0) or not. The symbol of addition + is used also to indicate the censored survivor times.

group=Maintained						
time	**n. risk**	**n.event**	**survival**	**std.err**	**lower 95% CI**	**upper 95% CI**
9	11	1	0.909	0.0867	0.7541	1.000
13	10	1	0.818	0.1163	0.6192	1.000
18	8	1	0.716	0.1397	0.4884	1.000
23	7	1	0.614	0.1526	0.3769	0.999
31	5	1	0.491	0.1642	0.2549	0.946
34	4	1	0.368	0.1627	0.1549	0.875
48	2	1	0.184	0.1535	0.0359	0.944

Table 4.2 : Kaplan-Meier estimated Survival times for Maintained group

group=Non-Maintained						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
5	12	2	0.833	0.108	0.647	1.000
8	10	2	0.667	0.136	0.447	0.995
12	8	1	0.583	0.142	0.362	1.941
23	6	1	0.486	0.148	0.268	0.883
27	5	1	0.389	0.147	0.185	0.816
30	4	1	0.292	0.139	0.115	0.714
33	3	1	0.194	0.122	0.057	0.664
43	2	1	0.097	0.092	0.015	0.620
45	1	1	0.000	NA	NA	NA

Table 4.3 : Kaplan-Meier estimated Survival times for Non-Maintained group

In the above tables ((4.2), (4.3)), the estimated survival times are given along with the standard error at each specific observation, using the equations (4.1) and (4.3). Confidence intervals, at a specific time point τ are approximated by the formula

$$\hat{S}(\tau) \pm z_{a/2} \times st.er\left(\hat{S}(\tau)\right) \tag{4.8}$$

Equation (4.8) may give survival bounds that are greater than the unity or less than zero. On the other hand confidence intervals based on the cumulative-hazard scale given by

$$\exp\left(\log S \pm z_{a/2} \times se\left(\hat{H}\right)\right) \tag{4.9}$$

have the best performance, even though they may sometimes be greater than unity. In equation (4.9) the quantity $z_{a/2}$ is the upper a/2 quantile of the Standard normal distribution. In our case α equals to 0.05 and 95% confidence intervals are calculated.

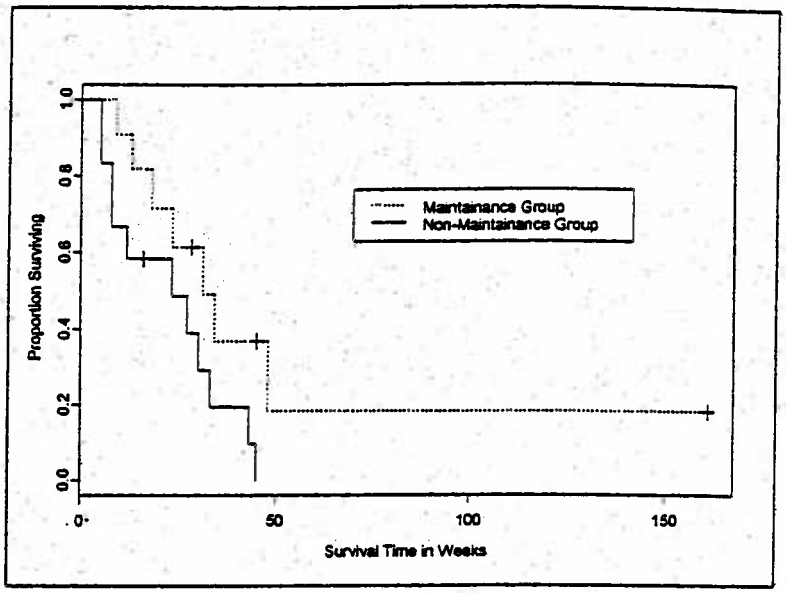

Figure 4-1: Kaplan Meier Survival curves

- Another release of the equation of the confidence intervals is on the log-hazard or log-log survival scale,

$$\exp\left(-\exp\left(\log\left(-\log S\right) \pm z_{a/2} \times se\left(\log \hat{H}\right)\right)\right) \tag{4.10}$$

- A further refinement to the confidence interval is suggested by Dorey and Korn in 1987. When the tail of the survival curve contains much censoring and few failures, there will be one or more long flat segments. However, intervals based on the above equations ((4.8), (4.9) & (4.10)) are constant across these censors. Dorey and Korn point out a correction to the lower limit which is now based on the *effective number at risk* between death times.

Similar results are derived, using the Fleming-Harrington estimators for the survival times. Particularly, the estimated values presented in tables (4.4), (4.5) are slightly different. Indeed for sufficiency large sample sizes the Fleming-Harrington and Kaplan-

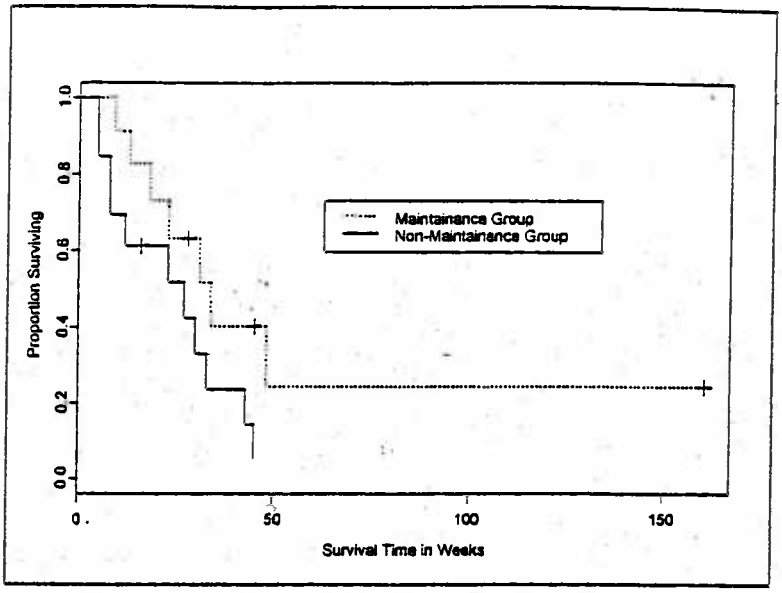

Figure 4-2: Fleming Harrington Survival curves

Meier estimators are arbitrarily close to one another. In addition the survival curves are plotted in Figure 4-2.

group=Maintained						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
9	11	1	0.913	0.0871	0.7575	1.000
13	10	1	0.826	0.1174	0.6253	1.000
18	8	1	0.729	0.1422	0.4974	1.000
23	7	1	0.632	0.1572	0.3882	1.000
31	5	1	0.517	0.1731	0.2687	0.997
34	4	1	0.403	0.1781	0.1695	0.958
48	2	1	0.244	0.2038	0.0477	1.000

Table 4.4 : Fleming-Harrington estimated Survival times for Maintained group

group=Non-Maintained						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
5	12	2	0.847	0.109	0.657	1.000
8	10	2	0.693	0.141	0.465	1.000
12	8	1	0.612	0.149	0.379	0.987
23	6	1	0.518	0.158	0.285	0.941
27	5	1	0.424	0.160	0.202	0.889
30	4	1	0.330	0.157	0.130	0.838
33	3	1	0.237	0.148	0.069	0.808
43	2	1	0.144	0.136	0.023	0.914
45	1	1	0.053	INF	0.000	1.000

Table 4.5 : Fleming-Harrington estimated Survival times for Maintained group

In order to compare the two different groups with respect to their survival distributions we can use the Mantel-Haenszel or else known as **log-rank** test, the Gehan-Wilcoxon test which is used modified as the **Peto** test.

Log rank test

The log-rank test is a chi-square which uses as its test criterion a statistic that provides an overall comparison of the Kaplan-Meier curves being compared. Like in other kind of chi-squared tests, *observed* and *expected* cell counts over categories of outcomes are required. The categories in our case are defined by each of the ordered failure times for the entire set of the data being analyzed.

Imagine p separate groups of patients. Let us denote by m_{ij} the number of failures in the **ith** group and in the jth failure time $t_{(j)}$. Also we can extend the notation of the subjects in risk set as r_{ij} which represents those in risk in the ith group and in the jth

failure time. Then the expected number of failures in the kth group is

$$E_k = m_{.j}\frac{r_{kj}}{r_{.j}}, \quad k = 1...p \ \& \ j = 1...k_k \tag{4.11}$$

Where the k_k is the number of observations in the kth group. The rank statistic is given by

$$\text{log-rank statistic} = \sum_i^p \frac{(O_i - E_i)^2}{Var\,(O_i - E_i)} \tag{4.12}$$

Where O_i is the sum of the observed failures in the ith group. Under the null hypothesis (that the KM survival curves are statistically equivalent) the log-rank statistic is approximately chi-square with p-1 degrees of freedom. An approximation formula of (4.12) derives by the substitution of the $Var\,(O_i - E_i)$ with the expected values E_i.

Thus with respect to the leukemia example, Table (4.6) includes the log-rank statistic for both groups. The p-value 0.065 suggests that there is mild evidence that the maintained group has better survival than the non-maintained group. Another test statistic is the Peto modification of the Wilcoxon test.

Wilcoxon and Peto statistics

The above procedure may be generalized by the inclusion of weights w_i for each failure time. The overall weighted vector is then $\sum w_i\,(O_i - E_i)$, where $O_i - E_i$ are the observed minus the expected values added in the ith group. Respectively the variance of this quantity is $\sum w_i^2 var\,(O_i - E_i)$ and the new statistic becomes now,

$$\sum_i^p \frac{w_i\,(O_i - E_i)^2}{w_i^2 Var\,(O_i - E_i)} \tag{4.13}$$

When w_k equals one, quantity (4.13) reduces to the Mantel-Haenszel or long-rank test, for w_k equals r_k, quantity (4.13) is the Gehan-Wilcoxon test, and for w_k equals $\hat{S}_{KM}$, quantity (4.13) is the Peto modification of the Wilcoxon test.

The distribution of the statistic in form (4.13) is approximately chi-square with p-1 degrees of freedom. Table (4.6) includes the Peto statistic for both groups giving a p-value 0.096. Thus, the previous result is verified in the absence of evidence that the two curves differ.

- The different formulas described above, indicate that the Peto test places more emphasis on the information at the beginning of the survival curve where the number at risk is large. In other words, early failures receive larger weights while failures in the tail of the survival curve receive smaller weights.

- On the other hand, the log-rank test emphasizes failures in the tail of the survival curve, where the number at risk decreases over time, yet equal weight is given to each failure time.

Group	N	Observed	Expected	long-rank	Peto
Maintained	11	7	10.689	1.273	0.859
Non-Maintained	12	11	7.311	1.862	1.081
Chi-Square test				3.396	2.78
P-value				0.065	0.096

Table 4.6 : Survival testing between the groups

4.3 The Accelerated Life Models

Assuming that the subjects are divided into two groups, we apply a different treatment in each group. We represent those two treatments by values 0 and 1 of the explanatory variable Z. Then the survivor function of the one group e.g. at $z = 1$ is a function of the survivor of the other group e.g. at $z = 0$. Particularly, there is a constant, namely ψ such that

$$\bar{F}_1(t) = \bar{F}_0(\psi t) \tag{4.14}$$

Thus, the density function and the hazard rate are given respectively as

$$f_1(t) = \psi f_0(\psi t) \tag{4.15}$$

$$h_1(t) = \psi h_0(\psi t) \tag{4.16}$$

More formally, suppose that there is a positive function $\psi(\mathbf{z})$, where $\mathbf{z}$ are the explanatory variables. The accelerated life model holds when the survivor function $\bar{F}(t;\mathbf{z})$ is of the form

$$\bar{F}(t;\mathbf{z}) = \bar{F}_0(t\psi(\mathbf{z})) \tag{4.17}$$

The density and hazard functions are

$$f(t;\mathbf{z}) = \psi(\mathbf{z}) f_0(t\psi(\mathbf{z})) \tag{4.18}$$

$$h(t;\mathbf{z}) = \psi(\mathbf{z}) h_0(t\psi(\mathbf{z})) \tag{4.19}$$

The last equation (4.19) is derived by (2.6) and (4.17) as follows

$$h(t;\mathbf{z}) = -\frac{\partial}{\partial t}\log \bar{F}(t;\mathbf{z}) = -\frac{\partial}{\partial t}\log \bar{F}_0(t\psi(\mathbf{z})) = \psi(\mathbf{z})\, h_0(t\psi(\mathbf{z}))$$

The function $\bar{F}_0(.)$ refers to the situation when $\mathbf{z}$ equals to zero, that is no values from the explanatory variables are used to evaluate the above functions. So, from (4.17) we can conclude that under the standard conditions $\mathbf{z} = \mathbf{0}$, $\psi(\mathbf{z} = 0) = 1$. Applying, the above equations (4.17), (4.18) and (4.19) in terms of random variables, we can infer that

$$T = T_0/\psi(\mathbf{z}) \tag{4.20}$$

Where T_0 has survivor function $\bar{F}_0(.)$. If $\mu_0 = E(\log T_0)$, we can rewrite (4.20) as

$$\log T = \log(T_0) - \log\psi(\mathbf{z}) \tag{4.21}$$

Assuming $log(T_0)$ as a random variable, we can write $log(T_0)$ as the sum of $\mu_0 + \varepsilon$, where ε is a new random variable with mean zero. After the last considerations, the model in (4.21) becomes

$$\log T = \mu_0 - \log\psi(\mathbf{z}) + \varepsilon \tag{4.22}$$

Note that ε is coming from the random variable $\log(T_0)$, which do not involve the vector $\mathbf{z}$. Thus, the distribution of ε is independent of the explanatory variables.

In problems in which the values of $\mathbf{z}$ are finite and distinct, it may be unnecessary to specify $\psi(.)$ further. In other cases, a parametric form $\psi(\mathbf{z};\boldsymbol{\beta})$ is essential For instance, as $\psi(\mathbf{z};\boldsymbol{\beta})$ should be non negative and $\psi(\mathbf{0};\boldsymbol{\beta}) = 1$, one possible option is to choose the exponential function. In other words we can write $\psi(\mathbf{z};\boldsymbol{\beta}) = e^{\boldsymbol{\beta}^T\mathbf{z}}$, and transform the equation (4.22) into the linear regression model

$$\log T = \mu_0 - \boldsymbol{\beta}^T\mathbf{z} + \varepsilon \tag{4.23}$$

4.3.1 Time dependent explanatory variables

There are cases where the treatment variables are non constant, with respect to the passage of time. In this spot, the hazard at any particular time t depends only on the explanatory variable at that time, as the last varies according to time. This usually involves the use as components of $z(t)$ of integrals, sums, derivatives and differences of the explanatory variables as originally recorded.

The essence of the accelerated life model is that the failure time is contracted or expanded relative to that at the situation where $\mathbf{z} = 0$. This suggests that for an individual characterized by $z(t)$, time $t^{(z)}$, say, evolves relative to the time $t^{(0)}$ for that individual as,

$$dt^{(z)} = dt^{(0)} / \psi \left[z \left(t^{(z)} \right) \right] \tag{4.24}$$

The last formula (4.24), actually suggests that the any change in the time of a system with $\mathbf{z}$ the explanatory vector, equals to the same change in the time of a system with no explanatory information, divided by the function ψ of the accelerated model. So

$$t^{(0)} = \int_0^{t(z)} \psi[z(u)] du = \Psi \left(t^{(z)} \right)$$

and the new relation of the failures is now

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

As a consequence, the survivor function, density and hazard are

$$\left.\begin{aligned} \bar{F}\left(t; \{\mathbf{z}(.)\}\right) &= \bar{F}_0 \Psi(t) \\ f\left(t; \{\mathbf{z}(.)\}\right) &= \psi\left(\mathbf{z}(t)\right) f_0\left(\Psi(t)\right) \\ h\left(t; \{\mathbf{z}(.)\}\right) &= \psi\left(\mathbf{z}(t)\right) h_0\left(\Psi(t)\right) \end{aligned}\right\} \tag{4.25}$$

For instance, we can consider the comparison of two groups. Then, in the place of

the binary values of z, we place the function

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

Where $\xi(t)$ is a function of time. If we take $\psi(\mathbf{z})$ to be in the exponential form $e^{\mathbf{z}}$, the survivor function for the second group will be from (4.25)

$$\bar{F}(t;\{\mathbf{z}(.)\}) = \bar{F}_0\left(\int_0^t e^{\xi(u)}du\right)$$

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

$$e^{\xi(t)} = \frac{d}{dt}\bar{F}_0^{-1}\bar{F}_1(t)$$

A fairly rich family of models are produced by choosing $\xi_j(t) = t^j$ for j lies between 0 and a suitable integer p $j = 0 \ldots p$. As follows in Cox and Oakes (1984), a suitable function of the explanatory vector $\mathbf{z}$ is the exponential, i.e. $\psi(\mathbf{z}) = e^{\beta^T \mathbf{z}}$, where β is a $q \times 1$ parameter vector, with q the number of the explanatory variables.

The inconsistency of the Accelerated life model is explained by the presence of several types of failure $\bar{F}$. Each failure follows an Accelerated model with a different modifying function ψ. As $\mathbf{z}$ varies, the balance between the types of failure changes. In other words the Accelerated models that represent different failure, are not compatible.

Relation with proportional hazards model: Cox & Oakes (1984) showed that when the initial distribution is the Weibull distribution (3.10) , with constant explanatory variables, the accelerated life and proportional hazards models coincide.

$$\bar{F}_0(t) = \exp\left[-(\rho t)^{\kappa}\right] \tag{4.26}$$

In this spot, using Weibull failure times in first place, we obtain survival curves

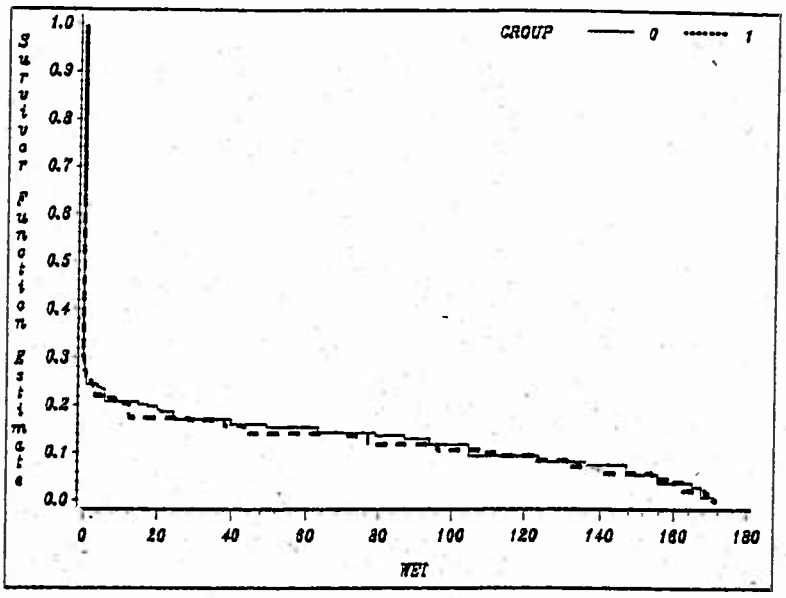

Figure 4-3: Survivor estimated curves

similar to those estimated using the proportional hazards model. In the example shown in Figure above (Fig 4-3) the failure times are generated from a Weibull distribution with parameter λ equal to 2, and the values are divided into two groups. There is no statistically significant difference in those curves, fact which is supported by both the p-value (0.8748) from the chi-square test, which exceeds the 0.05 limit and the hazard ratio which is almost unity (1.023).

4.4 Proportional Odds Model

The Proportional Odds Model is expressed by the formula

$$\left(1-\bar{F}\left(t;\mathbf{z}\right)\right)/\bar{F}\left(t;\mathbf{z}\right)=\psi_{\mathbf{z}}\left(1-\bar{F}_0\left(t\right)\right)/\bar{F}_0\left(t\right) \tag{4.1}$$

Bernett (1983). The defining ratio $\left(1-\bar{F}\left(t;\mathbf{z}\right)\right)/\bar{F}\left(t;\mathbf{z}\right)$ is actually the ratio $\frac{pr(T<t)}{pr(T\geq t)}$, which is the odds on the event $T<t$. So last equation illustrate that the odds under explanatory variables $\mathbf{z}$ is the product of the baseline odds $\left(1-\bar{F}_0\left(t\right)\right)/\bar{F}_0\left(t\right)$ times the

function of z ψ_z. The greater the ψ_z, the greater the probability of a shorter lifetime. Similarly to the Proportional Hazards model (we refer in next chapter), the current model continues to hold under transformation of the time scale. Differentiating last equation with respect to the time t yields $\frac{h(t;z)}{\bar{F}(t;z)} = \frac{h_0(t)\psi_z}{\bar{F}_0(t)}$. Thus from (4.31) the hazard ratio satisfies

$$\frac{h(t;\mathbf{z})}{h_0(t)} = \frac{\psi_z \bar{F}(t;\mathbf{z})}{\bar{F}_0(t)} = \frac{\left(1-\bar{F}(t;\mathbf{z})\right)}{\left(1-\bar{F}_0(t)\right)} \tag{4.27}$$

It follows from (4.27) that the hazard ratio equals to ψ_z at t=0, and tends to 1 as t$\rightarrow \infty$. The last indicates that the effect of the explanatory variables z on the hazard diminishes as time goes on. That is, either the system adjusts to the factors imposed on it, or the factors operate only in the earlier stages, referring to time.

4.4.1 Two-sample case

A class of consistent and asymptotically normal estimates in the two-sample case is given by Dabrowska & Doksum (1988). In particular, the definition of the **generalized** odds ratio for T is given by

$$H_T(t|c) = \begin{cases} \frac{1}{c}\left[\frac{1-\bar{F}^c(t)}{\bar{F}^c(t)}\right], & c>0 \\ -\log \bar{F}(t), & c=0 \end{cases} \tag{4.28}$$

According to (4.28), $H_T(t|0)$ is the integrated hazard, whereas $H_T(t|1)$ is the odds of the response T occurring before time t. For a value of c other than unity, $H_T(t|c)$ also has an interpretation as an odds ratio.

We consider next, two independent samples $\{X_i\ i = 1 \ldots m$ and $Y_j\ j = 1 \ldots n\}$ of lifetimes with distribution functions F and G, respectively. To tackle the problem of modeling herein, we can assume (as in the literature) that the log of the generalized odds rate given in (4.28) is linear in the explanatory variable, which identifies the sample. In other words, the relation of the integrated hazards of the two samples is

$$H_Y(t) = \theta^{-1} H_X(t), \qquad \text{for all t>0, and } \theta > 0 \tag{4.29}$$

An estimator used for θ is

$$\hat{\theta} = \left(\int_0^1 \Psi(u) (1-u)^{-(c+1)} du \right)^{-1} \int_0^\infty \Psi(G_n(t)) [1 - F_m(t)]^{-(c+1)} dF_m(t) \tag{4.30}$$

where F_m and G_n are the left-continuous empirical distribution functions based on the X and Y samples, respectively. Also, Ψ is some score function. Considering the leukemia data, given in Table 4.1, where there is no absolute evidence that the survivors in the patients of the maintained group are longer than those of the non-maintained group. Thus, we can use the fully efficient estimate of Ψ,

$$\Psi(t) = 2(1-t)^3$$

and substituting in the formula (4.30) we have as in (Dabrowska & Doksum (1988)) that,

$$\hat{\theta}_0 = \frac{2m}{n^3} \sum_{i=1}^{m} \frac{N_{(i)}^3}{(m+1-i)^2} \tag{4.31}$$

where $N_{(i)} = \sum_{j=1}^{n} I\left[Y_j \geq X_{(i)}\right]$ is the number of $Y_j's$ at risk at time $X_{(i)}$. As before $X_{(i)}$ is the ith order statistic among the X's and I is the indicator function.

Referring to the leukemia data, formula (4.31) gives $\hat{\theta}_0 = 0.0137828$.

Next, in order to test the null hypothesis $(H_0 : \theta = 1)$ versus the alternative one $(H_1 : \theta > 1)$, we use the statistic

$$(mn/(m+n))^{1/2} \left(\hat{\theta}_0 - 1\right) 3^{-1/2}$$

which follows the standardized normal distribution under the H_0. Here, the test statistic

equals to -0.569393 which is greater than the $1-a$th quantile of the standard normal distribution ($z_{1-a}=1.64$). So the null hypothesis cannot be rejected, and there is no clear evidence for the reliability of the effect of the maintenance chemotherapy. This result agrees with the results emerge form both the long rank and the Peto tests.

4.5 Additive risk Model

The additive risk models has been studied, in various forms by numerous authors. Aalen (1980, 1989) introduced a regression model, with response variable the conditional hazard function $h_i(t)=h_i(t|Z_i)$ for subject i, where Z_i are its covariates, given by the p vector $Z_i=(Z_{i1},\ldots,Z_{ip})^T$. Aalen's model stipulates that

$$h_i(t|Z_i)=Z_i^T a(t) \tag{4.32}$$

where $a=(a_1,\ldots,a_p)^T$ is an unknown vector of hazard functions.

An OLS estimator is then given, by applying the regression analysis techniques. Later on, Huffer & Mckeague (1991), based the inference of the vector a on a weighted least squares (WLS) estimator. Also, confidence intervals and bands were calculated, in both grouped and continuous data. As grouped data, they perceived, the time points at risk and the number of uncensored deaths taken over successive time intervals, for various levels of covariates.

Another approach is made by Buckley (1984). The model suggested is

$$h_i(t)=\psi+h^*(t;z_i) \tag{4.33}$$

where ψ is the disease effect, which reflects to the mortality of the patient. Also, $h^*(t;z_i)$ is the hazard function for the ith individual ($i=1\ldots n$), for causes of failure (here the death) other than that under study, and finally, z_i are the covariate of the particular patient. In the current analysis, the effect of the disease is assumed either

constant throughout the entire follow-up period, or piecewise constant within k follow-up intervals. Maximum likelihood estimates along with related statistics, based on moment estimators of the disease effect, are presented. Finally, the testing of homogeneity of r arbitrary groups is enlightened.

Similar, models have been eloquently advocated and successfully utilized by other authors, like Breslow & Day (1980, 1987), Pocock et al (1982), Pierce & Preston (1984), Thomas (1986). In Lin & Ying (1994, 1995), an augmentation form of the previous models is used. Particularly, the hazard function $h(t; Z)$ for the failure time T is associated with a p-vector of possibly time-varying covariates $\mathbf{Z}(.)$, as

$$h(t; Z) = h_0(t) + \beta_0^T \mathbf{Z}(t) \tag{4.34}$$

where β_0 is the p-vector of regression parameters, and $h_0(t)$ is the **baseline hazard** function, which we review in the next chapter. The difference on the approach of the current model (4.34), is that the inference for the baseline hazard function and the unknown parameters β_0 is based on the martingale feature of the partial likelihood score function (see Gill (1980), Anderson & Gill (1982) and Cox (1975)). So, the problem of elimination or estimation of the nuisance function $h_0(.)$ with the direct application of the partial likelihood, is surpassed.

4.5.1 Construction of the Estimators

Firstly, let $\{N_i(t);\ t \geq 0\}$ be the counting process for the ith subject in the set, which records the number of observed events up to time t. Then, the **intensity function** or else the mean number of events occurring in time unit, for $N_i(t)$ is given by

$$Y_i(t)\, dH(t; Z_i) = Y_i(t) \left(dH_0(t) + \beta_0^T Z_i(t)\, dt\right) \tag{4.35}$$

where $Y_i(.)$ is a 0-1 predictable process, and $Y_i(t) = 1$ if the ith subject is at risk at time t, whereas equals to zero otherwise. Also $H(t; Z_i)$ is the **cumulative** or **integrated**

hazard function In this spot, the baseline cumulative hazard function is given similarly by

$$H_0(t) = \int_0^t h_0(u)\, du \tag{4.36}$$

The counting process $N_i(.)$ is uniquely decomposed for every i and t, in the form

$$N_i(t) = M_i(t) + \int_0^t Y_i(u)\, dH(u; Z_i) \tag{4.37}$$

where $M_i(.)$ is a local square integrable martingale process (Anderson & Gill, (1982)). From the last equation (4.37), an estimate of $H_0(t)$, is

$$\hat{H}_0\left(\hat{\beta}, t\right) = \int_0^t \frac{\sum_{i=1}^n \left\{dN_i(u) - Y_i(u)\,\hat{\beta}^T Z_i(u)\, du\right\}}{\sum_{j=1}^n Y_j(u)} \tag{4.38}$$

The partial likelihood score function for β_0 as given in Lin & Ying (1994) is,

$$U(\beta) = \sum_{i=1}^n \int \left\{Z_i(t) - \bar{Z}(t)\right\} \left\{dN_i(t) - Y_i(t)\,\beta^T Z_i(t)\, dt\right\} \tag{4.39}$$

where

$$\bar{Z}(t) = \frac{\sum_{j=1}^n Y_j(t)\, Z_j(t)}{\sum_{i=1}^n Y_i(t)}$$

The solution of the equation $U(\beta) = 0$ gives an estimator for β, which can be written in the explicit form

$$\hat{\beta} = \left(\sum_{i=1}^n \int_0^\infty Y_i(t) \left\{Z_i(t) - \bar{Z}_i(t)\right\}^{\otimes 2} dt\right)^{-1} \left(\sum_{i=1}^n \int_0^\infty \left\{Z_i(t) - \bar{Z}_i(t)\right\}^{\otimes 2} dN_i(t)\right) \tag{4.40}$$

where $u^{\otimes 2} = uu^T$.

The random vector $n^{\frac{1}{2}}\left(\hat{\beta} - \beta_0\right)$ converges weakly to a p-variate normal distribution with zero mean and covariance matrix, which can be consistently estimated by the

quantity $A^{-1}BA^{-1}$, where

$$A = n^{-1}\sum_{i=1}^{n}\int_0^{\infty} Y_i(t)\left\{Z_i(t) - \bar{Z}(t)\right\}dt$$

Similarly, $n^{\frac{1}{2}}\left\{\hat{H}_0\left(\hat{\beta},.\right) - H_0(.)\right\}$ converges weakly to a zero-mean p-variate normal distribution, and covariance $\mathrm{Cov}(t,s)$ $(t \geq s)$, approximated by

$$Cov(t,s) = \int_0^s \frac{n\sum_{i=1}^{n} dN_i(u)}{\left(\sum_{j=1}^{n} Y_j(u)\right)^2} + C'(t)A^{-1}BA^{-1}C(s) - C'(t)A^{-1}D(s) - C'(s)A^{-1}D(t) \tag{4.41}$$

where $\mathrm{C}(t) = \int_0^t \bar{Z}(u)\,du$, and $D(t) = \int_0^t \frac{\sum_{i=1}^{n}\left\{Z_i(u)-\bar{Z}(u)\right\}dN_i(u)}{\sum_{j=1}^{n} Y_j(u)}$. Next, if we denote the survival function for an individual with a given covariate vector $z(.)$, as $\bar{\mathrm{F}}(t,z,)$, then a plausible estimate of $\bar{\mathrm{F}}(.,.)$ is

$$\hat{F}(t;z) = \exp\left\{-\hat{H}_0\left(\hat{\beta},t\right) - \int_0^t \hat{\beta}^T z(u)\,du\right\} \tag{4.42}$$

The process $\mathrm{n}^{\frac{1}{2}}\left\{\hat{F}(.;z) - \bar{F}(.,z)\right\}$ converges weakly to a zero-mean p-variate normal distribution, with covariate function at point (t,s) $(t \geq s)$ equal to

$$\begin{aligned}\hat{S}(t;z)\,\hat{S}(s;z)\Bigg[&\int_0^s \frac{n\sum_{i=1}^{n} dN_i(u)}{\left\{\sum_{j=1}^{n} Y_j(u)\right\}^2} + G'(t;z)A^{-1}BA^{-1}G(s;z)\\ &+G'(t;z)A^{-1}D(s) + G'(s;z)A^{-1}D(t)\Bigg]\end{aligned} \tag{4.43}$$

where $G(t;z) = \int_0^t \left\{z(u) - \bar{Z}(u)\right\}du$.

4.5.2 Two-sample example

Applying an additive model for the leukemia data, mentioned previously (Section 4.2), we can utilize the above estimates. In our case, the covariate Z is a binary one, defined as 1 for the non-maintenance group and as 0 for the maintenance group. So the model

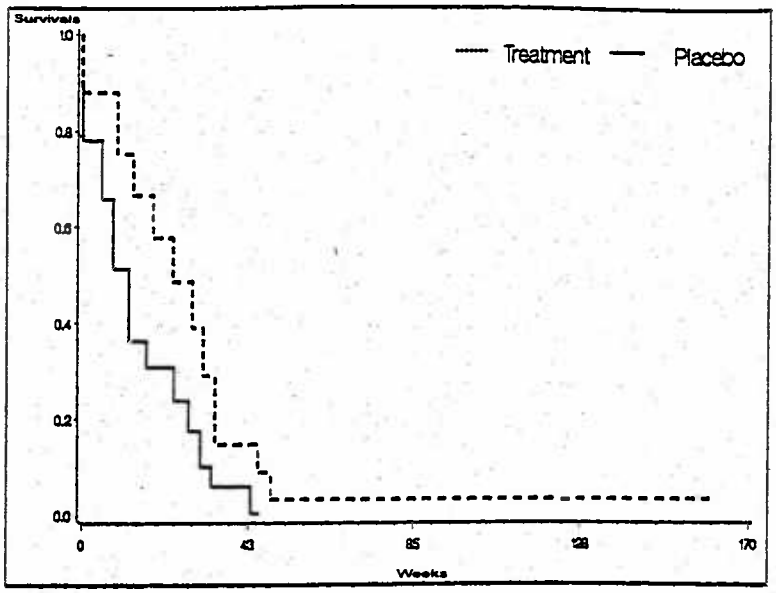

Figure 4-4: Kim & Ying Survivor estimates for leukemia two sample data.

(4.34) reduces to

$$h(t;Z) = h_0(t) + \beta_0 Z \tag{4.44}$$

The estimates of the baseline hazard $h_0(t)$, the unknown parameter β_0 and the survivals for each group are given respectively in the Tables (4.7) & (4.8).

$\hat{h}_0(t;\hat{\beta})$	$\hat{S}(t;0)$	$\hat{S}(t;1)$
0	1	1
0.031	0.864	0.969
0.066	0.779	0.936
0.053	0.771	0.936
0.207	0.617	0.813
0.133	0.617	0.813
0.320	0.502	0.726
0.231	0.502	0.726
0.525	0.349	0.592
0.230	0.349	0.592
0.615	0.284	0.541
0.478	0.284	0.541
0.833	0.213	0.435
0.521	0.213	0.435
0.998	0.169	0.369
0.761	0.169	0.369
1.476	0.081	0.228
1.396	0.081	0.228
4.281	0.081	0.014

Table 4.7: Lin & Ying estimates for the leukemia data.

$\hat{\beta}$	0.020279
$st.er$	0.0125112
$t-statistic$	1.838832
$p-value$	0.078862

Table 4.8: Parameter β_0 estimate and t-statistic.

The survivor estimates are also plotted in Fig 4-4. The dashed stepped line referes to the maintenance group, whereas the solid one, describes the survival curve of the non-maintenance, known else in the literature as **placebo** or **control** group. The failure time is measured in weeks, since the patient get out of the remission time. The range between the two lines, is rather small, like in Figures Fig(4-1) and Fig (4-25) in the previous analysis and there is no intense evidence that the chemotherapy treatment, had success. This initial assumption can also be drawn from the small value of the $\hat{\beta}$ estimator $\left(\hat{\beta} = 0.020279\right)$ given in Table (4.8). Testing the null hypothesis $(H_0 : \beta = 0)$, we conclude that the probability value of 0.078862 is not minor enough to reject the hypothesis in 95% significant level. Thus the presumption of high effects in the use of the chemotherapy in the leukemia patients cannot be supported, once again.

4.5.3 Goodness of fit

Kim & Lee (1998) suggested two goodness of fit tests for the two-sample additive risk models (4.44) with censored observations. The first optimize test is based on the martingale residuals and has similarities with that proposed by Wei (1984) . The second is based on the difference between weighted estimators of the excess risk, idea originated in Gill & Schumacher (1987) and Lin (1991).Both the test statistics are asymptotically normal under appropriate regularity conditions and consistent under any model misspecifications, given by Kim & Lee (1998) .

- In few words, the optimize test statistic S which tests for constant difference between two hazard functions, is given by

$$S = \sup_{0 \leq t < \infty} \hat{\xi}(\infty)^{-\frac{1}{2}} \left| n^{-\frac{1}{2}} U_0\left(\hat{\beta}, t\right) \right| \tag{4.45}$$

where $\hat{\xi}(\infty)$ is the limit as $t \to \infty$ of a score function that contains the cumulative hazard function while, $U_0\left(\hat{\beta}, t\right)$ is the quantity (4.39) modified to contain the local square integrable martingale process $M_i(.)$. This quantities are obtained by simulation techniques given by Kim & Lee (1998). The availability of the additive risk model is checked graphically and numerically by the probability of the simulated statistic value $\hat{S}$ be greater or equal to the observed value of $S_0\left(\hat{S} \geq s_0\right)$.

- Another test statistic for the additive risk assumption is given by

$$Q_w = n\left(\hat{\beta}_w - \hat{\beta}\right)^T \hat{D}_w\left(\hat{\beta}\right)^{-1}\left(\hat{\beta}_w - \hat{\beta}\right) \tag{4.46}$$

where $\hat{D}_w\left(\hat{\beta}\right)$ is a consistent estimator for a covariance matrix and $\hat{\beta}_w$ is a weighted estimator given in Kim & Lee (1998). The statistic Q_w has an asymptotic central *chi − square* distribution with 1 degree of freedom under model (4.34).

Chapter 5

Proportional Hazards Model

5.1 Introduction

In the current chapter, regression models are considered where the response variable is the hazard or else age-specific failure rate. In this spot the hazard rate is a function of the explanatory variables and unknown regression coefficients multiplied by an arbitrary and unknown function of time called the baseline function. We then, present the proportional hazard model, introduced by Cox (1972). The tool of the likelihood is used every time to obtain inferences about the unknown regression coefficients. Particularly, condition probabilities of the failure times, given the subjects still in risk, are used for the unknown parameters' estimation. Finally, we describe its basic properties, including considerations for its popularity.

5.2 General Description

Assuming a set of explanatory variables denoted by the bold $\mathbf{Z}$. In other words, the set $\mathbf{Z}$ represents a collection of predictor variables that is being modeled to predict individual's hazard. That is we can correspond the proportional hazard model as a regression model with the hazard rate in the role of the response variable which can be expressed by the

product of a function of the explanatory vector Z times a function considered under the standard conditions, $\mathbf{Z} = 0$. In particular, the simple form of the proportional hazards model introduced by Cox and Oakes (1984) is:

$$h(t, \mathbf{z}) = \psi(\mathbf{z}; \beta) h_0(t) \tag{5.1}$$

The last formula emerges that the hazard at time point t is the function of two quantities in (5.1). The second of these, is called the **baseline hazard** function. It is the hazard rate at time t for an individual taking into account the information of no explanatory variables. The first component in the product in the left of the (5.1) equation, is the expression of the explanatory variables contained in the vector Z, along with the unknown parameters β to be estimated. Three parameterizations are considered for the last expression Namely, the first and most common is the log linear form $\psi(\mathbf{z}; \beta) = e^{\beta^T \mathbf{z}}$,

$$h(t, \mathbf{z}) = e^{\beta^T \mathbf{z}} h_0(t) \tag{5.2}$$

The second is the linear form $\psi(\mathbf{z}; \beta) = 1 + \beta^T \mathbf{z}$

$$h(t, \mathbf{z}) = \left(1 + \beta^T \mathbf{z}\right) h_0(t) \tag{5.3}$$

and the logistic form $\psi(\mathbf{z}; \beta) = \log\left(1 + e^{\beta^T \mathbf{z}}\right)$

$$h(t, \mathbf{z}) = \log\left(1 + e^{\beta^T \mathbf{z}}\right) h_0(t) \tag{5.4}$$

The first two parameterizations can be found in a more augmented form like the following:

$$\psi(\mathbf{z}; \beta) = \left(1 + k\beta^T \mathbf{z}\right)^{1/k}$$

Especially, the linear form is deducted from the last formula by putting k equal to unity, while the log linear expression is obtained by assuming $k \rightarrow 0$.

An important feature of the above formula (5.1) is that considering the explanatory vector Z invariable, as far as the time is concerned, the functional expression $\psi(\mathbf{z};\beta)$, involves the Z's but does not involve t. In contrast, the baseline hazard $h_0(t)$, is only a function of the time t.

A time independent variable is defined to be any variable whose values for a given subject do not change over time. For instance, SEX and weight are such variables. Note that even a person's weight may actually change over time, it may be appropriate to treat this variable, for analysis purposes, as time independent if its values do not change much over time or if the effect of the variable (weight here), on survival risk depends essentially on the value at only one measurement.

5.3 The Likelihood function

Leaving the baseline hazard function $h_0(t)$ arbitrary, the loss of information about the unknown parameters β is usually slight. On the other hand, the analysis of the relative efficiency of inferences about β under the various assumptions about $h_0(t)$, is a major problem. Arbitrariness of $h_0(t)$ may contribute little or no information about β by the intervals in which no failures occur. One example is that the component $h_0(t)$ might be identically zero in such intervals. For that reason, the probabilities used for the likelihood manipulation are

conditioned on the I_j set, defined in Section 2.3.

Initially we consider n failures $t_1, t_2, \ldots, t_n$, and their order types $\tau_1, \tau_2, \ldots, \tau_n$. The $\{\tau_j\}$ and $\{I_j\}$ are jointly equivalent to the original data, namely the unordered failures t_i. The conditional probability that I_j=i given the entire history

$$\bar{H}_j = \{\tau_1, \tau_2, \ldots, \tau_j, i_1, i_2, \ldots, i_{j-1}\}$$

up to the jth ordered failure time τ_j can be written down explicitly. In the history expression above, the i_k is the index of the subject that failed at τ_k time point. The

probability already mentioned is actually the condition probability that i fails at τ_j given that one subject from the risk set $\Re(\tau_j)$fails at τ_j, which is simply, from the definition of probability

$$\frac{h_i(\tau_j)}{\sum\limits_{k\in\Re(\tau_j)} h_k(\tau_j)} = \frac{\psi(i)}{\sum\limits_{k\in\Re(\tau_j)} \psi(k)} \tag{5.5}$$

The baseline hazard function $h_0(\tau_j)$ is cancelled out because of the form of the hazard rate given in formula (5.1). For notational reasons, $\psi(k)$ in (5.5) denotes $\psi(Z_k,\beta)$, that is, the multiplier ψ for the *kth* subject.

Although (5.5) was derived as the conditional probability that I_j=i given the entire history $\bar{H}_j$, in fact it is functionally independent of the ordered failures τ_j. Therefore, it equals to $pr_j(I_j = i|I_1 = i_1, I_2 = i_2, \ldots, I_{j-1} = i_{j-1})$.

The joint distribution $pr_j(i_1,i_2,\ldots,i_{j-1})$ can therefore be obtained by the usual chain rule for conditional probabilities as

$$pr(i_1,i_2,...,i_n) = \prod_{j=1}^{n} pr_j(i_j|i_1,i_2,...,i_{j-1}) = \prod_{j=1}^{n} \frac{\psi(i_j)}{\sum\limits_{k\in\Re(\tau_j)} \psi(k)} \tag{5.6}$$

In the presence of censoring, a similar argument applies if it can be assumed that censoring can only occur immediately after failures. The last assumption does not conflict with the model in which the censoring times are fixed constants. In contrast, it can be considered as a reasonable approximation, as the information about β when observed censoring time c_i is involved will generally be small. The fixed censoring model can be handled explicitly through the partial likelihood, as we will see in the next section.

Supposing now that there are d observed failures from the sample of size n, and using the same notations with the censoring absence case, we have that equation (5.5) follows exactly as before, where the history set $\bar{H}_j$ now includes the censoring in $(0,\tau_j)$ as well as the failures. The risk set $\Re(\tau_j)$, and hence the expression (5.6), does not depend on τ_j, as no censoring is assumed to occur in the time interval (τ_{j-1},τ_j). Combination of

these conditional probabilities gives the overall likelihood:

$$lik = \prod_{j=1}^{d} \frac{\psi(i_j)}{\sum_{k \in \Re(\tau_j)} \psi(k)} \tag{5.7}$$

Terms that determine which subjects should be censored from among the survivors of each risk set are omitted. In this sense, (5.7) can be considered as likelihood rather than a probability. In addition, unless the censored mechanism itself depends on β, these terms, in the likelihood, do not depend functionally on â and can be ignored.

5.4 Inference for the parameters

Even though a precise form of $\psi(z;\beta)$ is not essential, for the derivation of the (5.7), we will focus on the log linear form, which is the most commonly used in the survival analysis. Besides, the results derived considering the pre mentioned form, can be generalized and further applied for the other forms.

5.4.1 Continuous case

In particular, using the property that the exponential function remains inalterable to any derivation we get that

$$\frac{\partial}{\partial \beta_r}\psi(z;\beta) = \psi_r(i) = z_{ir}\psi(i) \; and \frac{\partial^2}{\partial \beta_r \partial \beta_s}\psi(z;\beta) = \psi_{rs}(i) = z_{ir}z_{is}\psi(i)$$

where z_{ir} denotes the value of the rth component of the explanatory vector $\mathbf{z}$ on the ith subject. From (5.7) we have also that

$$l = \log lik = \log \prod_{j=1}^{d} \frac{\psi(i_j)}{\sum_{k \in \Re(\tau_j)} \psi(k)} = \sum_{j=1}^{d} l_j = \sum_{j=1}^{d} \left[\log[\psi(i_j)] - \log\left(\sum_{k \in \Re(\tau_j)} \psi(k) \right) \right] \tag{5.8}$$

Where

$$l_j = \log\left[\psi\left(i_j\right)\right] - \log\left(\sum_{k\in\Re(\tau_j)} \psi\left(k\right)\right) \tag{5.9}$$

By specifying the $\psi(i)$ expression and substituting both the ordered failures by random ones and the risk sets $\Re\left(\tau_i\right) = \Re_i$, the derivatives of the components of the likelihood l_i and the same for l_j are now become

$$l_i = \log\left[\psi\left(i\right)\right] - \log\left(\sum_{k\in\Re_i} \psi\left(k\right)\right) = \beta^T z_i - \log\left(\sum_{k\in\Re_i} \exp\left(\beta^T z_k\right)\right) \tag{5.10}$$

$$\frac{\partial l_i}{\partial \beta_r} = z_{ir} - \frac{\sum\limits_{k\in\Re_i} z_{kr} e^{\left(\beta^T z_k\right)}}{\sum\limits_{k\in\Re_i} e^{\left(\beta^T z_k\right)}} \tag{5.11}$$

$$\frac{\partial^2 l_i}{\partial \beta_r \partial \beta_s} = -\frac{\sum\limits_{k\in\Re_i} z_{kr} z_{ks} e^{\left(\beta^T z_k\right)}}{\sum\limits_{k\in\Re_i} e^{\left(\beta^T z_k\right)}} + \frac{\sum\limits_{k\in\Re_i} z_{kr} e^{\left(\beta^T z_k\right)}}{\sum\limits_{k\in\Re_i} e^{\left(\beta^T z_k\right)}} \frac{\sum\limits_{k\in\Re_i} z_{ks} e^{\left(\beta^T z_k\right)}}{\sum\limits_{k\in\Re_i} e^{\left(\beta^T z_k\right)}} \tag{5.12}$$

The expectations of these quantities when i is sampled, from the risk set $\Re_i$, with probability proportional to $\exp\left(\beta^T z_k\right)$ are calculated. Taking into account that the first order derivatives (5.11) are score functions, we can conclude that $E\left(\frac{\partial l_i}{\partial \beta_r}\right) = 0$. As required the expectation is taken as the same value of β over a single risk set as is used in evaluation of $\frac{\partial l_i}{\partial \beta_r}$. Taking expectations of the second derivatives we have that:

$$-E\left(\frac{\partial^2 l_i}{\partial \beta_r \partial \beta_s}\right) = \frac{\sum\limits_{k\in\Re_i} z_{kr} z_{ks} e^{\left(\beta^T z_k\right)}}{\sum\limits_{k\in\Re_i} e^{\left(\beta^T z_k\right)}} - \frac{\sum\limits_{k\in\Re_i} z_{kr} e^{\left(\beta^T z_k\right)}}{\sum\limits_{k\in\Re_i} e^{\left(\beta^T z_k\right)}} \frac{\sum\limits_{k\in\Re_i} z_{ks} e^{\left(\beta^T z_k\right)}}{\sum\limits_{k\in\Re_i} e^{\left(\beta^T z_k\right)}} = cov\left(\frac{\partial l_i}{\partial \beta_r}, \frac{\partial l_i}{\partial \beta_s}\right) \tag{5.13}$$

Notice that the observed (5.12) and the expected values of $\frac{\partial^2 l_i}{\partial \beta_r \partial \beta_s}$ above, taken over a risk set as already mentioned, are identical. Summing over all risk sets, we get the rth

element of the score function $U_r(\beta)$

$$U_r(\beta) = \sum_{i \in \wp} \left(z_{ir} - \frac{\sum_{k \in \Re_i} z_{kr} e^{(\beta^T z_k)}}{\sum_{k \in \Re_i} e^{(\beta^T z_k)}} \right) \tag{5.14}$$

Moreover the (r,s) element of the Fisher information matrix is

$$I_{rs}(\beta) = \sum_{i \in \wp} C_{irs}(\beta) \tag{5.15}$$

where

$$C_{irs}(\beta) = \frac{\sum_{k \in \Re_i} z_{kr} z_{ks} e^{(\beta^T z_k)}}{\sum_{k \in \Re_i} e^{(\beta^T z_k)}} - \frac{\sum_{k \in \Re_i} z_{kr} e^{(\beta^T z_k)}}{\sum_{k \in \Re_i} e^{(\beta^T z_k)}} \frac{\sum_{k \in \Re_i} z_{ks} e^{(\beta^T z_k)}}{\sum_{k \in \Re_i} e^{(\beta^T z_k)}}$$

The $\wp$ set used in the last formula is defined as those subjects who fails. These expectations and covariances are conditional on the composition of the risk set, in which they are considered. Calculation of fully unconditional expectations would require a fuller specification of the censoring mechanism. For instance, given a data set, involving the times at which individuals, who in fact failed, would have been censored for the calculation of the expectation of (5.15), is an irrelevant approach.

The expectations just calculated, can, though, be taken as conditional on the entire history of failures and censoring up to t_i, and this, in turn, allows a direct verification that the terms l_i behave closely, up to a degree, as a log likelihood function. That is asymptotic arguments for hypothesis testing and definitions of confidence intervals procedures are utilized. Thus, use of $I(\beta)$ rather than $E(I(\beta))$ is appropriate.

The maximum-likelihood estimates of β can be obtained by iterative use of (5.14) and (5.15). Significance tests about subsets of parameters can be derived by the use of the maximum log likelihood techniques. For instance, the likelihood ratio test, the score test and the direct use of the likelihood estimates can be applied.

In the absence of censoring or if the censoring mechanism is independent of the

explanatory variables, an exact test of the null hypothesis β=0 can be obtained by using the score statistic or else **Wald** statistic

$$W = U(0)^T (I(0))^{-1} U(0) \tag{5.16}$$

Particularly W is an expression of $U(0)$ which is asymptotically normal distributed with zero mean vector and covariance matrix $I(0)$. Therefore, the W statistic, under the null hypothesis, has an asymptotic chi-squared distribution with degrees of freedom equal to the number p of the explanatory variables. The score function $U(0)$ is given by:

$$U_r(0) = \sum_{i \in p} \left[z_{ri} - \frac{\sum_{k \in \Re_i} z_{kr}}{r_i} \right]$$

Actually (5.16) expresses the distribution of $U_r(0)$ generated when the ordered failure times $\tau_{(1)},\ldots,\tau_{(d)}$ and the sizes of the corresponding risk sets $r_1,\ldots,r_d$ are taken as fixed and the n values $z_1,\ldots,z_n$ of the explanatory variables are permuted randomly among the n subjects. Also the expression $\frac{\sum_{k \in \Re_i} z_{kr}}{r_i}$ is the mean of z_r over $\Re(\tau_i) = \Re_i$.

Further, from (5.15) it is implied that,

$$I_{rs}(0) = \sum_{i \in p} C_{irs}(0)$$

Where $C_{irs}(0)$ is the covariance of z_r and z_s in the finite population $\Re(\tau_i) = \Re_i$.

For computational reasons, the distribution of U(0) is re-expressed by

$$U(0) = \sum_{i=1}^{n} q_i z_i,$$

Given by Cox & Oakes (1984), where $q_i = \delta_i - \sum_{j:\tau_j \leq \tau_i} \frac{1}{r_j}$ and $\delta_i = 0$ or 1 accordingly as the ith individual is censored or not. Also by the definition of the sum $\sum_{i=1}^{n} z_i \left(\sum_{j:\tau_j \leq \tau_i} \frac{1}{r_j} \right) \equiv \sum_{j=1}^{n} \frac{1}{r_j} \left(\sum_{j:\tau_j \leq \tau_i} z_i \right)$. Thus by setting z_i=1 in the last quantity,

we see that $\sum q_i = 0$, so that $E[U(0)] = 0$.

Finally the covariance matrix of new $U(0)$ is

$$\tilde{I}(0) = \frac{1}{n-1}\left(\sum_{i=1}^{n}(z_i - \bar{z})(z_i - \bar{z})^T\right)\left(\sum_{i=1}^{n} q_i^2\right),$$

Where $\bar{z}$ is the mean covariate among the patients.

The last covariance matrix differs from I(0) and is valid, only under the assumption that the censoring mechanism is the same in the two groups.

5.4.2 Discrete case

Even though, the presence of ties complicates the log likelihood manipulation, they are usually recorded in survival analysis. Based on Cox & Oakes (1984), the model (5.5) can be generalized to discrete time as

$$\frac{h(t,z)}{1-h(t,z)} = \psi(z;\beta)\frac{h_0(t)}{1-h_0(t)} \tag{5.17}$$

Where, h(t,z) is given by the (2.5). Assuming k ordered failures $\tau_1,\tau_2,\ldots,\tau_k$ and the definition of the multiplicity m_i=$m_{(i)}$ (section 2.3) , the history $\mathcal{H}_j$ now includes the multiplicity of failure times up to and including τ_j. The conditional probability p that individuals $i_1,i_2,\ldots,i_k$ fail from the risk set $\Re(\tau_i)$ given $\mathcal{H}_j$ is, considering the log linear type for ψ:

$$p = \frac{\exp\left(\beta^T z_{i1}\right)\exp\left(\beta^T z_{i2}\right)\ldots\exp\left(\beta^T z_{im}\right)}{\sum_{d\in s(j;m)}\exp\left(\beta^T z_{d1}\right)\exp\left(\beta^T z_{d2}\right)\ldots\exp\left(\beta^T z_{dm}\right)} \tag{5.18}$$

Where $s(j;k)$ denotes the set of all selections of m_j items from the risk set $\Re(\tau_i)$ of size r_j=r. The probability in (5.18) contributes a single failure time. Due to the dependence on multiplicity $m = m_j$, the product of all such terms is no longer a marginal likelihood of ranks, but the method of partial likelihood must be used to justify the asymptotic theory. In our case, where the log linear form of the ψ expression is used, the likelihood

is simplified, as previously, in

$$l = \sum_{j=1}^{k} \left[\beta^T s_j - \log \left(\sum_{d \in s(j;m)} \exp\left(\beta^T s_{jd}\right) \right) \right] \tag{5.19}$$

Where $s_j = z_{i1} + z_{i2} + \ldots + z_{im}$ is the sum of the vectors z_I over the individuals who actually fail at $t_{(j)}$, each s_{jk} is the corresponding sum over a m length vector $(d_1, d_2, \ldots, d_m)$ of subjects who might have failed at $t_{(j)}$. Like before, estimates of the unknown parameters are obtained recursively by the equation (5.19), even when r, the card of the risk set and m are too large.

Furthermore, another approach to the case where ties are present is to consider ties as arising out of the grouping of survival times that are generated from the continuous-time model. Unfortunately, in this case, the resulting likelihoods are different. For that reason the summation of all the terms of the marginal likelihood (5.7) is suggested. Those terms are used whose ranks are coming from the continuous model and are consistent with the observed data. Unfortunately, the log likelihood is difficult to compute if several risk sets have values of r and d that are all large.

Instead the approximate likelihood is obtained by multiplying all the sums in the denominator to include all terms in the corresponding risk set.

$$\frac{d! \psi(i_1) \psi(i_2) \cdots \psi(i_m)}{\left(\sum_{k \in \Re(\tau_j)} \psi(k) \right)^m} \tag{5.20}$$

This approximation is widely used and is quite satisfactory except when the data exhibit heavy ties. Also the multiple counting of failed individuals does result in a conservative bias.

Two others simple estimators have been suggested to overcome this problem. Both involves the replace of the denominator of the (5.19) equation by the quantities

$$\prod_{j=1}^{m}\left(\sum_{k\in\Re(\tau)}\psi(k)-\frac{(j-1)}{d}\sum_{i\in\wp_j}\psi(k)\right)^{m}$$
$$\left(\sum_{k\in\Re(\tau)}\psi(k)-\frac{(d-1)}{2d}\sum_{k\in\wp_j}\psi(k)\right)^{m}$$

Where $\wp_j$ is the set of individuals who fail at τ_j. Either type does not support the case when ties are between reported censoring and failures. The usual convention is to assume that all failures reported at random time τ precede any censoring reported at τ, so that the censored subjects contribute fully to the corresponding risk sets.

In addition, in the absence of ties all the pre-mentioned suggestions give the same likelihood as this considering the continuous model. With very heavy ties, so that the survival times are grouped into a small number of intervals, it becomes feasible to devote a separate nuisance parameter π_j to the conditional baseline survivor function for each interval and to carry out a full maximization of the log likelihood in both the π_j and the regression parameters β. Note that, for sensible results, the total number of reported failures d must be much larger than the number g of grouping intervals. Asymptotic results require that the sample size n$\rightarrow\infty$ under the fixed grouping intervals.

5.5 Inference for the Baseline Hazards Functions

The estimation of the baseline hazard function $h_0(t)$ is essential when the survivor function is used in the model for the analysis. Moreover, $h_0(t)$ can be used in graphical procedures for goodness of fit checking.

5.5.1 Estimation

Firstly, we assume that $h_0(t)$ is expressed parametrically, as $h_0(t,\phi)$. The m.l.e. estimator for ϕ is obtained by the joint log likelihood $l(\beta,\phi)$ by either a direct joint maximization

or the conditional likelihood maximization of $l(\beta,\phi;\beta=\hat{\beta})$.

Secondly, we can use nonparametric estimation techniques for the baseline hazard integration $H_0(t)=\int_0^t h_0(u)\,du$.

Particularly, the estimator suggested by Cox & Oakes (1984) is

$$\hat{H}_0(t)=\sum_{\tau_j<t}\frac{d_j}{\sum_{i\in\Re(\tau_j)}\hat{\psi}(l)}$$

Where the $\hat{\psi}(l)$ are the estimated values of $\psi(l)$. Then the baseline survivor function $\bar{F}_0(t)$ can be estimated by $\bar{F}_0(t)=\exp\left[-\hat{H}_0(t)\right]$. Moreover, the estimators for the hazard rate and the survivor function for subject i are respectively

$$\hat{H}_i(t)=\hat{\psi}(i)\,\hat{H}_0(t) \text{ and } \bar{F}_i(t)=\left[\bar{F}_0(t)\right]^{\hat{\psi}(i)}$$

5.5.2 Goodness of fit test

- Using the exponential behavior of the cumulative hazard function, we can compare it with a sample from the unit exponential distribution. Especially, when the survival times T_i are transformed by the true cumulative hazard functions $H_i(t)$, then the vector of the values of $H_i(T_i)$ establish a sample from the exponential distribution. The same results are derived when the corresponding estimated cumulative hazard is involved. The last are called generalized residuals; see Cox & Oakes(1984)

- Familiar to the p-p plot used for goodness of fit testing in OLS case, is the plot of the ordered values of $\hat{H}_i(T_i)$ versus the expected values of the unit exponential distribution. More usefully, separate plots may be made for subjects of the data defined by the explanatory variables. Such plots are rather useful in the choice of the suitable explanatory variables for the model. On the other hand, the presence of correlation between the residuals, introduced for instance by the estimation of the multipliers ψ and the cumulative hazard, indicates the caution is needed in the

interpretation of the results from them.

- For the two-sample case, Wei (1984) proposed the test statistic

$$\begin{aligned} T_n &= \left\{\hat{\theta}\hat{\eta}(\infty)\right\}^{-1/2} \sup_{0\leq t<\infty} |W_n(t)| \\ W_n(t) &= n^{-1/2} U_n\left(\hat{\theta}; t\right) \end{aligned}$$

which tests the null hypothesis

$$H_0 : H_1(t) = \theta H_2(t)$$

that the cumulative hazards functions of the two-samples are proportional. The score function $U_n(\theta; t)$ is derived from the log-likelihood and equals to

$$\begin{aligned} U_n(\theta; t) &= \int_0^t dN_1(s) - \int_0^t \frac{Y_1(s)\,\theta}{Y_1(s)\,\theta + Y_2(s)} \times d\{N_1(s) + N_2(s)\} \\ &= \sum_{j=1}^{n_1} \delta_{1j} I\left(\tilde{X}_{1j} \leq t\right) - \sum_{i=1}^{2}\sum_{j=1}^{n_1} \frac{Y_1\left(\tilde{X}_{ij}\right)\theta}{Y_1\left(\tilde{X}_{ij}\right)\theta + Y_2\left(\tilde{X}_{ij}\right)} I\left(\tilde{X}_{ij} \leq t\right) \end{aligned}$$

Also the estimator of $\hat{\theta}$ is the solution of the equation

$$\theta^{-1} U_n(\theta; t) = 0$$

whereas, $\tilde{X}_{ij}$ is the observed data from the ith group (i=1,2) and the jth individual.

$$\begin{aligned} N_i(t) &= \#\left\{j : \tilde{X}_{ij} \leq t \text{ and } \delta_{ij} = 1\right\} \\ Y_i(t) &= \#\left\{j : \tilde{X}_{ij} \geq t\right\} \end{aligned}$$

$$y_i(t) = (1 - F_i(t))\left(1 - L_i\left(t^-\right)\right)$$
$$L_n(\theta) = \max_{H_2} L_n(\theta; H_2)$$

and $L_n(\theta; H_2)$ is a joint likelihood for the unknown parameters θ and H_2. In addition, $\hat{\eta}(t)$ can be written explicitly in the form

$$\hat{\eta}(t) = n^{-1}\left\{\int_0^t \frac{Y_1(s)\,Y_2(s)}{\left(Y_1(s)\,\hat{\theta} + Y_2(s)\right)^2} d\left\{N_1(s) + N_2(s)\right\}\right\}$$
$$= n^{-1}\sum_{i=1}^{2}\sum_{j=1}^{n_1} \frac{\delta_{ij} Y_1\left(\tilde{X}_{ij}\right) Y_1\left(\tilde{X}_{ij}\right)}{\left(Y_1\left(\tilde{X}_{ij}\right)\hat{\theta} + Y_2\left(\tilde{X}_{ij}\right)\right)^2} I\left(\tilde{X}_{ij} \le t\right)$$

- Another, standardized test statistic testing the null hypothesis

$$H_0 : h_1(t) = \theta h_2(t)$$

is presented by Gill and Schumacher (1987), and is denoted by

$$T_{K_1K_2} = \left\{est\ var\left(Q_{K_1K_2}\right)\right\}^{-1/2} Q_{K_1K_2}$$

As $Q_{K_1K_2}$ is a symmetric quantity

$$Q_{K_1K_2} = \hat{K}_{11}\hat{K}_{22} - \hat{K}_{21}\hat{K}_{12}$$

where K_{ij} is the integral of some predictable random weight functions, as given in the aforementioned article.

The statistic $T_{K_1K_2}$ follows asymptotically the standardized normal distribution.

- Moreover, Lin (1991), suggests the consistent statistic

$$Q_w = n\left(\hat{\beta}_w - \hat{\beta}\right)^T D_W\left(\hat{\beta}\right)^{-1}\left(\hat{\beta}_w - \hat{\beta}\right)$$

which has an asymptotic central chi-square distribution on p degrees of freedom. $\hat{\beta}_w$ and $\hat{\beta}$ are the estimators of the real $p \times 1$ parameter vector value β_0. Despite the fact that both are derived from the likelihood equations of the score functions, $\hat{\beta}_w$ is taking place when unequal weights are assigned to different failures according to the times of their occurrences.. As $D_W\left(\hat{\beta}\right)$, the covariance matrix of the random variable $n^{1/2}\left(\hat{\beta}_w - \hat{\beta}\right)$, is defined. The null hypothesis that is tested in this case, can be written as

$$H_0 : h(t; Z) = h_0(t)\, e^{\beta_0 Z(t)}$$

against the alternative

$$H_0 : h(t; Z) = h_0(t)\, e^{\theta(t) Z(t)}$$

Where $\theta(t)$ is an unspecified monotone function of t.

- Based on the differences between the counting processes and their respective integrated intensity functions, Lin et al. (1993), proposed model checking techniques for the Cox's proportional model. Especially, they make use of the statistics

$$\sup_t \left\| U\left(\hat{\beta}, t\right) \right\| \text{ and}$$
$$\sup_t \sum_{j=1}^{p} \left\{ \Im^{-1}\left(\hat{\beta}\right)_{jj} \right\}^{1/2} \left| U_j\left(\hat{\beta}, t\right) \right|$$

to check the proportional hazards assumption.

Where $U\left(\hat{\beta},t\right)$ and $\Im\left(\hat{\beta}\right)_{jj}$ are the partial likelihood score function, and the minus derivative of the previous, respectively, as well defined in Lin et al. (1993).

5.6 Some Properties and Considerations of the popularity of the Cox Proportional Hazard Model

The Cox model (5.1), considering the log linear form of ψ, has the property that when all the explanatory variables are equal to zero, that is we have information up to the failure times and the censoring, it reduces to the baseline hazard function. In other words, the baseline hazard when no Z's are in the model can be regarded as a starting or "baseline" version of the hazard function, prior to considering any of the Z's. Furthermore, by the definition of (5.1) model, the baseline function $h_0(t)$ is unspecified. This property allows Cox model to be a **non-parametric** one. The randomness of $h_0(t)$, does not halter the estimation of neither the unknown parameters β, nor the hazard and the survival functions $h(t,Z)$, $S(t,Z)$. Thus, with the Cox model, using a minimum of assumptions, primary information desired from a survival analysis can be obtained.

A key reason for the popularity of the Cox model is that, even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained for a wide variety of data. Indeed, the Cox model is a "robust" model, so that the results from using the Cox model will closely approximate the results for the correct parametric model. The use of a specific parametric model is preferable when the exact statistical representation of the data. Although there are various methods for assessing goodness of fit of a parametric model, the final model selection can always be disputed. Hence, when in doubt, as usually, the Cox model will give reliable enough results, so that the choice of the model is now "safer".

Chapter 6

Generalization of the linear transformation models

6.1 Introduction

In this chapter a class of (recently developed by Cheng et al (1995)) semi-parametric transformation models, under which an unknown transformation of the survival probability equals the sum of an increasing function plus the linear predictor, are presented. The proportional hazards and proportional odds models are included in this class. A class of generalized survival time is linearly related to the covariates with various completely specified error distribution. Nearby, estimating equations is proposed to examine the covariate effects with censored observations. Also, a simple modification of these methods is used, to face the asymptotically bias, caused when the support of the censoring variable is shorter than that of the failure time. An example study is used to investigate the properties of the proposals.

6.2 Model structure

We may rewrite the Cox proportional model of (5.2) in the form

$$\log\left(-\log\left(\bar{F}_Z(t)\right)\right) = h(t) + Z^T\beta \tag{6.1}$$

where h(t) is a completely unspecified strictly increasing function, $\bar{F}_Z(t)$ is the survival function of failures T given the explanatory variables z and β is a p×1 vector of unknown regression coefficients. Inference about β can be based on the partial likelihood function, as shown in Chapter 5.

An alternative is the proportional odds model, presented in equation (4.31). The latter can be rewritten in another more convenient form, using the logit function Pettitt (1982). Thus, the relation

$$-\log it\left\{\bar{F}_Z(t)\right\} = h(t) + Z^T\beta \tag{6.2}$$

is equivalent to the previous definition (4.31), regarding that the logit function is exactly the logarithm of the ratio of variable divided by its complement $\left(\log it(z) = \log\left(\frac{z}{1-z}\right)\right)$. Despite the fact (6.2) is a model commonly used, scarcely is found any theoretical justification for the large sample properties of inference procedures for parameter β in the literature, apart from the simple two-sample case (Dabroswka & Doksum, (1988)).

The generalization of the two last models given by Cheng et al (1995) is

$$g\left\{\bar{F}_z(t)\right\} = h(t) + Z^T\beta \tag{6.3}$$

where g(.) is a known decreasing function. Applying the random variable of the failures T in the model (6.3), and solving in respect to the $h(.)$ function, we get

$$h(T) = -Z^T\beta + g\left\{\bar{F}_z(T)\right\} \tag{6.4}$$

In (6.4) the $g\left\{\bar{F}_z(T)\right\}$expression is a random variable and can be renamed in a random error form. In other words

$$h(T) = -Z^T\beta + \varepsilon \tag{6.5}$$

The distribution function of the error component ε is $F(.) = 1 - g^{-1}(.)$. We can verify that by applying the last function to the random error $g\left\{\bar{F}_z(T)\right\}$.

$$F\left(g\left\{\bar{F}_z(T)\right\}\right) = 1 - \bar{F}_z(T) = F_T$$

which is a distribution cumulative function. If $F(t)$ equals to $1 - e^{\left(-e^{(t)}\right)}$, then (6.5) is the proportional hazards model, while if F is coming from the standard logistic distribution, (6.5) is the proportional odds model.

A class of simple estimating functions for β in the linear transformation model (6.5) will be presented in the sequel. The hypothesis of censored data may complicate the analysis but should be included as is mostly the case in survival data.

6.3 Estimation for the Linear Transformation Model

6.3.1 General remarks

We consider the symbols and notions introduced in Section (2.3). The censoring time c_i is assumed to be independent of T_i. Again, let the $p \times 1$ vector Z_i be the corresponding covariate vector for the ith individual. In addition the survival function of the censored lifetime $\bar{\mathrm{C}}_i$ is assumed to be independent of the failures T_i. In the case the covariate vector Z has a finite number of possibly values, the latter assumption can be relaxed.

The function h(.) maps the positive half-time onto the whole real line, like the natural logarithm. So the set $\{h(T_i), i = 1, \ldots, n\}$ has the same rank configuration of that of $\{T_i\}$. In other words, $h(T_i)$ is a one-one transformation of the failure times into a new random variable. Thus, as the $h(T_i)$ does not involve the parameter of interest β, it seems natural to use the marginal likelihood (Cox & Hinkley (1996)) to make inferences about

β. On the other hand, the maximum likelihood estimate and its variance, are difficult to be obtained numerically.

6.3.2 Generalized Estimating Equations

Using the indicator function $I(.)$ as in Cheng et al (1995), we consider the dichotomous variables

$$\{I(T_i \geq T_j), i \neq j = 1, \ldots, n\} = \begin{cases} 1, & if\ T_i \geq T_j \\ 0, & \text{otherwise} \end{cases} \tag{6.6}$$

Then,

$$E(I(T_i \geq T_j) | Z_i, Z_j) = pr(h(T_i) \geq h(T_j) | Z_i, Z_j)$$

Using (6.5) the last expected value becomes

$$E(I(T_i \geq T_j) | Z_i, Z_j) = pr(\varepsilon_i - \varepsilon_j \geq Z_{ij}^T \beta_0)$$

where β_0 is the true value for β and so can replace the values of β_1 and β_2. Also, the new symbol Z_{ij} equals to the difference of the explanatory variables $Z_i - Z_j$.

Considering the probability $pr(\varepsilon_i - \varepsilon_j \geq Z_{ij}^T \beta_0)$ as a function of $Z_{ij}^T \beta_0$, i.e.

$$pr(\varepsilon_i - \varepsilon_j \geq Z_{ij}^T \beta_0) = \xi(Z_{ij}^T \beta_0)$$

we can calculate this probability by using its geometric meaning. Thus, since F a well defined and differentiable function, we find that

$$\xi(s) = \int_{-\infty}^{\infty} \left(\int_{s+\varepsilon_j}^{\infty} f(\varepsilon_i) f(\varepsilon_j) d\varepsilon_i \right) d\varepsilon_j$$

$$= \int_{-\infty}^{\infty} \left(\int_{s+\varepsilon_j}^{\infty} f(\varepsilon_i)\, d\varepsilon_i \right) f(\varepsilon_j)\, d\varepsilon_j$$
$$= \int_{-\infty}^{\infty} (1 - F(s+\varepsilon_j))\, dF(\varepsilon_j)$$

which means

$$\xi(s) = \int_{-\infty}^{\infty} \{1 - F(t+s)\}\, dF(t)$$

where F is the completely specified cumulative distribution function of ε. Note, that the dichotomous variables in (6.6) are dependent, as for instance the two sets $\{T_1 \geq T_2\}$ and $\{T_2 > T_1\}$ are complemental. Despite the dependence of the variables, one may make inferences about β_0 based on generalized estimating equations (Liang & Zeger (1986)).

As a consequence, assuming that the dichotomous variables are independent, the resulting estimating function is

$$\bar{U}(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} w\left(Z_{ij}^T \beta\right) Z_{ij} \left[I(T_i \geq T_j) - \xi\left(Z_{ij}^T \beta\right) \right] \tag{6.7}$$

where w(.) is a weight function. Even though the dichotomous variables are dependent, $E\left\{\bar{U}(\beta_0)\right\} = 0$. the last equation, suggests that a solution to $\bar{U}(\beta) = 0$ is a reasonable estimator for β_0. A common way to tackle the weight component, is to assume w(.) = 1, and to proceed with a linear regression technique. Another approach, is that of the quasi likelihood, for independent observations. Especially, we take

$$w(.) = \frac{\xi'(.)}{v(.)} \tag{6.8}$$

where $v(.) = \xi(.)\{1 - \xi(.)\}$.

In case of censored data, the indicators $\{I(T_i \geq T_j)\}$ in (6.7) are not always observable. It can be shown that

$$E\left\{\frac{V_j I\left(X_i \geq X_j\right)}{\bar{C}^2\left(X_j\right)} \left| Z_i, Z_j \right.\right\}$$

$$= E\left(E\left[\frac{I\left(T_i \geq T_j\right)\left\{I\left(\min\left(C_i, C_j\right)\right) \geq T_j\right\}}{\bar{C}^2\left(T_j\right)} \left| T_j, Z_i, Z_j \right.\right]\right) \tag{6.9}$$

where V_j is the indicator variable, specifying whether the jth observation is censored or not. Also, X_i is the observed information of an individual. That is, X_i coincides with the failure time T_i in the uncensored case, whilst, X_i equals to C_i in the censored case. So, equation (6.9) is derived by the use of $X_i = \min\left(T_i, C_i\right)$.

Also,

$$E\left(E\left[\frac{I\left(T_i \geq T_j\right)\left\{I\left(\min\left(C_i, C_j\right)\right) \geq T_j\right\}}{\bar{C}^2\left(T_j\right)} \left| T_j, Z_i, Z_j \right.\right]\right) = E\left[I\left\{h\left(T_i\right) \geq h\left(T_j\right)\right\} \left| Z_i, Z_j \right.\right]$$

The last formula indicates that the two expectation are equal, so instead of using the dichotomous variable $I(T_i \geq T_j)$ in the estimating function (6.7), the expression $\frac{V_j I(X_i \geq X_j)}{\bar{C}^2(X_j)}$ may be used. A plausible candidate of the survival function $\bar{C}(.)$ of the censoring variable is the Kaplan-Meier estimator $\hat{G}$. On that account, the resulting estimating function is now denoted by

$$U(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} w\left(Z_{ij}^T\beta\right) Z_{ij}\left[\frac{V_j I\left(X_i \geq X_j\right)}{\hat{G}^2\left(X_j\right)} - \xi\left(Z_{ij}^T\beta\right)\right] \tag{6.10}$$

In (Cheng et al (1995)) is shown that, when the weights w(.) are positive, then the equation $U(\beta) = 0$ has asymptotically, a unique solution $\hat{\beta}$. When w=1 and the observed matrix $\sum\sum Z_{ij} Z_{ij}^T$ is positive definite, which is trivially satisfied for most practical situations, the equation $U(\beta) = 0$ has a unique solution. In the situation where the F function has the extreme form $1 - e^{\left(-e^{(t)}\right)}$, the weight function (6.8) becomes identical to 1.

Also Cheng et al (1995) proved that the distribution of $n^{-\frac{1}{2}} U\left(\beta_0\right)$ can be approximated by a normal distribution with mean 0 and variance-covariance matrix $\hat{\Gamma}$, given

by

$$\hat{\Gamma} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n} d_{ij}d_{ik}Z_{ij}Z_{ik}^{T}$$

$$-\frac{4}{n^3}\sum_{l=1}^{n}\frac{1-V_l}{\{\sum_k I(X_k \geq X_l)\}^2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} w\left(Z_{ij}^{T}\hat{\beta}\right) Z_{ij}\frac{V_j I(X_i \geq X_j)}{\hat{G}^2(X_j)} I(X_j \geq X_l)\right\}^{\otimes 2}$$

where $d_{ij} = \left\{w\left(Z_{ij}^{T}\hat{\beta}\right)\hat{e}_{ij}\left(\hat{\beta}\right) - w\left(Z_{ji}^{T}\hat{\beta}\right)\hat{e}_{ji}\left(\hat{\beta}\right)\right\}$, $\hat{e}_{ij}\left(\hat{\beta}\right) = \frac{V_j I(X_i \geq X_j)}{\hat{G}^2(X_j)} - \xi\left(Z_{ij}^{T}\beta\right)$, and $u^{\otimes 2} = uu^{T}$ for a vector u. From the Taylor series expansion of $\mathrm{U}\left(\hat{\beta}\right)$ around β_0, we can conclude that $\mathrm{n}^{1/2}\left(\hat{\beta} - \beta_0\right)$ is asymptotically equivalent to $\mathrm{n}^{-3/2}\hat{\Lambda}U(\beta_0)$, where

$$\hat{\Lambda}^{-1} = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} w\left(Z_{ij}^{T}\hat{\beta}\right)\xi'\left(Z_{ij}^{T}\hat{\beta}\right)Z_{ij}^{\otimes 2}$$

So, the distribution of $\mathrm{n}^{1/2}\left(\hat{\beta} - \beta_0\right)$ can be approximated by a normal distribution with mean 0 and covariance matrix $\hat{\Sigma} = \hat{\Lambda}\hat{\Gamma}\hat{\Lambda}$.

Assumptions in the parameter β can be made, using the above procedures, only when the distribution of the censoring variable C is free of the covariate vector Z. This assumption, is often satisfied in randomized controlled clinical trials.

Now suppose that we can partition Z into k possible values. Then an analogue of the estimating function (6.10) is

$$U^{*}(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} w\left(Z_{ij}^{T}\beta\right) Z_{ij}\left[\frac{V_j I(X_i \geq X_j)}{\hat{G}_{Z_i}(X_j)\,\hat{G}_{Z_j}(X_j)} - \xi\left(Z_{ij}^{T}\beta\right)\right] \qquad (6.11)$$

where $\hat{G}_Z(.)$ is the Kaplan-Meier estimator for the survival function of the censoring variable C based on those pairs $\{X_l, V_l\}$ whose $\mathrm{Z}_l = Z\,(l = 1, ..., n)$.

In the same article Cheng et al (1995) showed that the distribution of $\mathrm{n}^{3/2}U^{*}(\beta_0)$ can be approximated by a normal distribution with mean zero and covariance matrix

$$\Gamma^* = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n} d_{ij}^* d_{ik}^* Z_{ij} Z_{ik}^T$$
$$-\frac{4}{n^3}\sum_{l=1}^{n}\frac{1-V_l}{\left\{\sum_{k=1}^{n} I\left(X_k \geq X_l, Z_k = Z_l\right)\right\}^2}$$
$$\times\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} w\left(Z_{ij}^T\hat{\beta}\right) Z_{ij}\frac{V_j I\left(X_i \geq X_j\right)}{\hat{G}_{Z_i}\left(X_j\right)\hat{G}_{Z_j}\left(X_j\right)} I\left(X_j \geq X_l\right)\right\}^{\otimes 2}$$

Consequently, $n^{1/2}\left(\hat{\beta}-\beta_0\right)$ can be approximated also by a normal distribution with mean 0 and covariance matrix $\Sigma^* = \hat{\Lambda}\Gamma^*\hat{\Lambda}$ as before.

By replacing the $\hat{G}_Z(.)$ function in (6.11), by a nonparametric functional estimate like the K-M based on subjects whose covariates are in a small neighborhood of univariate Z, the corresponding estimator $\hat{\beta}$ is still consistent Cheng et al. (1995). Changing the size of the neighborhood, $\hat{\beta}$ can be showed to be asymptotically normal. In this way, we can deal the difficulties in discrediting the covariates.

6.4 A Modification in the Estimates

A modification of the previous estimator of β is given by Fine et al (1998). The last alteration, is shown to perform well in the case of heavy censoring, in which the Cheng et al (1995) estimators presented previously (6.10) & (6.11) are asymptotically biased.

$$h(T) = Z^T\beta + \varepsilon \tag{6.12}$$

Particularly, using as starting point the equation (6.12), Fine et al (1998) suggested that a consistent estimator for the parameters vector $\theta = \left(a, \beta^T\right)^T$ is the solution $\hat{\theta}_\omega$ of the equation $\tilde{U}_W(\theta) = 0$. The solution is unique for large sample size $(n \to \infty)$. The score function $\tilde{U}_w(.)$ is defined now as:

$$\begin{aligned}\tilde{U}_w(\theta) &= -\frac{1}{2}\frac{dQ_w(\theta)}{d\theta} \\ &= \sum\sum_{i\neq j} w_{ij}\left(\hat{\theta}\right)\check{\eta}_{ij}(\theta)\left[\frac{V_j I(\min(X_i,t_0)\geq X_j)}{\hat{G}^2(X_j)} - \eta_{ij}(\theta)\right]\end{aligned} \tag{6.13}$$

Giving justifications for the terms appeared in (6.13), we start with t_0, which is a known constant such that the probability $pr\{\min(T,C) > t_0\}$ is greater than zero. In addition, $Q_w(\theta)$ is the sum of squares which we want to minimize and equals.

$$Q_w(\theta) = \sum\sum_{i\neq j} w_{ij}\left(\hat{\theta}\right)\left[\frac{V_j I(\min(X_i,t_0)\geq X_j)}{\hat{G}^2(X_j)} - \eta_{ij}(\theta)\right]^2$$

As before, instead of using only the dichotomous variables, we can use the quantity $\frac{V_j I(\min(X_i,t_0)\geq X_j)}{\hat{G}^2(X_j)}$, as its expected value (conditioned on $Z_i$and Z_j) denoted by $\eta_{ij}(\alpha_0,\beta_0)$ is equal to the integral

$$\begin{aligned}\eta_{ij}(\alpha_0,\beta_0) &= \xi\left(Z_{ij}^T\beta\right) - pr(T_i \geq T_j \geq t_0 | Z_i, Z_j) \\ &= \int_{-\infty}^{a_0}\left\{1 - F\left(t - Z_{ij}^T\beta\right)\right\} dF\left(t - Z_{ij}^T\beta\right)\end{aligned} \tag{6.14}$$

Note here that as a_0 is taken the true value of the function range in t_0 $a = h(t_0)$ whereas as $h_0(.)$ the true function values $h(.)$, in general. Despite the same notation, in (6.13), the symbol V_i is 1 in the censor case and 0 otherwise. Also, $\check{\eta}_{ij}(.)$ is the vector of partial derivatives, of the expected values $\eta_{ij}(\theta)$ with respect to θ, given in (6.14) and equals to

$$\check{\eta}_{ij}(\theta) = \begin{pmatrix} 1 \\ -Z_j^T \end{pmatrix}\int_{-\infty}^{a}\left\{1 - F\left(t - Z_i^T\beta\right)\right\} df\left(t - Z_j^T\beta\right)$$

$$-\begin{pmatrix} 1 \\ -Z_i^T \end{pmatrix} \int_{-\infty}^{a} f\left(t - Z_i^T\beta\right) dF\left(t - Z_j^T\beta\right) \tag{6.15}$$

Finally, w_{ij} is a weight function. The distribution of $\hat{\theta}_\omega$ is approximated by a normal distribution with mean θ_0 and covariance matrix

$$n\hat{D}^{-1}\hat{\Gamma}\hat{D}^{-1} \tag{6.16}$$

where

$$\hat{D} = n^{-1}\sum\sum_{i\neq j} w_{ij}\left(\hat{\theta}_\omega\right)\check{\eta}_{ij}\left(\hat{\theta}_\omega\right)\check{\eta}_{ij}^T\left(\hat{\theta}_\omega\right)$$

$$\begin{aligned}\hat{\Gamma} = & \frac{1}{n^3}\sum\sum\sum_{i\neq j\neq k}\left(\hat{e}_{ij}+\hat{e}_{ji}\right)\left(\hat{e}_{ik}+\hat{e}_{ki}\right)^T - \frac{4}{n^3}\sum_{l=1}^{n}\frac{1-V_l}{\left\{\sum_k I\left(X_k \geq X_l\right)\right\}^2} \\ & \left\{\sum\sum_{j\neq i} w_{ij}\left(\hat{\theta}_\omega\right)\check{\eta}_{ij}\left(\hat{\theta}_\omega\right)\frac{V_j I\left(\min\left(X_i,t_0\right)\geq X_j\right)}{\hat{G}^2\left(X_j\right)} I\left(X_j \geq X_l\right)\right\}^{\otimes 2}\end{aligned}$$

$\hat{e}_{ij}$ is obtained by replacing θ and G in

$$e_{ij}\left(\theta\right) = w_{ij}\left(\theta\right)\check{\eta}_{ij}\left(\theta\right)\left[V_j I\left(\min\left(X_i,t_0\right)\geq X_j\right)\hat{G}^2\left(X_j\right) - \eta_{ij}\left(\theta\right)\right]$$

with $\hat{\theta}_\omega$ and $\hat{G}$, respectively.

6.5 Survivor Estimates

The non-decreasing function estimator $\hat{h}\left(t\right)$ is necessary to calculate the survival function $S_{Z_0}\left(t\right) = g^{-1}\left\{h\left(t\right) - Z_0^T\hat{\beta}_w\right\}$ for a given covariate vector Z_0. The root of the equation $V^*\left\{h\left(t\right)\right\} = 0$ is a non decreasing and consistent estimator of the true function $h_0\left(t\right)$ for $t \in [0,\tau]$, where $pr\left(X \geq \tau | Z\right) > 0$ for all Z, see Cheng et. al. (1997).

The score function $V^*\{h(t)\}$, in the case we consider the linear transformation model in (6.12), is given by the next formula, see Fine et. al. (1998)

$$V^*\{h(t)\} = \sum_{i=1}^{n}\left[\frac{I(X_i \geq t)}{\hat{G}_{Z_i}(t)} - g^{-1}\left\{h(t) - Z_i^T\hat{\beta}_w\right\}\right] \tag{6.17}$$

whereas, in the case the linear transformation model in (6.5) is utilized, the adjusted score function of the h function is now

$$V^*\{h(t)\} = \sum_{i=1}^{n}\left[\frac{I(X_i \geq t)}{\hat{G}_{Z_i}(t)} - g^{-1}\left\{h(t) + Z_i^T\hat{\beta}_w\right\}\right] \tag{6.18}$$

and the survival probabilities are given from the equation

$$S_{Z_0}(t) = g^{-1}\left\{h(t) + Z_0^T\hat{\beta}_w\right\}$$

The process

$$n^{\frac{1}{2}}\left[g\left\{\hat{S}_{Z_0}(t)\right\} - g\left\{S_{Z_0}(t)\right\}\right] = n^{\frac{1}{2}}\left\{\hat{h}(t) - h_0(t) - Z_0^T\left(\hat{\beta}_w - \beta_0\right)\right\}$$

can be approximated by the new process $\hat{\mathrm{W}}_{Z_0}$ given by

$$\hat{W}_{Z_0}(t) = \frac{1}{\hat{a}(t)}\left[\begin{array}{c}\left\{\hat{b}(t) + \hat{a}(t)Z_0\right\}^T\hat{H}\left\{\begin{array}{c} n^{\frac{-3}{2}}\sum\sum_{i\neq j}\hat{e}_{ij}(Y_i + Y_j) \\ +n^{\frac{1}{2}}\sum_{i=1}^{n}\int_0^{t_0}\frac{\hat{q}_{Z_i}(u)}{\hat{\pi}_{z_i}(u)}d\hat{M}_{Z_i}(u)\,Y_i\end{array}\right\} \\ +n^{\frac{1}{2}}\sum_{i=1}^{n}\hat{r}_i(t)\,Y_i + n^{\frac{1}{2}}\sum_{i=1}^{n}\frac{\hat{\pi}_{Z_i}(t)}{\hat{G}_{Z_i}(t)}\int_0^t\frac{1}{\hat{\pi}_{Z_i}(u)}d\hat{M}_{Z_i}(u)\,Y_i\end{array}\right]$$

where

$$\begin{aligned}\hat{r}_i(t) &= I(X_i \geq t)\left\{\hat{G}_{Z_i}(t)\right\}^{-1} - g^{-1}\left\{\hat{h}(t) - Z_i^T\hat{\beta}_w\right\} \\ \hat{M}_{Z_i}(t) &= I(X_i \leq t, V_i = 0) - \int_0^t I(X_i \geq u)\,d\hat{H}_{Z_i}(u) \\ \hat{e}_{ij}(\theta) &= \hat{e}_{ij}(Y_i + Y_j)\end{aligned}$$

and $\hat{H}_Z$ is the Nelson-Aalen estimator for the cumulative hazard H_Z based on observations $(X_i, 1-V_i)$ whose $Z_i = Z$ and i,j=1,...,n.

In order to construct confidence intervals for the survivals $S_{Z_0}(t)$, Fine et al (1998) suggested the quantity

$$g^{-1}\left[g\left\{\hat{S}_{Z_0}(t)\right\} \pm \phi_\xi n^{-\frac{1}{2}}\sigma_{Z_0}(t)\right] \tag{6.19}$$

where ϕ_ξ is the 100ξ upper percentage point of the standard normal distribution and the sample variance $\sigma_{Z_0}(t)$ is equal to

$$\sigma^2_{Z_0}(t) = J^{-1}\sum_{k=1}^{J}\hat{W}^2_{k,Z_0}(t)$$

The realization $\hat{W}^2_{k,Z_0}(t)$ is taken over J independent samples of $\{Y_i\}$.

In addition, in order to construct a $(1-2\xi)$ simultaneous confidence interval for $\{S_{Z_0}(t), a_1 \le t \le a_2\}$, the fixed quantiles d_ξ are calculated such that

$$pr\left[\sup_{t\in[a_1,a_2]}\left|\hat{W}_{Z_0}(t)\right|\{\sigma_{Z_0}(t)\}^{-1} \le d_\xi\right] = 1-2\xi \tag{6.20}$$

Then a $(1-2\xi)$ confidence band for $\{S_{Z_0}(t), a_1 \le t \le a_2\}$ is

$$g^{-1}\left[g\left\{\hat{S}_{Z_0}(t)\right\} \pm d_\xi n^{-\frac{1}{2}}\sigma_{Z_0}(t)\right]$$

The above probability and d_ξ are approximated with those J realizations of $\left\{\hat{W}_{Z_0}(t)\right\}$.

Last, when G is independent of the explanatory vector Z, t confidence intervals and bands for $S_{Z_0}(t)$ can be obtained by replacing

$$\hat{G}_{Z_i}(.) \Rightarrow \hat{G}(.)$$
$$\hat{\pi}_{Z_i}(u) \Rightarrow n^{-1}\sum_j I(X_j \ge u)$$

$$\hat{q}_{Z_i}(u) \Rightarrow 2n^{-2}\sum\sum_{i\neq j} w_{ij}\left(\hat{\theta}_w\right)\check{\eta}_{ij}\left(\hat{\theta}_w\right)V_j I\left\{\min\left(X_i,t_0\right)\geq X_j\right\}\left\{\hat{G}\left(X_j\right)\right\}^{-2} I\left(X_j\geq u\right)$$

$$\hat{M}_{Z_i}(u) \Rightarrow I\left(X_i\leq t, V_i=0\right)-\int_0^t I\left(X_i\geq u\right)d\hat{H}\left(u\right)$$

respectively, where $\hat{H}(u)$ is the Nelson estimate of the common cumulative hazard function for the censoring variable.

6.6 Choosing the correct decreasing function g(.)

When it is not clear that either the proportional odds or the proportional hazards model fit the dataset well, a simple graphical method to select an appropriate model is suggested by Cheng et al. (1997). In particular, the following class of transformations g(.) indexed by λ for models (6.5) & (6.12) is considered:

$$g(s)=\left.\begin{array}{ll}\log\left(\lambda^{-1}\left(s^{-\lambda}-1\right)\right) & \lambda>0\\ \log\left(-\log\left(s\right)\right) & \lambda=0\end{array}\right\} \tag{6.21}$$

When $\lambda=0$ then the new model is reduced to the proportional hazards one, while if $\lambda=1$ then the new model is reduced to the proportional odds one. We assume, that an unknown function of the survival time T is truly linearly related to the covariates, and the error distribution depends on the parameter λ. Using a realization of λ, we estimate the parameters θ_0 and the hazards $h_0(t)$. If the choice of λ is appropriate, then the distribution of $\left\{\hat{h}\left(X_i\right)+Z_i^T\hat{\beta}\right\}$ or $\left\{\hat{h}\left(X_i\right)-Z_i^T\hat{\beta}\right\}$ respectively to the initial model, would be very close to the error distribution. Hence, a P-P plot based on the fitted error distribution and the Kaplan-Meier estimate constructed from $\left\{\left(\hat{h}\left(X_i\right)+Z_i^T\hat{\beta},V_i\right),i=1,\ldots,n\right\}$ or $\left\{\left(\hat{h}\left(X_i\right)-Z_i^T\hat{\beta},V_i\right),i=1,\ldots,n\right\}$ respectively, would be approximately a straight line through the origin.

6.7 Numerical Example

We now apply the leukemia data, presented in Table (4.1) to the two-sample proportional hazards model (5.2). Even though, censoring in the **placebo** group is little, in the chemotherapy **treatment** group censoring is quite evident in the range of 36.36%. Particularly, only one patient from those who did not receive any treatment (strata 2), was observed not to experience the event as shown in the Fig (6-1). His tracks were lost in the 16 week from his entrance to the study. Consequently, we can say that this group is 8.33% censored, in the sense that 1 out of 11 patients, may not experience the event until the end of the study or not at all; see also Table 6.1. Likewise, in the Fig. (6-1), the four weeks time points represents the weeks period (13, 28, 45 and 161 weeks respectively), after which the traces of the four patients, that belong to the treatment group (strata 1), were lost. Hence, the maintenance group is said to be 36.36% censored in the sense that a AML patient received the same kind of treatment under the same circumstances, is subjected to censorship with this probability 0.364. In Table 6.1 there is a summary of the number of censored and uncensored patients included in the current study. The total percentage of censoring is somewhere in between the percentage of censoring of the two groups, and refers exclusively to the total censored ratio of the patients who were censored versus those who participate in the study.

Group	Total	Failed	Censored	%Censored
0	11	7	4	36.3636
1	12	11	1	8.3333
Total	23	18	5	21.7391

Table 6.1: Summary of the Number of censored patients

Moreover, we notice on the graphs (6-2) and (6-3) that one can obtain several descriptive statistics for each group. Firstly, the median is obtained by proceeding horizontally from the 0.5 point (marked by an arrow) on the vertical axis each time until the survivor curve is reached, and then proceeding vertically downward until the horizontal axis

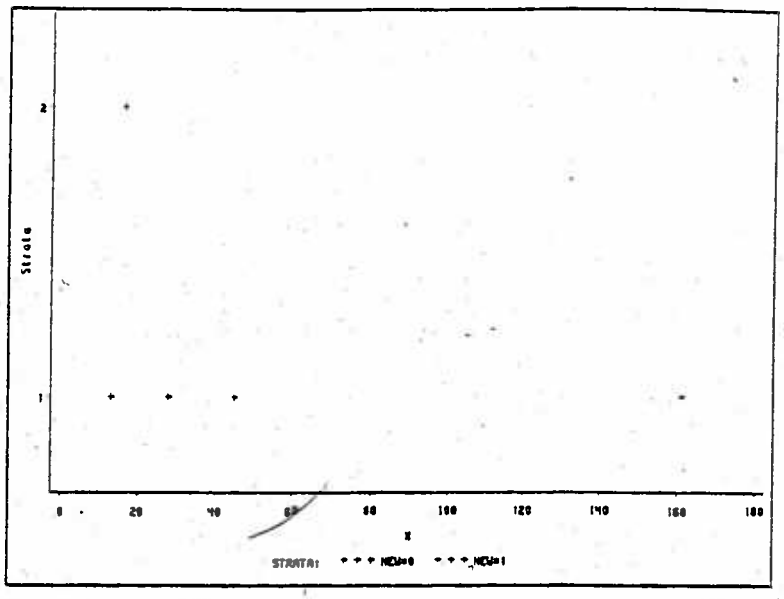

Figure 6-1: The censored times for the different groups. Strata 1: Treatment Group Strata 2: Placebo Group

is crossed at the median survival time. In other words we choose as median the week time t_0 which gives probability to survive equals to 0.5. Since, the estimated survivors are just point estimates and not an explicit continuous function, we choose the closest to 0.5 probability value and keep the corresponding time point such that $\bar{F}(t_0) \leq 0.5$. Secondly, the 25% and 75% quartiles, marked in the two graphs Fig (6-2) & (6-3) are calculated in a similar way $\bar{F}(t_{0.25}) \leq 0.25$ and $\bar{F}(t_{0.75}) \leq 0.75$. Moreover, the confidence interval for each survivor (which are denoted by stars on the step functions) are given by $\bar{F}(t) \pm 1.96se\left(\bar{F}(t)\right)$ and are plotted in the pre-mentioned figures. The fitted standard errors are calculated using formula (4.13), and can be found in Tables (4.2) & (4.3). The censored patients' survivors are also reported on the graphs

For the treatment group the median is 31 weeks; for the placebo group, the median is 23 weeks. Comparison of the two medians reinforces the survivor curves inspection that the treatment has an overall effect on patients, but how significant is this effect. This difference enlarges when the two means 31.843 for the treatment group against 22.7 for

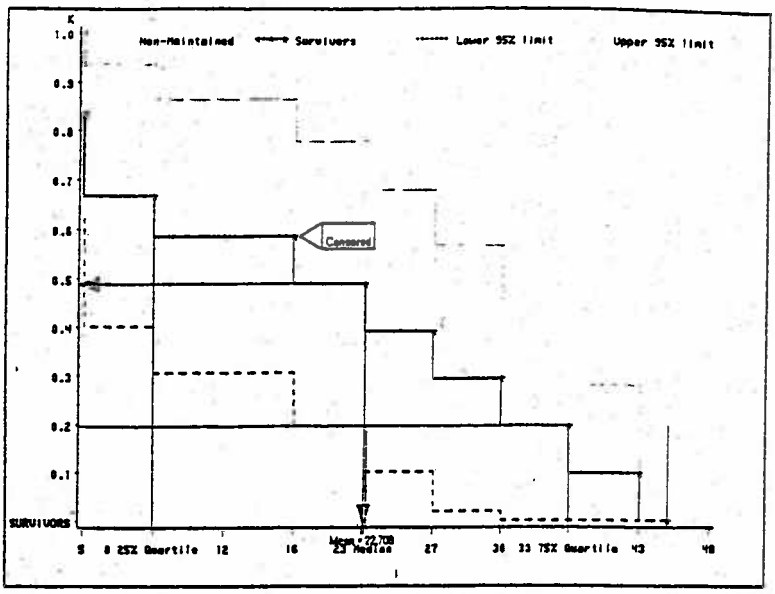

Figure 6-2: Confidents intervals for the K-M survivors and the quartile survival times for the placebo group

the placebo one, are to be related.

The group indicator is the only covariate in the model; that is, $Z = 0$ if the observation is from the first or placebo group and $Z = 1$ otherwise. The maximum partial likelihood estimate for the parameter β_0 manipulated by solving the equation $U_r(\beta) = 0$ is equal to 0.904 and the corresponding standard error does not exceed the value 0.51. The score function $U_r(\beta)$ is given in (5.14). Also, analyzing the p-value of 0.0775 from the Wald statistic given in (5.16) which tests the null hypothesis $\beta = 0$ we can conclude that in 95% significant level, the coefficient of the treatment variable is not significant and the hazards of the two groups should not be considered "separate enough" so as, the chemotherapy treatment to effect in point. Nevertheless, we can speak for an effect of the treatment in 92% significant level.

In addition, the risk ratio given in Table (6.1) states that a patient who received no treatment has an estimated 2.47 times higher risk of leaving the remission period, than the one that did receive chemotherapy. In math terms, the Risk ratio or else hazard ratio

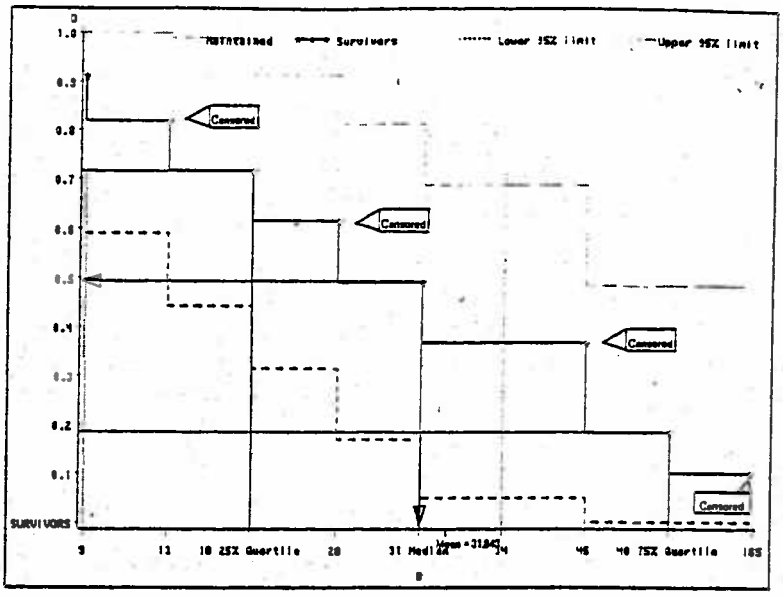

Figure 6-3: Confidents intervals for the K-M survivors and the quartile survival times for the treatment group

can be expressed as

$$HR = \frac{h(t, z^*)}{h(t, z)} \tag{6.22}$$

When the above quantity is constant over time, then the hazard for one individual or a cluster of objects with the specification Z^* of the explanatory variable (which here is the placebo group) is proportional to the hazard for any other individual or cluster of objects with the specification Z respectively (which here is the treatment group). This is also known as the PH (proportional hazard) assumption (see Kleinbaum (1997)). Whether or not the PH assumption is met, we can decide the suitability of the proportional hazards model (5.2) in the data. Excluding the case where the explanatory variables are **time-dependent** which is not the case here, we will make use of three general approaches for assessing the PH assumption in the current data. All of them are popular graphical techniques available in the survival analysis and are discussed also in the chapter 5. In particular, we compare the $\log(-\log)$ **survivor curves** over the explanatory

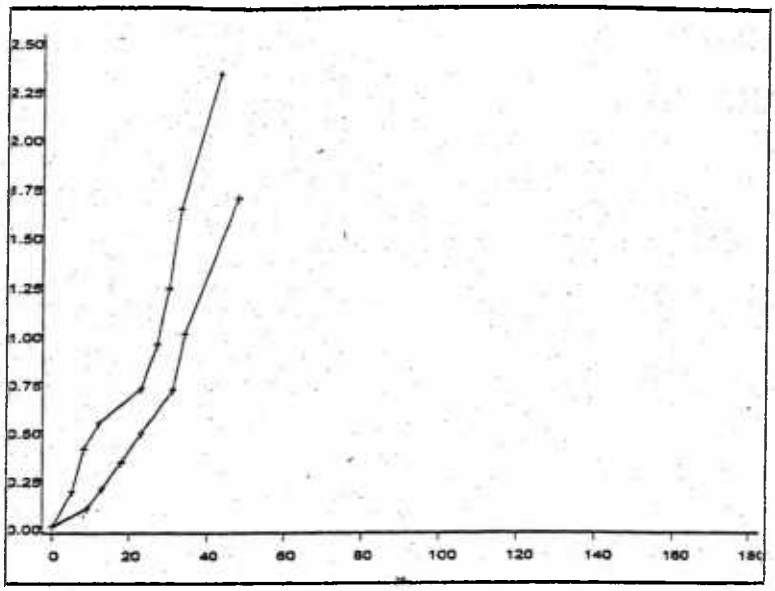

Figure 6-4: $-log\left(\bar{F}\left(t\right)\right) = \hat{H}\left(t\right)$: The estimated cumulative hazard function

variables being investigated.. The log $(-\log)$ survival curve is simply a transformation of an estimated survival curve that results from applying the natural logarithm on the survivors twice. The last transformation appeared to map the $[0, 1]$ interval to the real numbers line. The validity of the PH assumption can be checked by evaluating whether or not log-log curves for the two group of patients are parallel. To show this we correspond the formula for the survival curve to the hazard function for the PH model (5.2).

$$h\left(t, Z\right) = h_0\left(t\right) e^{\beta Z^T} \Longrightarrow \bar{F}\left(t, Z\right) = \left[\bar{F}_0\left(t\right)\right]^{e^{\beta Z^T}}$$

Recalling the mathematical link between any hazard function and its corresponding survival function, see paragraph 2.3, we obtain the survival curve for the PH model. The term $\bar{F}_0\left(t\right)$ denotes the baseline function that corresponds to the baseline hazards function $h_0\left(t\right)$. Applying the log$(-\log)$transformation in the last expression we have that

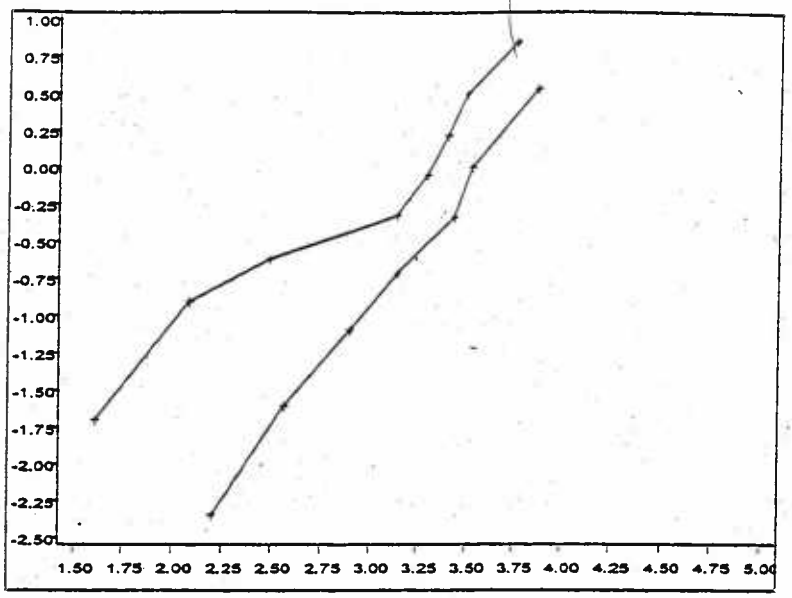

Figure 6-5: $log\left(-log\left(\bar{F}(t)\right)\right) = \log\left(\hat{H}(t)\right)$: The logarithm of the estimated cumulative hazard function

$$\log\left(-\log\left(\bar{F}(t,Z)\right)\right) = \beta Z^T + \log\left[-\log \bar{F}_0(t)\right]$$

The right hand side of the formula, consists of the sum of the expression βZ^T which is actual the sum $\sum_{i=1}^{p} \beta_i Z_i$ where p is the number of explanatory variables (here $p = 1$), with the log of the negative log of the baseline survival function which takes values on the real number line. Now considering our example situation, where Z has two values Z_1 and Z_2 for the two group of patients. Then the corresponding log-log curves for these groups are given by

$$\begin{aligned}
\log\left(-\log\left(\bar{F}(t,Z_1)\right)\right) &= \beta Z_1 + \log\left[-\log \bar{F}_0(t)\right] \\
\log\left(-\log\left(\bar{F}(t,Z_2)\right)\right) &= \beta Z_2 + \log\left[-\log \bar{F}_0(t)\right]
\end{aligned}$$

While their difference is

$$\log\left(-\log\left(\bar{F}\left(t, Z_1\right)\right)\right) - \log\left(-\log\left(\bar{F}\left(t, Z_2\right)\right)\right) = \beta\left(Z_1 - Z_2\right)$$

The above formula says that if we use a Cox PH model and we plot the estimated log-log survival curves for the different group of patients on the same graph, the two plots would be approximately parallel, since $\beta\left(Z_1 - Z_2\right)$ involves only the differences in predictor values and does not depend on time. In Fig. (6-5) the empirical plots of log-log survival curves, based on the Kaplan-Meier estimates, are presented. As shown in the graph the two curves for the placebo group ($new = 1$) and the treatment group ($new = 0$) can be considered to have approximately a constant difference.. Even though this assumption can be thought quite subjective for this small data set, we adapt the conservative strategy that the PH assumption is satisfied unless there is strong evidence of nonparalelism of the log-log curves. The cumulative hazard curves for the two group of patients are also schematized in Fig. (6-4) and their difference is considered steady as they do not cross, apart from the zero time point, in which there is no hazard for any patient.

A second check of the PH assumption is made by comparing the observed with expected survival curves. To obtain expected plots, we fit a Cox PH model containing the predictor being assessed, whereas as observed survival curves the Kaplan-Meier curves are utilized. Observed and expected curves in Figures (6-6) & (6-7) appear to be quite close for each group. Thus we would conclude using this graphical approach that the predictor variable satisfies the PH assumption and therefore the Cox procedure for these kind of data should be used. Similarly to the previous approach, the last conclusion was made considering the conservative strategy which suggests that the PH assumption is not met only when observed and expected plots are strongly discrepant.

Another examination of the PH assumption is made through the graph of the rescaled Schoenfeld residuals (given below) versus the treatment variable (Fig (6-8)). The existence of a significant slope in the scatter plot would be evidence of the unsuitability of

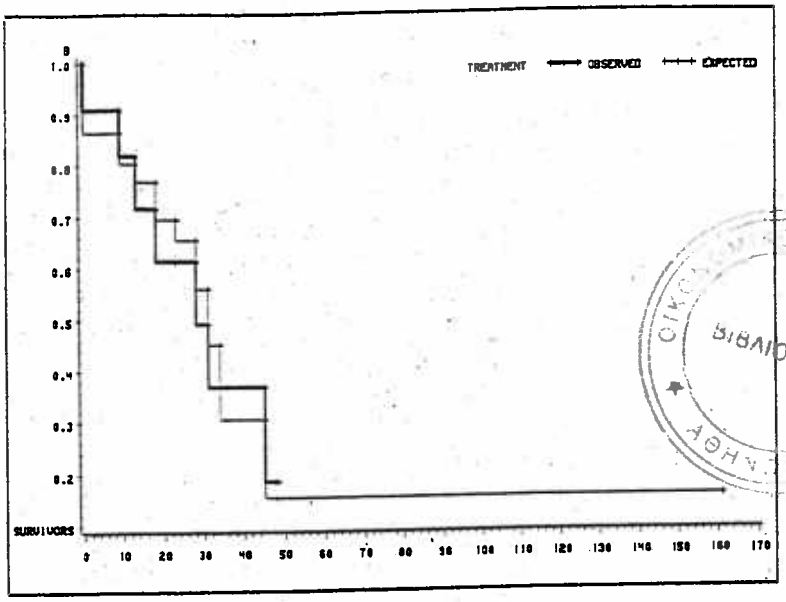

Figure 6-6: Observed and Expected Survival curves for the treatment group

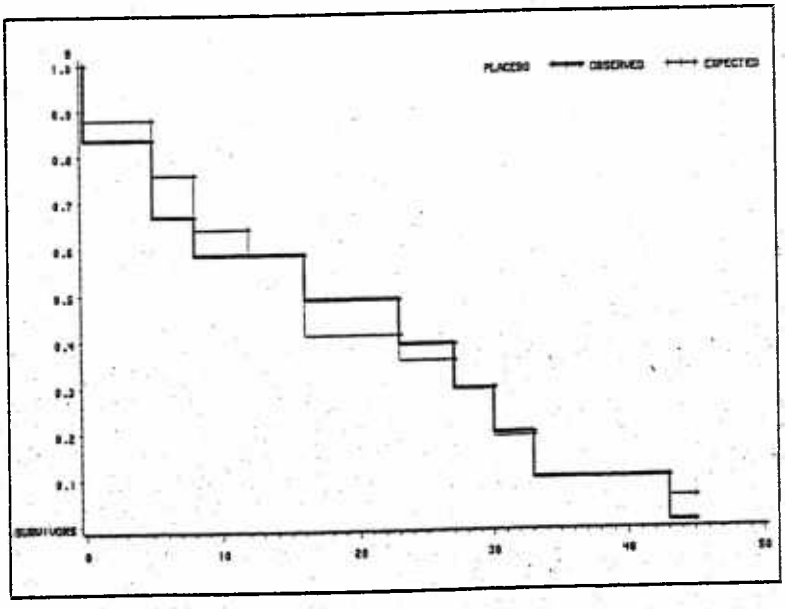

Figure 6-7: Observed and Expected Survival curves for the placebo group

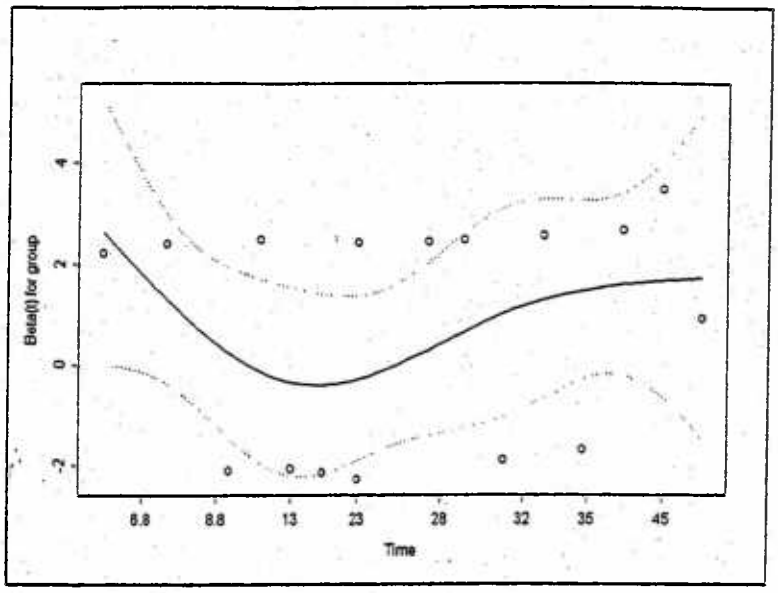

Figure 6-8: Schoenfeld residuals used to check the PH assumption

the PH assumption which is not the case here, strictly speaking.

$$s_{ij}(\beta) = Z_{ij}(t_i) - \bar{Z}_j(\beta, t_i)$$

Where i and t_i are the subject and the time the event occured, whereas j refers to the variable, $j = 1 \ldots p$ and p is the total numer of the variables which here is 1.

- Analyzing the residuals of the Cox PH method, we see a reasonably linear relationship of the group variable versus the martingale residuals (given below as a counting process notation), see Fig. (6-9). Therefore, no special functional form of the group or else treatment variable should be utilized in our model.

$$M_i(t) = N_i(t) - \int_0^t r_i(\beta, s)\, Y_i(s)\, d\hat{H}_0(\beta, s)$$

Where $N_i(t)$ is a counting process for the i patient, which increases by 1 at the observed event, $Y_i(t)$ is an indicator function which is 1 if person i is still at risk,

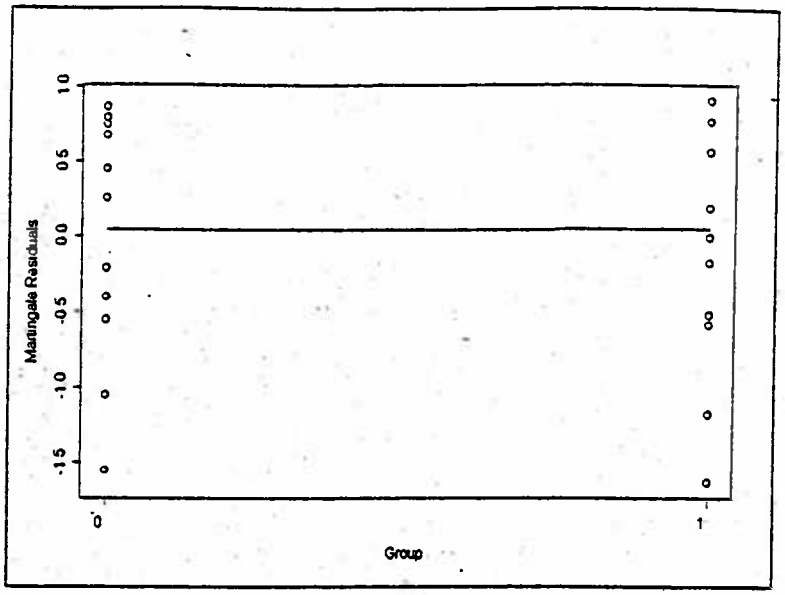

Figure 6-9: Martingale residuals for discovering the predictor form

$r_i(\beta, t)$ is the risk function and $\hat{H}_0(\beta, s)$ is the baseline hazard estimation.

- Moreover, From the deviance residuals (given below) scatter plot, we notice no widely deviant observation. Even though there one can identify some doubt of configuration of the residuals which is excused due to the scarcity of data Fig (6-10).

$$d_i = sign\left(\hat{M}_i\right)\sqrt{-\hat{M}_i - N_i \log\left(\frac{\left(N_i - \hat{M}_i\right)}{N_i}\right)}$$

Where $\text{sign}\left(\hat{M}_i\right)$ is the sign of the martingale residuals, taking values $-1, 0, 1$ when M_i is negative, zero and positive respectively.

- Furthermore, the largest changes in the regression coefficient are 0.3 (Fig. 6-11), which are reasonable, and give no suspicion for influential points. In other words there is no point that seems to influent the regression estimates. Particularly, the

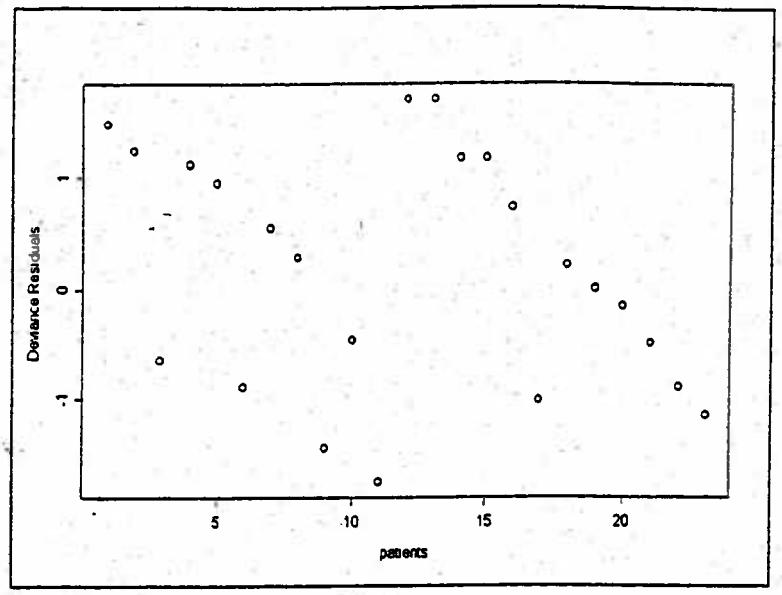

Figure 6-10: Deviance residuals for identifying poorly predicted subjects

change that would occur in coefficient β if observation i were dropped from the model is $-Il_i$, where I^{-1} is the Cox model variance matrix, as defined in Chapter 5. These changes in β are plotted in Fig. (6-10), where there is no distinct great change in the coefficient.

Finally the bounds of a 95% confidence interval of the risk ratio are calculated. The formula used is

$$e^{\hat{\beta}_0 \pm 1.96(S.error)}$$

D.F.	Parameter Estimate	S. Error	Risk Ratio	Lower	Upper	Lik. Ratio	P-value
1	0.904	0.51	2.47	0.90	6.74	3.296	0.069

Table 6.2: Maximum Likelihood Estimates

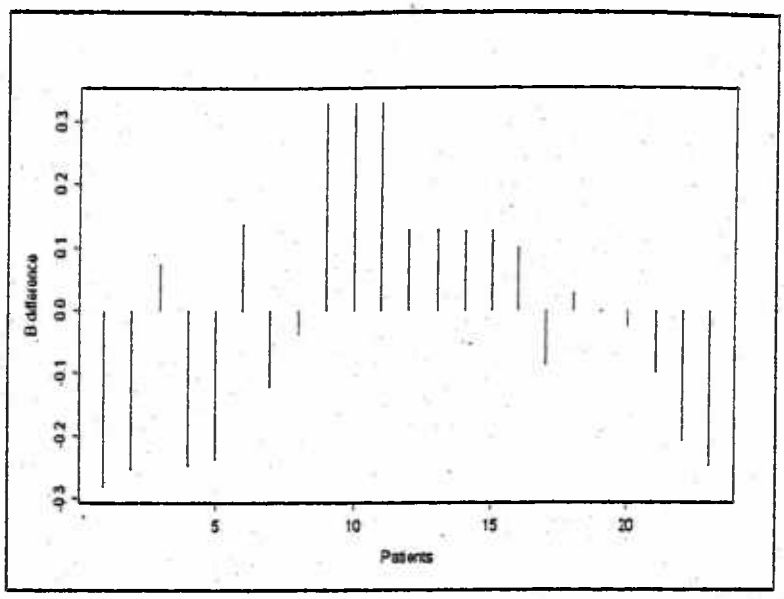

Figure 6-11: Identifying influential points

With the estimating function of $U^*(\beta)$ as defined in (6.11), the estimate for the parameter β_0 is 0.93, with an estimated standard error of 0.42, as shown in Table (6.3), below.

D.F.	Parameter Estimate	S. Error	Risk Ratio	Lower	Upper
1	0.93	0.42	2.53	1.11	5.76

Table 6.3: Cheng's estimates

Also following the steps of obtaining the score functions on (6.17) and (6.18), we can estimate the survival probabilities of each group Fig-(6-12). The individuals who take the treatment have greater survival probabilities from those who do not.

D.F.	Parameter Estimate	S. Error	Risk Ratio	Lower	Upper
1	0.88	0.45	2.429	0.99	5.82

Table 6.4: Modified estimates

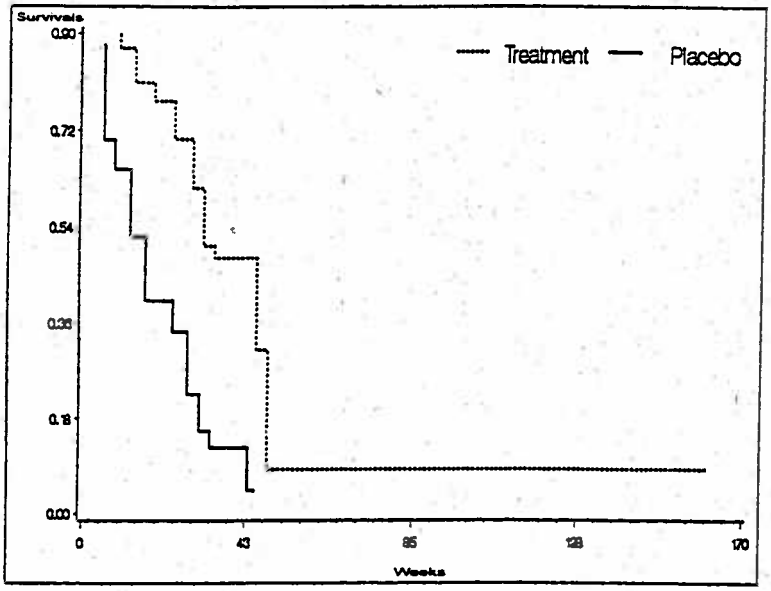

Figure 6-12: Predicted Survivals Probabilities

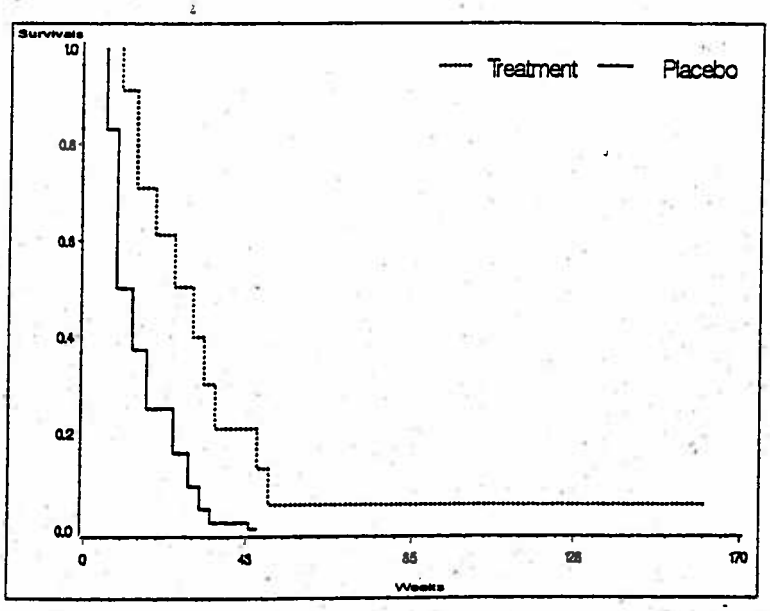

Figure 6-13: Corrected Predicted Survivals Probabilities

Applying, thereupon, the corrected estimators, found in the paragraph 6.4, the results included in Table (6.4) are obtained. The parameter estimate $\hat{\beta}_w$ is again close to the Cox's proportional estimate, like the one derived from the score function in (6.11). The choice of the truncation point t_0, which is assumed to be data-independent, is not deterministic. On the contrary, a quite stable estimator $\hat{\beta}_w$ is obtained when t_0 is chosen adoptively to the current data of the study. Regarding the present leukemia data, a suitable time t_0 is such that about 15% of the observed failure times fall beyond it. The derived survivors for the last method are in Fig (6-13).

The two-sample proportional hazards model fits the data well, as shown above. Thus, inference about the new models can be based on their comparison with the Cox's procedure. Initially, we can say that following the Cheng's et al. (1995) procedure, the null hypothesis ($H_0 : \beta = 0$) is rejected in 95 significant level, which differ from the results so far. Nevertheless, assuming a more strict significance level as 99%, the same hypothesis is not rejected. Hence, the statistical inference of the Cheng's et al. (1995) procedure cannot said to diverge from that of Cox's procedure. According to Fine et al. (1998), with heavy censoring, the coverage probabilities for the interval estimation procedure based on (6.10) may be substantially smaller than the nominal levels. Here, the censoring is not heavy enough to observe clearly this bias.

Following Cheng et al. (1997), the graph of the fitted error distribution against the empirical probability of the quantity $\left\{\hat{h}(T_i) + Z_i^T\hat{\beta}\right\}$ is given in Figures (6-14), (6-15), (6-16). Although, the probabilities do not coincide on the diagonal line, there is no strong evidence of the propriety of the proportional hazards model, due to the small number of observations. In addition, the points in the cases of $\lambda = 0.5$ and $\lambda = 1$ are not far from the diagonal line as shown in the Figures (6-15) and (6-16) respectively. The choice of λ to be either 0, 0.5 or 1 seem all appropriate for the leukemia data analysis.

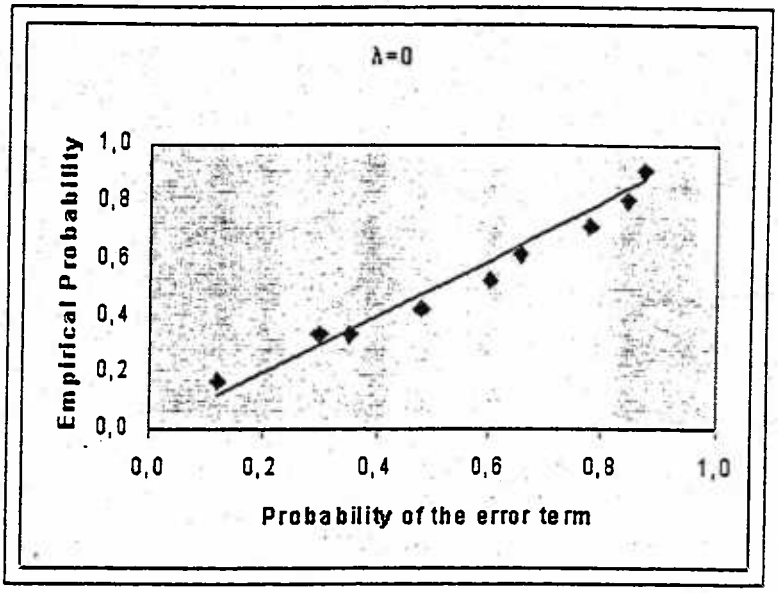

Figure 6-14: P-P Plot for $\lambda = 0$, and the proportional hazards model

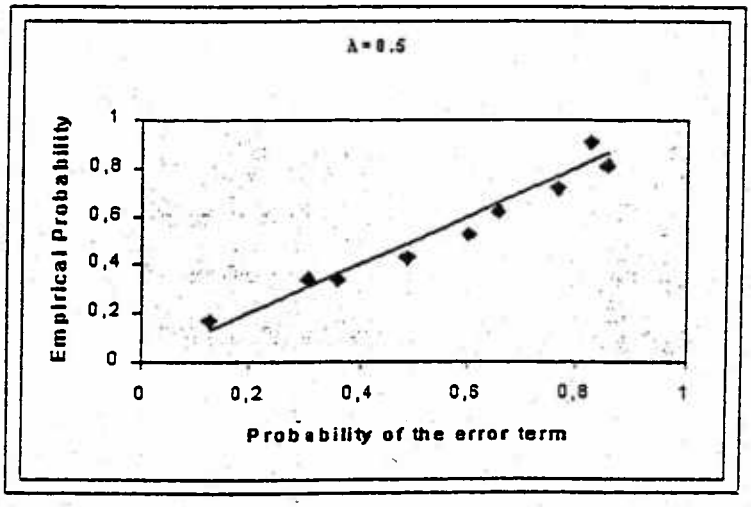

Figure 6-15: P-P Plot for $\lambda = 0.5$

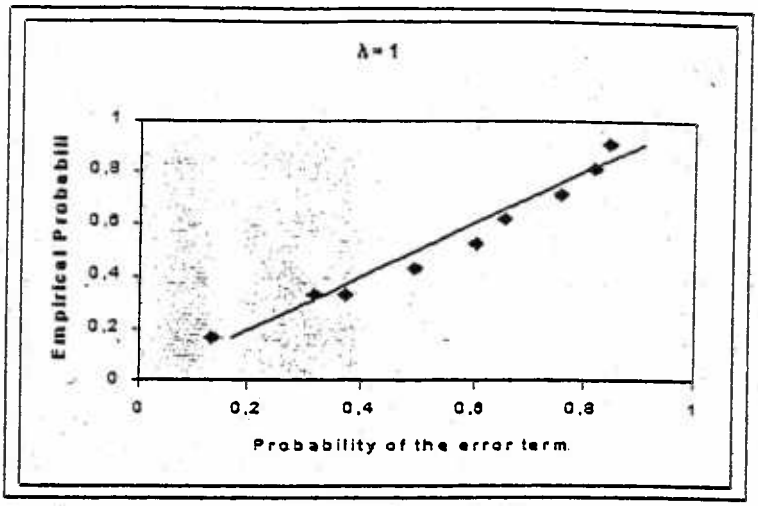

Figure 6-16: P-P Plot for $\lambda = 1$, and the proportional odds model

Parameter Estimate	S. Error
1.2897054	0.6958963

Table 6.5: Standard logistic case

Models	-2log L	Wald	P-value	Efron R^2
Proportional hazard	82.500	3.12	0.078	
Additive	420.063	3.38	0.066	
Generalized hazard	328.295	4.97	0.026	0.0902
Adjusted Generalized hazard	121.605	3.76	0.052	0.0044
Generalized Odds	328.172	3.43	0.064	0.0562

Table 6.6: Model diagnostic tests

$$R^2 = 1 - \frac{\sum_i \sum_j \left(y_{ij} - \hat{\eta}_{ij}\right)^2}{\sum_i \sum_j \left(y_{ij} - \bar{y}\right)^2} \tag{6.23}$$

Moreover, in the Table 6.6 The Efron goodness of fit test (6.23) is included for the three new models. In addition the minus log-likelihood (-2logL) is included for each model The Cox's procedure seems to have a better fit to the current example (smallest -2logLikelihood), fact which states the suitability of the proportional hazard model to the data. In contrast, the additive model has a great -2logL value, which does not suggest so well convergence to the real values of the estimates. All the models' analysis agrees (with the small variation of the Generalized hazard model) that the null hypothesis that the parameter of the grouping variable is not significant, is not rejected. Fact that supports the unreliability of the chemotherapy treatment Under the generalized proportional hazards model (as here we used the extreme value cumulative function $F(s) = 1 - e^{-e^s}$), both the new proposed generalized procedures are not expected to be as efficient as the Cox procedure; however, in the presence of moderate censoring they perform fairly well.

In the current example, where the censoring level is only 21.74% (5 out of 23 patients have been censored), both the plain and even more the modified generalized procedures by Cheng et al (1995) and Fine et al (1998), provide very close estimates to those of the Cox procedure. Finally, the Efron R square which is defined in (6.22) is in favor of the modified model, which seems preferable in our example. The choice of the standard logistic distribution, which leads to the proportional odds model Table 6.5, can be a sound alternative as its Efron statistic is less than the one of the Generalized hazard model. However its log-likelihood value (328.172), leads us, not to choose it for the current dataset.

Chapter 7

Conclusions and Suggestions for further Research

The current report presents standard and new Survival Analysis techniques and their applicability to already studied data in the literature.

The first part is a brief historical report of the survival analysis. Afterwards the concepts of **failure time, left-truncated data, competing risk** problem, **censorship, hazard rate** and **survivor function** are introduced. The latter two functions are calculated for the most common used survival random variables in the third part. Moreover, fourth part includes the **Kaplan-Meier** estimator for the survival function, the **Accelerated Life Model**, the **Proportional Odds Model** and the **Additive risk Model**, followed by a two-sample application in AML medical data.

In the fifth part, statistical inference of the Cox's proportional hazard model is developed in the continuous and the discrete case, whereas, a recent generalized survival model and its correction are found in the last part of this report. The results obtained by the new suggested model are quite satisfactory in the sense that they are close to the ones obtained using the standard survival models. The fact that the Cheng's generalized survival model is an augmentation of the rest known survival models by altering the distribution of the error component, makes it a fruitful and practicable choice for the

statistician who can trace in advance a suitable model using the same algorithmic steps.

All of the models used in this report, give adequate estimates for the parameters and fit the data satisfactory. Having the results of the Cox's proportional model, which gives efficient results, no matter the nature of the response variable, the adjusted generalized model seems to be closer in estimates. Moreover, detecting that the model checking tests (Table 6.6) are in favour of the latter model in relevance to the simple Cheng's generalized model, we can conclude that Fine et al (1998) procedure can be used equally with the Cox's proportional hazard procedure and provide significance statistical inference.

An area of further research will be the use of the new generalized models in cases of categorical explanatory factors, where the individuals can be grouped along the strata defined by the number of possible factor combinations. Also, goodness of fit tests, applied in generalized binary data like Efron statistic (6.23), can be utilized not only for checking the adequacy of each explanatory factor, but for model checking between the different choices of the distribution of the error term. Finally, these generalized procedures seem also adequate for the analysis of non-medical data like in Regional and Social issues area.

APPENDIX

A. CODE IN SAS SOFTWARE FOR THE EVALUATION OF THE UNKNOWN PARAMETER β AND ITS VARIANCE USING THE SCORE FUNCTION GIVEN IN (6.11). ALSO THE SURVIVOR ARE ESTIMATED BY (6.18)

```
proc iml;
reset autoname;

/* Put the data into matrix R*/
use sasuser.leukem2;
read all into R;

/* Put the K-M estimates for
the censor variable into R2 */
use sasuser.marios2;
read all into R2;

/*Put the probabilities used for the calculation of the fitted survivals*/
use sasuser.mar;
read all into T2;

/*Put the distinct times */
use sasuser.mar2;
read all into coll;

bmin=-3;
bmax=2;
b=(bmin+bmax)/2;
usum=1;

/*Algorythm stops when the score function is near zero*/
do while(abs(usum)>=1E-5);
usum=0;
    do i=1 to 23;
        do j=1 to 23;
        tie=1;
        if i=j then tie=0;
        Z2=R[i,3]-R[j,3];
        Z3=Z2*b;
        a1=R[i,1];
        a2=R[j,1];
        ll=0;
        if a1>=a2 then ll=1;
```

```
        Z2=R[i,3]-R[j,3];
        lh=Z2*Z2*(-1)*exp(Z2*b)/((exp(Z2*b)+1)**2)+lh;
        end;
end;

lh=(1/529)*lh;


gen=0;
do l=1 to 23;
        one=1-R[l,2];
        ppsum=0;
        do k=1 to 23;
                if (R[k,1]>=R[l,1] & R[k,3]=R[l,3]) then pp=1; else pp=0;
                ppsum=ppsum+pp;
        end;
        part2=0;
        do i=1 to 23;
                do j=1 to 23;
                        if R[i,1]>=R[j,1] then qq=1; else qq=0;
                        if R[j,1]>=R[l,1] then qq2=1; else qq2=0;
                        Z4=R[i,3]-R[j,3];
                        if i>12 then k=2; else k=1;
                        part2=(((-1)*Z4*R[j,2]*qq*qq2)/(R2[j,k]*R2[j,3]))+part2;

                end;
        end;

        gen=(one*(part2**2))*(ppsum**-2)+gen;
end;

gen=(4/12167)*gen;

serror=(gamsum-gen)/(lh*lh);

/*The st. error of b*/
serror=sqrt(serror/23);

/*The Wald statistic*/
W=(b/serror)**2;

/*The p-value of the statistic*/
pvalue=1-cdf("chisquared",W,1);

/*End of the SAS procedure IML*/
quit;
```

- Manipulation of the Likelihood and the Efron R^2

```
proc iml;
reset autoname;
use sasuser.leukem2;
read all into R;

use sasuser.marios2;
read all into R2;

/*The estimated value of the unknown parameter is*/
b=0.93126;

/*The mean of the y_ij*/
ymean=0.50;

/*The log-likelihood accumulator */
lik=0;
rsq=0;
rsq2=0;


      do i=1 to 23;
            do j=1 to 23;
            tie=1;
            if i=j then tie=0;
            weight=-1;
            Z2=R[i,3]-R[j,3];
            Z3=Z2*b;
            a1=R[i,1];
            a2=R[j,1];
            ll=0;
            if a1>=a2 then ll=1;
            intr=1/(exp(Z3)+1);

/*The third column of R2 contains G_j(T_j), whereas the first is equal to
G_1group(T_j) and the second is equal to G_2group(T_j)*/
            if i>12 then k=2; else k=1;
            quan=((R[j,2]*ll)*tie/(R2[j,k]*R2[j,3]));
            lik=lik+(quan*log(intr)+(1-quan)*log(1-intr));
            rsq=rsq+((quan-intr)*(quan-intr));
            rsq2=rsq2+((quan-ymean)*(quan-ymean));
          end;
      end;

/*Efron R square*/
rsq=1-rsq/rsq2;
print rsq,lik;
```

```
quit;
```

B. CODE IN SAS SOFTWARE FOR THE EVALUATION OF THE UNKNOWN PARAMETER β AND ITS VARIANCE USING THE SCORE FUNCTION GIVEN IN (6.13). ALSO THE SURVIVOR ARE ESTIMATED BY (6.17)

```
proc iml;
reset autoname;
use sasuser.leukem2;

/* Put the data into matrix R*/
read all into R;

/* Put the K-M estimates for
the censor variable into R2 */
use sasuser.marios2;
read all into R2;

/*Put the probabilities used for the calculation of the fitted survivals*/
use sasuser.mar;
read all into T2;

/*Put the distinct times into matrix named coll*/
use sasuser.mar2;
read all into coll;

R[,2]=ABS(1-R[,2]);

/*Choice of the unknown time parameter*/
num=17;
tsur=coll[num,2];

/*By adjusting the values of the two parameters we are lead to the root of the
derivative of the sum of squares*/
bmin=-1;
bmax=3;
b=(bmin+bmax)/2;
usum=1;
usum1=1;
do while(abs(usum)>=1E-6);

amin=-2.7;
amax=-2;
a=(amin+amax)/2;
usum1=1;
do while(abs(usum1)>=1E-7);

/*Usum is the partial derivative with respect to the parameter b*/
```

```
usum=0;

/*Usum2 is the sum of squares*/
usum2=0;

/*Usum is the partial derivative with respect to the parameter a*/
usum1=0;
      do i=1 to 23;
            do j=1 to 23;
            tie=1;
            if i=j then tie=0;
            Z2=R[j,3]-R[i,3];
            Z3=Z2*b;

            a1=R[i,1];
            a2=R[j,1];
            a3=min(a1,tsur);
            ll=0;
            if a3>=a2 then ll=1;

/*Using unkown parameter a into the algorithm*/
            intr=(1-exp((-1)*(exp(alpha))*(exp(Z3)+1)))/(exp(Z3)+1);
            intr3=(exp(alpha-(exp(alpha)*(exp(Z3)+1))));
            intr3=(intr-intr3)/(exp(Z3)+1);
            if i>12 then k=2; else k=1;
            quan=((R[j,2]*ll)/(R2[j,k]*R2[j,3]));
            quan=quan-intr;
            ksum=intr-intr3;
            ksum2=exp(Z3)*intr3;
            new=(-1)*(R[j,3]*ksum)+(R[i,3]*ksum2);
            weight=ksum-ksum2;
            usum=usum+new*weight*quan*tie;
            usum1=usum1+new2*quan*tie;
            usum2=usum2+quan*quan*tie;

          end;
      end;
if usum1<0 then amax=a; else if usum1>0 then amin=a;
a=(amin+amax)/2;
end;
if usum<0 then bmax=b; else if usum>0 then bmin=b;
b=(bmin+bmax)/2;
end;

/*SURVIVAL ESTIMATES*/
/*Calculation of the h function estimates which here are denoted as hsim */
hsim=J(20,1);
```

```
Ssim=J(20,1);

do j=1 to 20;
t=coll[j,2];
hmin=-30;
hmax=31;
h=(hmin+hmax)/2;
usum=1;
do while(abs(usum)>=1E-5);
usum=0;
      do i=1 to 23;
      if R[i,1]>=t then quan=1; else quan=0;
      k=1;
      if i>12 then k=2;
      first=quan/T2[j,k];
      second=exp(-exp(h-b*R[i,3]));
      usum=usum + (first-second);
      end;
heta=h;
if usum>0 then hmax=h; else if usum<0 then hmin=h;
h=(hmin+hmax)/2;
end;
h=heta;
hsim[j]=h;

/*Calculation of the survival estimates*/
ssim[j]=exp(-exp(h-b*coll[j,3]));
end;

num=num-1;

/*Using the parameter estimates*/
alpha=-2.130802;
b=0.88;

/*CALCULATION OF THE STANDARD ERROR OF THE b ESTIMATOR*/
/*The e bivariate  function */
e=J(23,23);
do i=1 to 23;
      do j=1 to 23;
      if j=i then sw=0; else sw=1;
      Z2=R[j,3]-R[i,3];
      Z3=Z2*b;
      a1=R[i,1];
      a2=R[j,1];
      a3=min(a1,tsur);
      if a3>=a2 then ll=1; else ll=0;
```

```
        intr=(1-exp((-1)*(exp(alpha))*(exp(Z3)+1)))/(exp(Z3)+1);
        intr3=(exp(alpha-(exp(alpha)*(exp(Z3)+1))));
        intr3=(intr-intr3)/(exp(Z3)+1);
        if i>12 then k=2; else k=1;
        quan=((R[j,2]*ll)/(R2[j,k]*R2[j,3]));
        quan=quan-intr;
        ksum=intr-intr3;
        ksum2=exp(Z3)*intr3;
        new=(-1)*(R[j,3]*ksum)+(R[i,3]*ksum2);
        weight=(1/(intr*(1-intr)));
        alpha2=new*quan;
        e[i,j]=alpha2;
        end;
end;

gamsum=0;
do i=1 to 23;
      do j=1 to 23;
            Z2=R[i,3]-R[j,3];
            do k=1 to 23;
                  if (k=j | i=j | i=k) then contr=0; else contr=1;
                  Z3=R[i,3]-R[k,3];
                  gam1=(e[i,j]+e[j,i]);
                  gam2=(e[i,k]+e[k,i]);
                  gamsum=gamsum+gam1*gam2*contr;
            end;
      end;
end;

gamsum=(23**-3)*gamsum;

lh=0;
do i=1 to 23;
      do j=1 to 23;
      sw=1;
      if i=j then sw=0;
      Z2=R[j,3]-R[i,3];
      Z3=Z2*b;
      a1=R[i,1];
      a2=R[j,1];
      a3=min(a1,tsur);
      if a3>=a2 then ll=1; else ll=0;
      intr=(1-exp((-1)*(exp(alpha))*(exp(Z3)+1)))/(exp(Z3)+1);
      intr3=(exp(alpha-(exp(alpha)*(exp(Z3)+1))));
      intr3=(intr-intr3)/(exp(Z3)+1);
      if i>12 then k=2; else k=1;
      quan=((R[j,2]*ll)/(R2[j,k]*R2[j,3]));
```

```
      quan=quan-intr;
      ksum=intr-intr3;
      ksum2=exp(Z3)*intr3;
      new=(-1)*(R[j,3]*ksum)+(R[i,3]*ksum2);
      weight=(1/(intr*(1-intr)));
      lh=sw*new*new+lh;
      end;
end;

lh=(lh*(23**-2));

gen=0;
do l=1 to 23;
      one=1-R[l,2];
      ppsum=0;
      do k=1 to 23;
            if (R[k,1]>=R[l,1] & R[k,3]=R[l,3]) then pp=1; else pp=0;
            ppsum=ppsum+pp;
      end;
      part2=0;
      do i=1 to 23;
            do j=1 to 23;
                  if i=j then sw=0; else sw=1;
                  Z2=R[j,3]-R[i,3];
                  Z3=Z2*b;
                  a1=R[i,1];
                  a2=R[j,1];
                  a3=min(a1,tsur);
                  if a3>=a2 then ll=1; else ll=0;
                  intr=(1-exp((-1)*(exp(alpha))*(exp(Z3)+1)))/(exp(Z3)+1);
                  intr3=(exp(alpha-(exp(alpha)*(exp(Z3)+1))));
                  intr3=(intr-intr3)/(exp(Z3)+1);
                  if i>12 then k=2; else k=1;
                  quan=((R[j,2]*ll)/(R2[j,k]*R2[j,3]));
                  ksum=intr-intr3;
                  ksum2=exp(Z3)*intr3;
                  new=(-1)*(R[j,3]*ksum)+(R[i,3]*ksum2);
                  weight=(1/(intr*(1-intr)));

                  if R[j,1]>=R[l,1] then qq2=1; else qq2=0;
                  if R[i,3]=R[l,3] then qq=1; else qq=0;
                  if R[j,3]=R[l,3] then qq3=1; else qq3=0;
                  qq=qq+qq3;
                  qq2=qq2*qq;
                  part2=new*weight*sw*qq2*quan+part2;

            end;
```

```
        end;

        gen=(one*(part2**2))*(ppsum**-2)+gen;
end;

gen=(4/(23**3))*gen;

serror=(gamsum-gen)/(lh*lh);

/*St. Error value of the b estimator*/
serror=sqrt(serror/23);

/*The Wald statistic*/
W=(b/serror)**2;

/*The p-value of the statistic*/
pvalue=1-cdf("chisquared",W,1);

quit;
```

- Manipulation of the Likelihood and the Efron R^2

```
proc iml;
reset autoname;
use sasuser.leukem2;
read all into R;

use sasuser.marios2;
read all into R2;

coll1=R[,1];
coll1=UNIQUE(coll1);
coll1=shape(coll1,18,1);

num=17;
tzero=coll1[num];
b=0.88749;
alpha=-2.130802;

/*The mean of the y_ij*/
ymean=0.3036616;

/*The log--ikelihood acumulator*/
lik=0;
rsq=0;
rsq2=0;
      do i=1 to 23;
            do j=i to 23;
            Z2=R[j,3]-R[i,3];
            Z3=Z2*b;
            if i=j then tie=0; else tie=1;
            alpha=Zzero*b+log(-log(surviv));
            a1=R[i,1];
            a2=R[j,1];
            a3=min(a1,tzero);
            ll=0;
            if a3>=a2 then ll=1;
            intr=(1-exp((-1)*(exp(alpha))*(exp(Z3)+1)))/(exp(Z3)+1);
/*The third column of R2 contains G_j(T_j), whereas the first is equal to
G_1group(T_j) and the second is equal to G_2group(T_j)*/
            if i>12 then k=2; else k=1;
            quan=((R[j,2]*ll)*tie/(R2[j,k]*R2[j,3]));
            lik=lik+(quan*log(intr)+(1-quan)*log(1-intr))*tie;
            rsq=rsq+((quan-intr)*(quan-intr));
            rsq2=rsq2+((quan-ymean)*(quan-ymean));

          end;
```

```
    end;

/*Efron R square*/
rsq=1-rsq/rsq2;
print rsq,lik;
quit;
```

C. CODE IN SAS SOFTWARE FOR THE EVALUATION OF THE UNKNOWN PARAMETER β and the estimating survivals setting λ=2 in (6.21)

```
/*Putting the data in the same way like before*/
proc iml;
reset autoname;
use sasuser.leukem2;
read all into R;

use sasuser.marios2;
read all into R2;

use sasuser.mar;
read all into T2;

use sasuser.mar2;
read all into coll;

bmin=1;
bmax=2;
b=(bmin+bmax)/2;
usum=1;
do while(abs(usum)>=1E-5);

usum=0;
      do i=1 to 23;
            do j=1 to 23;
            tie=1;
            if i=j then tie=0;
            Z2=R[i,3]-R[j,3];
            Z3=Z2*b;
            a1=R[i,1];
            a2=R[j,1];
            ll=0;
            if a1>=a2 then ll=1;

/*Adjusting the generalized functions in the error distribution, excluding the case
where b=0 or Zi=Zj*/
            intr=(-1)*(6*Z3*exp(2*Z3)-2*exp(3*Z3)-1+6*exp(Z3)-3*exp(2*Z3));
            if (Z3<>0) then intr=intr/(((-1)*exp(2*Z3)+2*exp(Z3)-1)**2); else intr=1;
            weight=(-2)*(-
6*Z3*exp(4*Z3)+6*Z3*exp(2*Z3)+exp(5*Z3)+8*exp(4*Z3)-
18*exp(3*Z3)+8*exp(2*Z3)+exp(Z3))*(-exp(2*Z3)+2*exp(Z3)-1);
```

```
            if (Z3<>0) then weight=weight/(6*Z3*exp(2*Z3)-2*exp(3*Z3)-
1+6*exp(Z3)-3*exp(2*Z3))*(exp(4*Z3)-
6*exp(3*Z3)+3*exp(2*Z3)+2*exp(Z3)+6*Z3*exp(2*Z3));
            else weight=1;
            IF i>12 then k=2; else k=1;
            quan=((R[j,2]*ll)/(R2[j,k]*R2[j,3]));
            quan=quan-intr;
            usum=usum+tie*weight*Z2*quan;
          end;
      end;
beta=b;
if usum>0 then bmin=b; else if usum<0 then bmax=b;
b=(bmin+bmax)/2;
end;

/*b estimate*/
b=beta;

/*SURVIVAL ESTIMATES*/
/*Calculation of the h function estimates which here are denoted as hsim */
hsim=J(20,1);
Ssim=J(20,1);

do j=1 to 20;
t=coll[j,2];
hmin=-20;
hmax=20;
h=(hmin+hmax)/2;
usum=1;
do while(abs(usum)>=1E-7);
usum=0;
      do i=1 to 23;
      if R[i,1]>=t then quan=1; else quan=0;
      k=1;
      if i>12 then k=2;
      first=quan/T2[j,k];
      second=((2/(2+exp(h+b*R[i,3])))**2);
      usum=usum+(first-second);
      end;
heta=h;
if usum>0 then hmax=h; else if usum<0 then hmin=h;
h=(hmin+hmax)/2;
end;
h=heta;
hsim[j]=h;

/*Survival probabilities for this model*/
```

```
ssim[j]=(2/(2+exp(h+b*coll[j,3])))**2;
end;

quit; /*Procedure ends*/
```

D. CODE IN SAS SOFTWARE FOR THE EVALUATION OF THE UNKNOWN PARAMETER β and the estimating survivals setting λ=1 (Proportional Odds model) in (6.21)

```
/*Putting the data in the same way like before*/
proc iml;
reset autoname;
use sasuser.leukem2;
read all into R;

use sasuser.marios2;
read all into R2;

use sasuser.mar;
read all into T2;

use sasuser.mar2;
read all into coll;

bmin=-1;
bmax=3;
b=(bmin+bmax)/2;
usum=1;
do while(abs(usum)>=1E-9);

usum=0;
      do i=1 to 23;
            do j=1 to 23;
            tie=1;
            if i=j then tie=0;
            Z2=R[i,3]-R[j,3];
            Z3=Z2*b;
            a1=R[i,1];
            a2=R[j,1];
            ll=0;
            if a1>=a2 then ll=1;

/*Adjusting the generalized functions in the error distribution, excluding the case
where b=0 or Zi=Zj*/
            if (Z3<>0) then intr=((1+(Z3-1)*exp(Z3))*(exp(Z3)-1)**-2); else intr=0;
            weight=(exp(Z3)-1)*(Z3+2+(Z3-2)*exp(Z3));
            if (Z3<>0) then weight=weight/((Z3+1-exp(Z3))*(1+(Z3-1)*exp(Z3))); else
weight=0;
            if i>12 then k=2; else k=1;
            quan=((R[j,2]*ll)/(R2[j,k]*R2[j,3]));
            quan=quan-intr;
            usum=usum+tie*weight*Z2*quan;
```

```
        end;
    end;
beta=b;
if usum>0 then bmin=b; else if usum<0 then bmax=b;
b=(bmin+bmax)/2;
end;

/*b estimate*/
b=beta;

/*SURVIVAL ESTIMATES*/
/*Calculation of the h function estimates which here are denoted as hsim */
hsim=J(20,1);
Ssim=J(20,1);

do j=1 to 20;
t=coll[j,2];
hmin=-30;
hmax=31;
h=(hmin+hmax)/2;
usum=1;
do while(abs(usum)>=1E-9);
usum=0;
    do i=1 to 23;
    if R[i,1]>=t then quan=1; else quan=0;
    k=1;
    if i>12 then k=2;
    first=quan/T2[j,k];
    second=(1/(1+exp(h+b*R[i,3])));
    usum=usum+(first-second);
    end;
heta=h;
if usum>0 then hmax=h; else if usum<0 then hmin=h;
h=(hmin+hmax)/2;
end;
h=heta;
hsim[j]=h;

/*Survival probabilities for this model*/
ssim[j]=(1/(1+exp(h+b*coll[j,3])));
end;

e=J(23,23);

do i=1 to 23;
```

```
      do j=1 to 23;
      if j=i then sw=0; else sw=1;
      Z2=R[i,3]-R[j,3];
      Z3=Z2*b;
      if R[i,1]>=R[j,1] then quan=1; else quan=0;
      if i>12 then k=2; else k=1;
      quan1=(R2[j,k]*R2[j,3]);
      if Z3<>0 then quan3=((1+(Z3-1)*exp(Z3))*((exp(Z3)-1)**-2)); else quan3=0;
      alpha=((R[j,2]*quan)/quan1)-quan3;
      alpha=alpha*sw;
      e[i,j]=alpha;
      end;
end;


gamsum=0;
do i=1 to 23;
      do j=1 to 23;

            Z2=R[i,3]-R[j,3];
            Z4=Z2*b;
            intr4=(exp(Z4)-1)*(Z4+2+(Z4-2)*exp(Z4));
            if Z4<>0 then intr4=intr4/((Z4+1-exp(Z4))*(1+(Z4-1)*exp(Z4))); else
intr4=0;
            Z6=R[j,3]-R[i,3];
            Z6=Z6*b;
            intr6=(exp(Z6)-1)*(Z6+2+(Z6-2)*exp(Z6));
            if Z6<>0 then intr6=intr6/((Z6+1-exp(Z6))*(1+(Z6-1)*exp(Z6))); else
intr6=0;


            do k=1 to 23;
                  if k=j then contr=0; else contr=1;
                  Z3=R[i,3]-R[k,3];
                  Z5=Z3*b;
                  intr5=(exp(Z5)-1)*(Z5+2+(Z5-2)*exp(Z5));
                  if Z5<>0 then intr5=intr5/((Z5+1-exp(Z5))*(1+(Z5-1)*exp(Z5))); else
intr5=0;
                  Z7=R[k,3]-R[i,3];
                  Z7=Z7*b;
                  intr7=(exp(Z7)-1)*(Z7+2+(Z7-2)*exp(Z7));
                  if Z7<>0 then intr7=intr7/((Z7+1-exp(Z7))*(1+(Z7-1)*exp(Z7))); else
intr7=0;

                  gam1=(intr4*e[i,j]+intr6*e[j,i]);
                  gam2=(intr5*e[i,k]+intr7*e[k,i])*Z2*Z3;
                  gamsum=gamsum+gam1*gam2*contr;
```

```
            end;
        end;
end;

gamsum=(1/12167)*gamsum;

lh=0;
do i=1 to 23;
        do j=1 to 23;
        Z2=R[i,3]-R[j,3];
        Z3=Z2*b;
        if (Z3<>0) then intr=(-1)*exp(Z3)*((2+Z3+(Z3-2)*exp(Z3))*(exp(Z3)-1)**-3);
else intr=0;
        weight=(exp(Z3)-1)*(Z3+2+(Z3-2)*exp(Z3));
        if Z3<>0 then weight=weight/((Z3+1-exp(Z3))*(1+(Z3-1)*exp(Z3))); else
weight=0;
        lh=Z2*Z2*intr*weight+lh;
        end;
end;

lh=(1/529)*lh;
print gamsum,lh;


gen=0;
do l=1 to 23;
        one=1-R[l,2];
        ppsum=0;
        do k=1 to 42;
                if (R[k,1]>=R[l,1] & R[k,3]=R[l,3]) then pp=1; else pp=0;
                ppsum=ppsum+pp;
        end;
        part2=0;
        do i=1 to 23;
                do j=1 to 23;
                        if R[i,1]>=R[j,1] then qq=1; else qq=0;
                        if R[j,1]>=R[l,1] then qq2=1; else qq2=0;
                        Z4=R[i,3]-R[j,3];
                        Z5=Z4*b;
                        weight=(exp(Z5)-1)*(Z5+2+(Z5-2)*exp(Z5));
                        if Z5<>0 then weight=weight/((Z5+1-exp(Z5))*(1+(Z5-1)*exp(Z5)));
else weight=0;
                        if i>12 then k=2; else k=1;

                        part2=((weight*Z4*R[j,2]*qq*qq2)/(R2[j,k]*R2[j,3]))+part2;

                end;
```

```
        end;

        gen=(one*(part2**2))*(ppsum**-2)+gen;
end;

gen=(4/12167)*gen;

serror=(gamsum-gen)/(lh*lh);

/*S. error value*/
serror=sqrt(serror/23);

/*The Wald statistic*/
W=(b/serror)**2;

/*The p-value of the statistic*/
pvalue=1-cdf("chisquared",W,1);
quit;
```

- Manipulation of the Likelihood and the Efron R^2

```
proc iml;
reset autoname;
use sasuser.leukem2;
read all into R;

use sasuser.marios2;
read all into R2;

/*The estimated value of the unknown parameter is*/
b=0.93126;

/*The mean of the y_ij*/
ymean=0.5072159;

/*The log-likelihood accumulator */
lik=0;
rsq=0;
rsq2=0;


      do i=1 to 23;
            do j=1 to 23;
            tie=1;
            if i=j then tie=0;
            Z2=R[i,3]-R[j,3];
            Z3=Z2*b;
            a1=R[i,1];
            a2=R[j,1];
            ll=0;
            if a1>=a2 then ll=1;
            if (Z3<>0) then intr=((1+(Z3-1)*exp(Z3))*(exp(Z3)-1)**-2); else intr=0.5;
/*The third column of R2 contains G_j(T_j), whereas the first is equal to
G_1group(T_j) and the second is equal to G_2group(T_j)*/
            if i>12 then k=2; else k=1;
            quan=((R[j,2]*ll)*tie/(R2[j,k]*R2[j,3]));
            lik=lik+(quan*log(intr)+(1-quan)*log(1-intr))*tie;
            rsq=rsq+((quan-intr)*(quan-intr))*tie;
            rsq2=rsq2+((quan-ymean)*(quan-ymean));
          end;
      end;

/*Efron R square*/
rsq=1-rsq/rsq2;
print rsq,lik;
quit;
```

E. CODE IN SAS SOFTWARE FOR THE EVALUATION OF THE UNKNOWN PARAMETER β AND ITS VARIANCE USING THE SCORE FUNCTION GIVEN IN (5.39). ALSO THE SURVIVOR ARE ESTIMATED USING (5.38)

```
proc iml;
use sasuser.leukem2;
reset autoname;

/* Put the data into the matrix D */
read all into D;
coll=D[,1];
coll=UNIQUE(coll);

/*Create the N matrix of the counts and the Y the indicator process*/
N=J(18,23);
Y=J(18,23);

/*Create the explanatory variable matrix*/
Z=D[,3];
Z=shape(z,1,23);
Z=repeat(z,18);

do i=1 to 23;
do j=1 to 18;
      if d[i,1]>coll[j] then N[j,i]=0; else N[j,i]=1;
      if d[i,1]>=coll[j] then Y[j,i]=1; else Y[j,i]=0;
end;
end;

gin=Y#Z;
zhat=gin[,+]/Y[,+];

/*Evaluate the denominator in the quantity (5.49)*/
s=0;
do i=1 to 23;
dif=z[,i]-zhat;
dif=dif#dif;
pos=0;
      do j=1 to 17;
      pos=pos+((coll[j+1]-coll[j])#Y[j+1,i]#dif[j+1]);
      end;
s=s+pos;
end;

/*Evaluate the nominator in the quantity (5.49)*/
s2=0;
```

```
do i=1 to 23;
dif=z[,i]-zhat;
pos2=0;
      do j=1 to 17;
      pos2=pos2+((N[j+1,i]-N[j,i])#dif[j+1]);
      end;
s2=s2+pos2;
end;

bhat=s2/s;
hal=J(18,12);
do i=1 to 12;
hal[,i]=n[,i];
end;
hal=hal[,+];
hal2=J(18,11);
do i=1 to 11;
hal2[,i]=n[,12+i];
end;
hal2=hal2[,+];
y1=J(18,12);
do i=1 to 12;
y1[,i]=y[,i];
end;
y1=y1[,+];
hbase=j(18,1,0);
j[1]=1;

/*Calculate the estimated hazards*/
do j=0 to 16;
s1=0;
s2=0;
s3=0;
      do i=1 to 17-j;
      s1=s1+(hal[i+1]-hal[i])/Y[i+1,+];
      s2=s2+(hal2[i+1]-hal2[i])/Y[i+1,+];
      s3=s3+bhat#y1[i+1]/Y[i+1,+];
      end;
hbase[18-j]=s1+s2-s3;
end;

/*and the corresponding survivors*/
surv=j(18,1,0);
do j=0 to 16;
s=0;
      do i=1 to 17-j;
      s=s+bhat#(coll[i+1]-coll[i]);
```

```
      end;
surv[18-j]=s;
end;
do i=1 to 18;
surv[i]=exp(-hbase[i]-bhat*coll[i]);
end;
surv2=exp(-hbase);
hder=J(18,1,1);
do i=1 to 17;
```

- Manipulation of the Likelihood and the Efron R^2

```
      hder[i]=(hbase[i+1]-hbase[i])/(coll[i+1]-coll[i]);
end;
s2=0;
do i=1 to 23;
pos=0;
pos2=0;
      do j=1 to 17;
      pos2=pos2+((N[j+1,i]-N[j,i])#log(hder[j+1]+bhat#Z[j+1,i]));
      pos=pos+((coll[j+1]-coll[j])#Y[j+1,i]#(hder[j+1]+bhat#Z[j+1,i]));
      end;
s2=s2+(pos2-pos);
end;
print s2;
quit; /*Quit procedure IML*/
```

REFERENCES

Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. *In Lecture notes in Statistics,* Springer, New York

Aalen, O.O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8, 907-925

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study, *Annals of Statistics,* 10, 1100-1120

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine,* 2, 273-277

Binder, D.A. (1983). On the variances of asymptotically normal estimators surveys. *Int. Statist. Rev.,* 51, 279-292

Binder, D.A. (1992). Fitting Cox's proportional hazards models from survey data. *Biometrika,* 79, 139-147

Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics,* 30, 89-100

Breslow, N.E. and Day, N.E. (1980). *Statistical Models in Cancer Research,* 1, The design and Analysis of Case-Control Studies, IARC, Lyon

Breslow, N.E. and Day, N.E. (1987). *Statistical Methods in Cancer Research,* 2, The design and Analysis of Cohort Studies, IARC, Lyon

Buckley, J.D. (1984). Additive and Multiplicative Models for Relative Survival Rates, *Biometrics,* 40, 51-62

Cheng, S.C., Wie, L.J. and Ying, Z. (1995). Analysis of transformation models with censored data, *Biometrika,* 82, 835-845

Cheng, S.C., Wie, L.J. and Ying, Z. (1997). Predicting Survival Probabilities With Semiparametric Transformation Models, *Journal of the American Statistical Association,* 92, 227-235

Cox, D.R. (1972). Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society,* B, 34, 187-220

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* Chapman and Hall, London

Lin, D.Y. (2000). On Fitting Cox's Proportional Hazards models to survey Data, *Biometrika,* 87, 37-47

Link, C.L. (1984). Confidence Intervals for the Survival Function using Cox's Proportional-Hazard Model with Covariates, *Biometrics,* 40, 601-610

Mann, N.R., Shafer, R.E. and Singpurwalla, N.D. (1974). *Methods for Statistical Analysis of Reliability and Life Data.* Wiley, New York

Mike´, V. and Stanley, K.E. (1982) *Statistics in Medical Research.* Wiley, New York

Miller, R.G. (1981). *Survival Analysis.* Wiley, New York

Oakes, D. (1986). An approximate Likelihood Procedure for Censored Data, *Biometrics,* 42, 177-182

O'Quigley, J. (1986). Confidence Intervals for the Survival Function in the Presence of Covariates, *Biometrics,* 42, 219-220

Peto, R. and Lee, P. (1973). Weibull distributions for continuous carcinogenesis experiments, *Biometrics,* 29, 457-470

Pett, M.A. (1997). *Nonparametric statistics for health care research. Statistics for small samples and unusual distributions.* Sage, California

Pettitt, A.N. (1982). Inference for the Linear Model Using a Likelihood Based on Ranks, *Journal of the Royal Statistical Society,* B, 44, 234-243

Pierce, D.A. and Preston, D.L. (1984). *Hazard function modeling for dose-response analysis of cancer incidence in A-bomb survivor data: Utilazation and Analysis.* SIAM, Philadelphia

Pike, M.C. (1966). A suggested method of analysis of a certain class of experiments in carcinogenesis, *Biometrics,* 22, 142-161

Pocock, S.J., Gore, S.M. and Kerr, G.R. (1982). Long Term Survival Analysis: The Curability of Breast Cancer, *Statistics in Medicine,* 1, 93-104

Prentice, R.L. (1973). Exponential survivals with censoring and explanatory variables, *Biometrika,* 60, 279-288

Rao, C.R. and Toutenburg, H. (1999). *Linear Models,* Springer, New York.

Thomas, D.C. (1981). General relative risk models for survival time and matched case-control analysis, *Biometrics,* 37, 673-686

Thomas, D.C. (1986). *Use of auxiliary information in fitting nonproportional hazards models. In Modern Statistical Methods in Chronic Disease Epidemiology.* Wiley, New York

Van Der Laan, M.J. and Hubbard, A.E. (1998). Locally efficient estimation of the survival distribution with right-censored data and covariates when collection of data is delayed, *Biometrika,* 85, 771-783

Wang, M.C. (1999). Hazards Regression Analysis for Length-Biased Data, *Biometrika,* 87, 343-354

Wei, L.J. (1984). Testing Goodness of Fit for proportional Hazards Model With Censored Observations, *Journal of the American Statistical Association,* 79, 649-652

Xekalaki, E. (1983a). A property of the Yule distribution and its application. *Communications in Statistics (Theory and Methods),* 12, 1181-1189

Xekalaki, E. (1983b). Hazard Functions and Life Distributions in Discrete Time, *Communications in Statistics (Theory and Methods),* 12, 2503 -2509

Yip, S.F., Zhou, Y., Lin D.Y. and Fang, X.Z. (1999). Estimation of Population Size Based on Additive Hazards Models for Continuous-Time Recapture Experiments, *Biometrics,* 55, 904-908

Δωρεά