Sensitivity Analysis for Correlated Survival Models

by

Fotios Siannis

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Statistics

University of Warwick, Department of Statistics
December 2001

Contents

List of Figures

List of Tables

Notation

$\mathbf{f}_{T,C}(\mathbf{t}, \theta, \gamma)$: Joint density function

$\mathbf{S}_{T,C}(\mathbf{t}, \theta, \gamma)$: Joint survival function

$\mathbf{f}_T(t, \theta)$: Density function of T–Process

$\mathbf{f}_C(c, \gamma)$: Density function of C–Process

$\mathbf{h}_T(t, \theta)$: Hazard function of T Process

$\mathbf{h}_T(c, \gamma)$: Hazard function of C–Process

$\mathbf{H}_T(t, \theta)$: Cumulative (integrated) hazard function of T–Process

$\mathbf{H}_T(c, \gamma)$: Cumulative (integrated) hazard function of C–Process

$\mathbf{f}_T^{\sharp}(t)$: Sub–density function of T–Process

$\mathbf{f}_C^{\sharp}(c)$: Sub–density function of C–Process

$\mathbf{h}_T^{\sharp}(t)$: Sub–hazard function of T–Process

$\mathbf{h}_C^{\sharp}(c)$: Sub–hazard function of C– Process

Acknowledgements

First of all I would like to thank my mother and father for their invaluable support not only during my years in Warwick but for all my academic years. I would also like to thank all my friends in Warwick University who supported me when I needed it and shared with me some great moments, especially Rania, Judith and of course Theodosis, with whom we came together all the way to the end. Furthermore, I would like to thank Guobing Lu for his precious statistical advice. Finally I would like to thank my supervisor Professor John Copas, not only for guiding me to the completion of this thesis, but also for introducing to me a whole new analytical way of thinking, which is a priceless lesson not only in statistics but in life in general.

This research was supported by the University of Warwick graduate award and the EPSRC postgraduate studentship award.

Declaration

I hereby declare that this thesis is based on my own research, except when stated otherwise.

Summary

In this thesis we introduce a model for informative censoring. We assume that the joint distribution of the failure and the censored times depends on a parameter δ, which is actually a measure of the possible dependence, and a bias function $B(t, \theta)$. Knowledge of δ means that the joint distribution is fully specified, while $B(t, \theta)$ can be any function of the failure times. Being unable to draw inferences about δ, we perform a sensitivity analysis on the parameters of interest for small values of δ, based on a first order approximation. This will give us an idea of how robust our estimates are in the presence of small dependencies, and whether the ignorability assumption can lead to misleading results.

Initially we propose the model for the general parametric case. This is the simplest possible case and we explore the different choices for the standardized bias function. After choosing a suitable function for $B(t, \theta)$ we explore the potential interpretation of δ through it's relation to the correlation between quantities of the failure and the censoring processes. Generalizing our parametric model we propose a proportional hazards structure, allowing the presence of covariates. At this stage we present a data set from a leukemia study in which the knowledge,under some certain assumptions, of the censored and the death times of a number of patients allows us to explore the impact of informative censoring to our estimates. Following the analysis of the above data we introduce an extension to Cox's partial likelihood, which will call "modified Cox's partial likelihood", based on the assumptions that censored times do contribute information about the parameters of interest. Finally we perform parametric bootstraps to assess the validity of our model and to explore up to what values of parameter δ our approximation holds.

Chapter 1

Introduction

The research for this thesis was made in the area of Survival Analysis. In survival data censoring occurs very often, and all the existing ways of analyzing these data conveniently assume that censoring is uninformative. There is no statistical methodology to be widely approved for analyzing survival data which allows the possibility that censoring is not a random procedure, and hence informative. We have focused on this particular problem, hoping that we will manage to propose a well stated and well supported solution.

The main aim of this thesis is to introduce a new model which will enable us to analyze survival data, considering the possibility that censoring might not be completely at random. This means that we assume that the censoring process might follow a specific pattern, containing information that we would like to extract. In this case the failure and the censoring processes are not independent, and the level of dependence is of major interest. It would be very optimistic to believe that we can reveal the particular pattern that the censoring process has or even to calculate the exact level of dependence between the two

processes. Therefore, what we really try to do is to propose a model with which we can perform a sensitivity analysis on our estimates, assuming that we have a small level of dependence. Our conclusions are based on the impact that this sensitivity analysis has on the parameters of interest or to quantities which can be more easily interpreted like the survival curve or the median survival time. For example, if we assume dependence between the two processes and the change in the median survival time is minor, compared to the estimate under independence, then this means that our estimate is robust in different levels of dependence. Therefore the analysis of these survival data using one of the existing statistical procedures can be considered adequate, because in any other case where we use complicated models to account the potential dependence the practical result is not so different from the one we already have. However, if the change in the median survival time is significantly large, then the results obtained from the analysis of the data using statistical methods which assume uninformative censoring might be very misleading. In this case, our model provides us with a possible range of values for the median survival time, for some assumed levels of dependence, giving in that way an idea of the error that we make by not assuming informative censoring.

This thesis consists of seven more chapters, apart from the introductory one. In chapter 2 we present some of the existing statistical methods of analyzing survival data. We talk about Competing Risks theory and Frailty models, exploring the similarities and the differences between these theories and what we are trying to do within the Survival Analysis framework. A literature review is made and a section is also included, explaining what motivated us to do this work. Chapter 3 is where we propose our model, explaining the

assumptions behind it. Initially we discuss the parametric case, although the principal ideas will not change later on when we will talk about the semi parametric case. In the model, a parameter δ and an unknown correlation function $B(t, \theta)$ are introduced. δ represents the level of dependence between the failure and the censoring processes, while $B(t, \theta)$ represents the way that the two processes are related to each other. Parameter δ is the most important quantity in the model and the basis of the sensitivity analysis. Therefore, we devote chapter 4 exploring the role of δ, and discussing what interpretation and what properties this parameter might have. In chapter 5 we explore $B(t, \theta)$, trying to see which function is the most appropriate. After making our choice, we generalize the model allowing for the presence of explanatory variables. An important section in this chapter is the last one, where we discuss the case where the censoring at the end of the study is uninformative. This introduces the idea that both informative and uninformative censoring might be present at the same time, hence the need for a model to take into account both types of censorings was unavoidable. Such a model is introduced at the end of chapter 6, while in the beginning a special data set is analyzed, to show that our theory works and that if we had additional information an estimate of parameter δ could have been feasible.

In chapter 7 the semi–parametric version of the model is presented. It is based on Cox's partial likelihood function, which is altered in order to introduce what we will call the Modified Cox's Partial Likelihood (MPL). Finally in chapter 8 a simulation study was made to prove the validity of our model. Several examples are included in all the chapters in order to show how are methods work.

Chapter 2

The Impact of Informative Censoring in the Analysis of Survival Data

2.1 Analyzing Survival Data and Non–Identifiability Issues

Clinical trials are designed to test new drugs or treatments and come up, if possible, with reliable answers to very important questions. Depending on the trial and what is being tested we might observe patients to have a remission period, to relapses or even to die, and the times to these events are of great importance. Unfortunately, we are not always able to observe the event of importance to all the patients. A major reason for that is that the trials cannot run for such a long time so that we are able to observe it, and an other equally important reason is that many patients may quit the trials for reasons

that are not always known. Hence, we end up with data sets which contain a number of incomplete observations, the censored times. Our aim is to use these data to obtain as much information as if we had an uncensored data set.

New statistical methods had to be introduced, and the way we deal with the censored observations is still a subject of great debate. Whether censoring happens at random and hence is non informative or not is something that we cannot detect from the data themselves. Therefore, assumptions need to be made and models need to be constructed in order to proceed with the analysis of the data. The most common approach is to assume non–informative censoring, and in reliability studies this type of censoring happens very often. In an experiment where we test a number of machines it is very difficult and time consuming to follow them up until all of them break down. Therefore, we follow them up to a certain time point where some have already failed and some are still working. Similarly in clinical trials patients drop out of the study for reasons which are not related to the study itself. In both the above cases censored observation are produced, which we can easily assume happen at random.

Kaplan & Meier(1958) introduced the product–limit estimator of the survival curve. It is non–parametric and the main assumption is that the censored times carry no information about the distribution of the failure times, and hence they are independent. This method has become a standard procedure for estimating the survival curve when the independence assumption seems reasonable. Even in the case where this assumption is questionable, the product–limit estimate is always obtained to show at least what the survival curve would look like if censoring was non–informative. This estimator is consistent for the class of

constant sum survival models introduced by Williams & Lagakos(1977). Cox(1972) introduced the proportional hazard model, where the hazard was proportional to an unknown baseline hazard function. This function was multiplied by a quantity which was dependent on the set of explanatory variables, that followed each patient in the trial. The partial likelihood provides us with estimates of the parameters that multiplied these variables, without the need for knowledge of the baseline hazard function, for which an estimation procedure is also suggested. The following year Kalbfleisch & Prentice(1973) provided some extra justification to Cox's partial likelihood, when no tied observations occur, and they also proposed a step function as an estimate to the baseline hazard function. Parametric models have also been considered and proven helpful. Cox & Oakes(1984) consider models such as the Weibull, exponential and Gompertz–Makeham.

Cox(1959) suggested that patients are exposed to more than one risks and hence if they die, their death might be due to any of these risks. In this work he proposed four models for bivariate data, and he immediately acknowledged the difficulties related to their interpretation. When we have many competing risks and a population is subject to k causes of death, and suppose that each individual is characterized by a vector $\mathbf{T} = (T_1, T_2, \ldots, T_k)$ of times at which he dies, respectively, of the k causes, then Moeschberger & David(1971) emphasized that only the minimum of these times along with the associated cause of death are observed. This approach with the latent failure times was adopted by Gail(1975) in his "Review and critique of some models used in competing risks analysis". A detailed discussion of this approach was presented in David & Moeschberger(1978). Moreover, the survival analysis problem can be regarded as a special case of the competing risks problem.

Although in this problem we observe failure times from only one cause of failure, censored times can be seen as a result of a second cause of failure, different from the one under investigation. Hence the problem can be transformed into a competing risks problem with only two competing risks.

However, the fact that the assumption of non–informative censoring, or independent risks, was untestable made statisticians feel very uncomfortable with this idea. Tsiatis(1975) proved in the competing risks framework the non identifiability of dependent risk models. In other words, if someone assumes a model with dependent risks then there is always a proxy model with independent risks which can reproduce exactly the same sub–densities. Crowder(1991) supports this idea, showing that even when the marginal distribution is known the joint distribution is still not identifiable. Lagakos(1979) gave real life examples where the non-informative censoring assumption was questionable, while Peterson(1976) had already argued that "serious errors can be made in estimating the (potential) survival functions in the competing risks problem if the risks are assumed to be independent when in fact they are not". Therefore, he was the first to introduce bounds on the joint and the marginal distribution functions, with fixed sub–distributions, allowing any kind of dependence, in order to investigate the potential error we make if our assumptions are wrong. People later on claimed that Peterson's bounds were very wide. Slud & Rubinstein(1983) claimed that they could "improve dramatically" Peterson's bounds, based on a weak non–parametric assumption. A few years later Klein & Moeschberger(1988) proposed a model where the joint distribution belonged to a family of distributions indexed by a dependence parameter θ, with arbitrary marginals. Specifying a range of possible values

for the dependence parameter would produce bounds on the net survival probabilities. In a slightly different context, using a frailty model, Link(1989) claimed that in the case where censoring indicates an unfavorable prognosis the Kaplan and Meier estimate (KME) of the survival curve will tend to overestimate the survival probabilities. Therefore he suggested that "when censoring carries an unfavorable prognosis for future survival, reasonable estimators should be bounded above by the KME and below by the empirical survival function of the observed random variable".

An early attempt to model dependence between death and censored times was made by Fisher & Kanarek(1974). They proposed a model in which for an individual with censored time $C = c$, a survival time of $t - c$ after censoring is equivalent to a survival $\alpha(t - c)$ if there had been no censoring, where $\alpha > 0$. More specifically $\alpha > 1$ indicates a poorer while $\alpha < 1$ indicates a more favorable prognosis for the individual. Heckman & Honore(1989), under some regularity conditions, showed that if the patients are followed by a vector of covariates $\mathbf{Z}$ then the joint distribution is identifiable. The following year Hoover & Guess(1990) proposed a parametric model for the response linked censoring, which is the censoring caused by the fact that the response is about to happen. Following this definition, a positive association between the censoring and the response time was introduced.

There were two articles that helped a lot in the compilation of this section. The first one was a review on the "Identifiability Crises in Competing Risks" by Crowder(1994). In this paper the identifiability problem is discussed extensively, explaining the main theoretical results. The second paper was written by Moeschberger & Klein(1995), and provided a massive literature review on statistical methods developed up to that time for dependent

competing risks.

2.2 New Approach to the Identifiability Problem

In the previous section we saw different approaches that statisticians had taken in order to deal with the presence of informative censoring and the identifiability problem in general. Assumptions need to be made, some times arbitrary and restrictive, in order to be able to model cases where, we think from the context that, dependence between the possible risks exists. If for example we assume that we have data for which additional information is available, like doctor's opinions or the reasons for which patients are censored, then we might have a good idea of the type of censoring being present. In this case reasonably safe assumptions can be made and ad hoc models can be constructed in order to analyze the data.

The problem we want to tackle is how to analyze data when things are not that clear. How safe is it to assume non-informative censoring when no information exist to imply the opposite or even when knowledge of the reasons why patients are censored is not clear enough to give a good idea of the type of censoring we have. In other words we need to find a way to test whether the analysis of survival data, when independence between the failure and censored times is assumed, is appropriate when small dependencies might exist. Therefore we want to explore the robustness of our estimates and see how misleading our inferences could be, if they are, under the independence assumption.

A data set was obtained from Klein & Moeschberger(1997), page 465, which is one

of the main data sets used to illustrate the methods discussed in this thesis. Details of the study are found in Copelan *et al.*(1991), and it is about bone marrow transplants to 137 patients. The treatment was given to two groups of patients with acute myeloctic leukemia (AML), which were divided into groups according to the risk of first remission (low–high), and to a third group of patients with acute lymphoblastic leukemia (ALL). Explanatory variables for each patient were recorded and the main purpose of the research was to compare the survival probabilities between these categories. In our case we focus only on the ALL group (38 patients), and these patients were followed up for a maximum period of 7 years. Within this period after the surgery 11 patients died, 13 were observed to relapse (and then die), and 14 were right censored at some point during the study (right censored). I need to mention that patients join the study at different time points, hence the follow up period is not the same for all the patients. The data are presented in Table 2.1, where

T_1 : disease–free survival time (time to relapse, death or end of study)

T_2 : time to death or on study time

S_1 : death indicator (1–dead, 0–alive)

R : relapse indicator (1–relapse, 0–disease free)

S_2 : disease free indicator (1–dead or relapsed, 0–alive disease free).

The fact that we are able to observe both the relapse and the death time for a number of patients makes this data set special, and I will immediately explain the reason why. In this particular example the treatment under investigation is surgery, which happens at the beginning of the follow up time of each patient. If a patient relapses after a certain time

	T_1	T_2	S_1	R	S_2
1	1	1	1	0	1
2	55	262	0	1	1
3	74	110	0	1	1
4	86	86	1	0	1
5	104	156	0	1	1
6	107	107	1	0	1
7	109	162	0	1	1
8	110	269	0	1	1
9	122	122	1	0	1
10	122	243	0	1	1
11	129	1279	0	1	1
12	172	172	1	0	1
13	192	262	0	1	1
14	194	194	1	0	1
15	226	226	0	0	0
16	230	371	0	1	1
17	276	276	1	0	1
18	332	350	0	1	1
19	383	417	0	1	1
20	418	418	1	0	1
21	466	466	1	0	1
22	487	487	1	0	1
23	526	526	1	0	1
24	530	530	0	0	0
25	609	781	0	1	1
26	662	716	0	1	1
27	996	996	0	0	0
28	1111	1111	0	0	0
29	1167	1167	0	0	0
30	1182	1182	0	0	0
31	1199	1199	0	0	0
32	1330	1330	0	0	0
33	1377	1377	0	0	0
34	1433	1433	0	0	0
35	1462	1462	0	0	0
36	1496	1496	0	0	0
37	1602	1602	0	0	0
38	2081	2081	0	0	0

Table 2.1: ALL Data

there is no alternative treatment, hence the only thing we can do is simply observe how long will this patient survives after being relapsed. However, we can see the above data set in a slightly different hypothetical context, which is very common in clinical trials. Assume that all the patients in the trial are under a specific treatment. They are followed up until they are observed either to die, withdraw from the study disease free (right censored, mainly due to the end of the study), or relapse. While the first two cases are straight forward to handle, the question is raised of how to deal with the relapsed cases. A common practice in clinical trials is that if a patient is not responding well in a new treatment then in order to prevent him from getting worst or even to save his life doctors might decide to take him off the treatment under investigation. Therefore, if we assume that the patients who relapse are taken out of the trial in order to receive an alternative treatment, then it is obvious that these patients were censored in an informative way. On the other hand, if we assume that these patients are not withdrawn from the trial in order to receive an alternative treatment, then they will be observed to die soon after their recorded relapse time, providing us with death times and not censored times any more. Summarizing the above we can see this data set from two different angles, which we will call viewpoints.

Viewpoint A: The patients who relapse are taken off the treatment and hence are considered to be censored observations. Our observed lifetimes are given by column T_1 in Table 2.1 and the corresponding death indicator variable is S_1, which for simplicity we will name them **Data** A. For example patient 19 relapses after 383 days in the trial and he eventually dies after 417 days. Under Viewpoint A we ignore his exact death time and we assume that this patients was censored after 383 days of follow up.

Viewpoint B : The patients who relapse are left in the trial and hence they are observed to die during their follow up time. In this case our observed lifetimes are given by column T_2 and the death indicator is S_2, forming **Data** B. Therefore, under Viewpoint B patient 19 will be observed to die after 417 days of follow up.

Patients who die disease free, eg. patient 9, or are censored disease free, eg. patient 15, maintain their status as death and censored observations under both Viewpoints. This means that the only difference between A and B is the way we treat the relapsed observations.

Making the common assumption that censoring is non–informative we get Figure 2.1 with the two Kaplan and Meier (KM) estimates of the survival curves of the above data. This is anyway the major assumption for the KM estimate for the survival curve, and this is what everybody would do in order to get an initial idea of the survival probabilities of these patients. The solid line is when we use Data A and the dotted line is the KM estimate of the survival when we use Data B. If we assume that Data A are the potential observed data then Data B are the "true" data, "true" in the sense that we are able to observe the exact death times of the informatively censored observations leaving us with only the random censoring, which will provide us with unbiased estimates of the parameters of interest. Note that, we observe a huge difference between the estimated survival curve for the observed data (solid line) and the "true" data (dotted line), indicating that the analysis of the data that include the informative censoring would give largely misleading results.

In this work we introduce the idea of sensitivity analysis, based on a paper by Copas

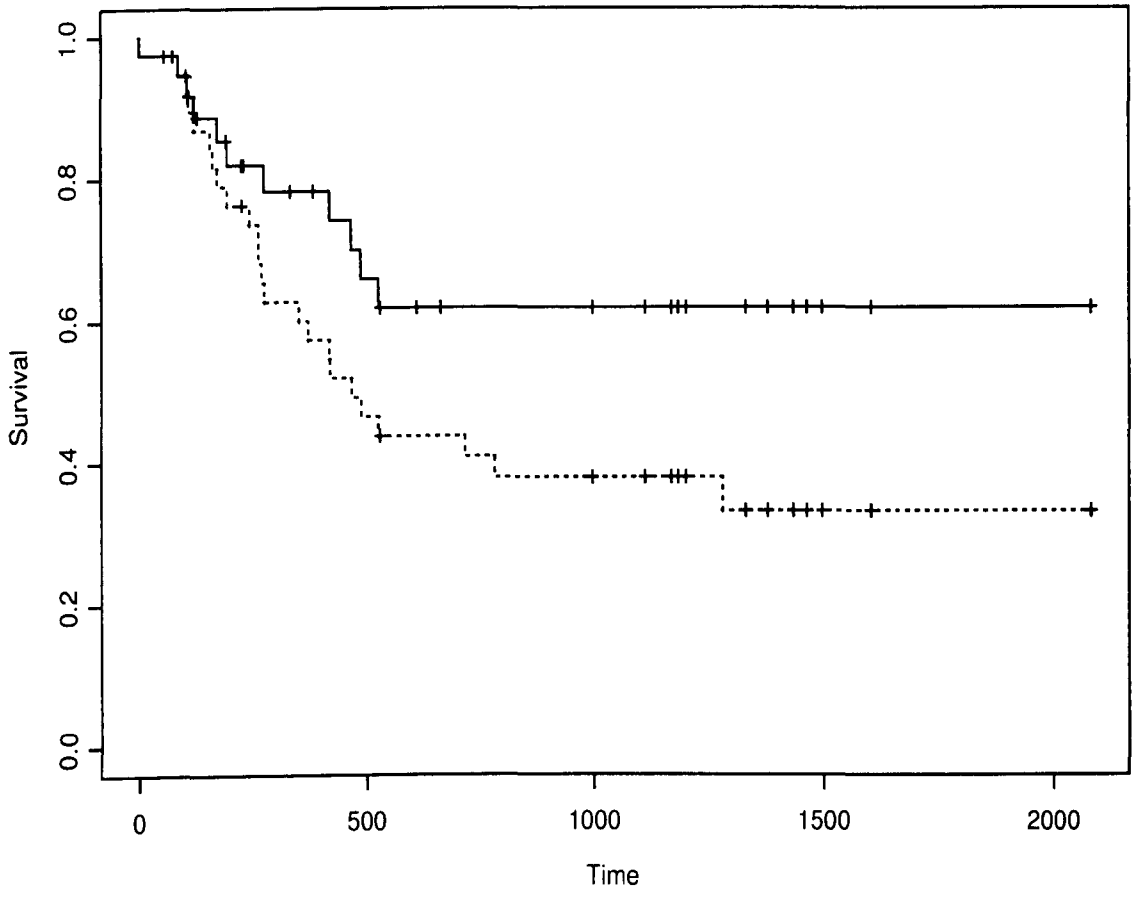

Figure 2.1: KM estimates of the survival curves for the two versions of the ALL data

& Eguchi(2001). In this paper the idea of sensitivity analysis was introduced, trying to deal with the problem of missing values. In our case we propose a model which allows for dependence in terms of a parameter δ and a bias function $B(t, \theta)$. Being unable to draw inferences about δ , we propose a sensitivity analysis on the estimate of the parameter of interest for small values of δ. The size of δ can be interpreted in terms of a correlation between the life time and the censoring mechanism.

Chapter 3

Model for Informative Censoring

3.1 The Definition of the Model

In the survival data, either we observe the time of occurrence of the event of interest, which is what we call a failure time, or this might be prevented by the occurrence of another event, and hence we observe a censored time. We assume that the two different kinds of observations form two different processes. The one with failure times, which will be named as T-process, and equivalently the one with censored times, which will be named as C-process. In terms of competing risk, it is like having two possible types of risk. The first one, which is the main risk under investigation, provides the failure times. The second one, which provides the censored times, summarizes all the other potential risks that might exist into one, only because we are not interested in any of these, and hence we allow them to be considered as one.

Our initial assumption is that the conditional density of the censored lifetimes given the exact failure times, which in any case we are not able to observe, has exactly the same

distributional form with the marginal distribution of C

$$P\big(C = c|T = t\big) \quad = \quad f_C\big(c, \gamma + \delta B(t, \theta)\big) \tag{3.1}$$

with the only difference being in the parameter of this distributions. $f_C(c, \gamma)$ is the marginal density of the C-process with parameter γ, and θ is the parameter of the distribution of the T process. We express the dependence between the two processes by allowing the parameter of the conditional distribution to depend on the failure times by using a function $B(t, \theta)$, multiplied by some quantity δ. Function $B(t, \theta)$ initially is assumed to be completely unknown and independent of γ. However, the most important part of the above equation is δ. We introduce this parameter as a measurement of the dependence between T and C processes. This doesn't mean necessarily that δ is equal to the statistical correlation, but we expect it to be strongly related to that, providing a better interpretation. This relationship will be explored in chapter 4. As we have mentioned several times before, the data do not provide enough information in order to estimate the level of dependence between the two processes and hence inferences about δ cannot be made. For that reason we assume that it is known. From (3.1) we see that when $\delta = 0$ the conditional distribution is exactly the same to the marginal implying independence, and this is an important possibility, that of ignorable censoring. Moreover, in our work we will allow δ to take only small values. This indicates that we are interested in small dependencies, what happens for values of δ around zero, something that will lead to mathematical approximations as the research goes on. Our aim is to perform a sensitivity analysis to all our estimates with respect to δ. Since δ is small, terms like δ^2, δ^3, ... are considered to be negligible and hence we omit them.

and the probabilities describe the two possible states that the data can be in, either failure of censored data. In order to have a consistent notation we use $f_*(*)$ for density, $h_*(*)$ for hazard, $H_*(*)$ for cumulative (integrated) hazard and $S_*(*)$ for survival functions. Hence we have

$$\begin{aligned} P(T = t \cap T < C) &= f_T(t,\theta) S_C\big(t, \gamma + \delta B(t,\theta)\big) \\ &\simeq f_T(t,\theta)\left[S_C(t,\gamma) + \delta B(t,\theta) \frac{\partial S_C(t,\gamma)}{\partial \gamma} \right] \\ &= f_T(t,\theta) S_C(t,\gamma)\left[1 - \delta B(t,\theta) \frac{\partial H_C(t,\gamma)}{\partial \gamma} \right] \end{aligned}$$

where

$$H_C(t,\gamma) = -\log S_C(t,\gamma)$$

is the cumulative hazard function and $S_C(t,\gamma)$ is the survival function of the censoring process. Furthermore we have

$$\begin{aligned} P(C = c \cap C < T) &= \int_c^\infty f_C\big(c, \gamma + \delta B(t,\theta)\big) f_T(t,\theta) dt \\ &\simeq \int_c^\infty \left[f_C(c,\gamma) + \delta B(t,\theta) \frac{\partial f_C(c,\gamma)}{\partial \gamma} \right] f_T(t,\theta) dt \\ &= f_C(c,\gamma) S_T(c,\theta)\left[1 + \delta \mu(c,\theta) \frac{\partial \log f_C(c,\gamma)}{\partial \gamma} \right] \end{aligned}$$

where

$$\mu(c,\theta) = \frac{\int_c^\infty B(t,\theta) f_T(t,\theta) dt}{S_T(c,\theta)}.$$

Therefore, the likelihood function is

$$L(t;\theta,\gamma,\delta) = \prod_{i=1}^{n}\left\{ f_T(t_i,\theta)S_C(t_i,\gamma)\left[\ 1 - \delta B(t_i,\theta)\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\right]\right\}^{I_i}$$

$$\left\{ f_C(t_i,\gamma)S_T(t_i,\theta)\left[\ 1 + \delta\mu(t_i,\theta)\frac{\partial \log f_C(t_i,\gamma)}{\partial\gamma}\right]\right\}^{1-I_i}$$

and the log-likelihood function, in first order approximation, will take the form

$$\begin{aligned} LL(t;\theta,\gamma,\delta) &\simeq \sum_{i=1}^{n}\Big[\ I_i \log h_T(t_i,\theta) + (1-I_i)\log h_C(t_i,\gamma) - H_T(t_i,\theta) - H_C(t_i,\gamma)\Big] \\ &+ \delta\sum_{i=1}^{n}\left[\ (1-I_i)\frac{\partial \log f_C(t_i,\gamma)}{\partial\gamma}\mu(t_i,\theta) - I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}B(t_i,\theta)\right] \\ &= LL(t;\theta,\gamma,\delta=0) \\ &+ \delta\sum_{i=1}^{n}\left[\ (1-I_i)\frac{\partial \log f_C(t_i,\gamma)}{\partial\gamma}\mu(t_i,\theta) - I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}B(t_i,\theta)\right]. \end{aligned} \tag{3.3}$$

For simplicity we set

$$LL(t;\theta,\gamma,\delta=0) = \sum_{i=1}^{n}\Big[\ I_i \log h_T(t_i,\theta) + (1-I_i)\log h_C(t_i,\gamma) - H_T(t_i,\theta) - H_C(t_i,\gamma)\Big] \tag{3.4}$$

which is the corresponding log-likelihood when $\delta = 0$. The unknown function $B(t,\theta)$ is included in the part of the likelihood multiplied by δ, which is essentially the correction factor. The fact the the whole term is multiplied by a small number, gives us some flexibility in the choice of $B(t,\theta)$, but still it's functional form is a major question. From the likelihood we intend to estimate θ, and at the same time we treat γ as a nuisance parameter. $\hat{\theta}_\delta$ is the estimate when δ is different form zero and $\hat{\theta}_0$ is the estimate when we

have independence. We differentiate the log-likelihood with respect to θ and we get

$$\begin{aligned}\frac{\partial LL(t;\theta,\gamma,\delta)}{\partial\theta} &= \left.\frac{\partial LL(t;\theta,\gamma,\delta=0)}{\partial\theta}\right|_{\theta=\hat{\theta}_\delta} \\ &+ \delta\sum_{i=1}^{n}\left[(1-I_i)\frac{\partial\log f_C(t_i,\gamma)}{\partial\gamma}\frac{\partial\mu(t_i,\theta)}{\partial\theta} - I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\frac{\partial B(t_i,\theta)}{\partial\theta}\right].\end{aligned}$$

Using Taylor's expansion we have that

$$\begin{aligned}\left.\frac{\partial LL(t;\theta,\gamma,\delta=0)}{\partial\theta}\right|_{\hat{\theta}_\delta} &\simeq \left.\frac{\partial LL(t;\theta,\gamma,\delta=0)}{\partial\theta}\right|_{\hat{\theta}_0} + (\hat{\theta}_\delta-\hat{\theta}_0)\frac{\partial^2 LL(t;\theta,\gamma,\delta=0)}{\partial\theta^2} \\ &= (\hat{\theta}_\delta-\hat{\theta}_0)\frac{\partial^2 LL(t;\theta,\gamma,\delta=0)}{\partial\theta^2}\end{aligned}$$

and finally what we get is

$$\hat{\theta}_\delta-\hat{\theta}_0 \simeq -\frac{\delta}{\frac{\partial^2 LL(t;\theta,\gamma,\delta=0)}{\partial\theta^2}}\sum_{i=1}^{n}\left[(1-I_i)\frac{\partial\log f_C(t_i,\gamma)}{\partial\gamma}\frac{\partial\mu(t_i,\theta)}{\partial\theta} - I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\frac{\partial B(t_i,\theta)}{\partial\theta}\right] \quad (3.5)$$

It is important to mention that no assumptions have been made so far about the distributions of the two processes and the unknown function $B(t,\theta)$, and (3.5) provides us with an expression for the difference between the two estimates. We see that

$$\imath(\theta_0) = -\frac{\partial^2 LL(t;\theta,\gamma,\delta=0)}{\partial\theta^2}$$

is the observed information, and hence

$$Var(\hat{\theta}_0) \simeq \imath(\theta_0)^{-1} = \left\{-\frac{\partial^2 LL(t;\theta,\gamma,\delta=0)}{\partial\theta^2}\right\}^{-1}$$

is the approximate variance of our estimate.

Therefore if we consider that T and C processes follow some convenient distribution,

then for a choice of $B(t,\theta)$ we can proceed and perform a sensitivity analysis for $\hat{\theta}_\delta$. What choice of $B(t,\theta)$ do we make is something that we explore in the next section. Moreover, it is obvious that an estimate of γ is required for our calculations. For that reason we obtain the independent estimate $\hat{\gamma}_0$ from (3.4) and we use it in order to proceed. The reason why we use the independent estimates in the correction factor is because in any other case terms of order $O(\delta^2)$ will be created, which we consider to be very small and we omit them.

3.3 Exploring the Unknown Function $B(t,\theta)$

3.3.1 Restrictions

We have assumed that we know the form of the conditional distribution $f_{C|T}(C = c|T = t)$. At the same time the marginal densities of T and C processes are $f_T(t,\theta)$ and $f_C(c,\gamma)$, with θ and γ being the parameters of the two distributions. For simplicity initially we assume that both are scalars. The joint density function is of the form

$$\begin{aligned} f_{T,C}(t,c) &= f_T\big(t,\theta\big) f_C\big(c,\gamma+\delta B(t,\theta)\big) \\ &\simeq f_T(t,\theta) f_C(c,\gamma)\left[1+\delta\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}B(t,\theta)\right]. \end{aligned} \tag{3.6}$$

The above expression is an equivalent expression of the definition of our model in (3.1). Therefore, the first requirement is that the joint density will provide us with the marginal

densities of the T and C processes. Hence, for the C-process we have

$$\begin{aligned} f_C(c,\gamma) &\simeq \int_0^\infty f_T(t,\theta) f_C(c,\gamma)\left[1+\delta\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}B(t,\theta)\right]dt \\ &= f_C(c,\gamma)+\delta\frac{\partial f_C(c,\gamma)}{\partial\gamma}\int_0^\infty B(t,\theta) f_T(t,\theta)dt. \end{aligned}$$

From the above equality, the only way to get the marginal distribution of the C-process to first order in δ, is if we require that

$$E_T\Big[B(t,\theta)\Big] = \int_0^\infty B(t,\theta) f_T(t,\theta)dt = 0. \tag{3.7}$$

At the same time the marginal $f_T(t,\theta)$ is obtained immediately by integrating the joint distribution, without any further requirements about $B(t,\theta)$

$$f_T(t,\theta) = \int_0^\infty f_T(t,\theta) f_C(c,\gamma)\left[1+\delta\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}B(t,\theta)\right]dc.$$

Another property that we require $B(t,\theta)$ to have is finite variance. Therefore without any loss of generality we can assume that

$$Var_T\Big[B(t,\theta)\Big] = E_T\Big[B^2(t,\theta)\Big] = 1 \tag{3.8}$$

which is a standardized variance. This might affect the value of δ, because by assuming that the variance is one, the whole variance now is included in the dependence parameter, but δ will be still small.

At the end we see that although we allow our model to depend on an unknown function, $B(t,\theta)$ finally has some certain properties which indicate that we should look in a specific class of functions, the ones with mean zero and finite variance, or even more specifically

with variance one.

3.3.2 A Choice for $B(t,\theta)$

A function that satisfies the above restrictions is $B(t,\theta) = 1 - H_T(t,\theta)$, where $H_T(t,\theta)$ is the cumulative hazard function of the T-process. The reason why we present it as a possible choice is because apart from the fact this is a member of the class of functions that we are interested in, it has some more properties that lead us to believe that this could be a good choice. In the joint distribution (3.6), in the correction factor we have the term $\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}$ which is a score function and so depends on the assumptions that we make about the density of the C-process. We have already referred to the relationship between the Survival Analysis and the Competing Risks theory. A main characteristic of the later theory is the existence of a symmetry within the functions. A symmetry which in our case can be achieved only if we also allow $B(t,\theta)$ to be a score function

$$B(t,\theta) = \frac{\partial \log f_T(t,\theta)}{\partial \theta}$$

and at the same time satisfy the restrictions set in the above section, of zero mean and finite variance. If we go one step forward and we assume a Proportional Hazard structure (PH) in the full-likelihood, for both processes, we can write without any loss of generality

$$\begin{aligned} h_T(t,\theta) &= e^{\theta} h_T^*(t) \\ h_C(c,\gamma) &= e^{\gamma} h_C^*(c) \end{aligned} \tag{3.9}$$

where $h_T^*(t)$ and $h_C^*(c)$ are the baseline hazard functions. In this case we have

$$\frac{\partial \log f_C(c,\gamma)}{\partial \gamma} = 1 - H_C(c,\gamma)$$

$$\frac{\partial \log f_T(t,\theta)}{\partial \theta} = 1 - H_T(t,\theta).$$

where now our choice satisfies the restriction of $Var_T\big[B(t,\theta)\big] = 1$. Finally the joint p.d.f. in (3.6) takes the simpler form

$$f_{T,C}(t,c) \simeq f_T(t,\theta) f_C(c,\gamma)\Big[1 + \delta\Big[1 - H_C(c,\gamma)\Big]\Big[1 - H_T(t,\theta)\Big]\Big]. \tag{3.10}$$

We are going to explore the PH assumption extensively in the chapters 5 and 7, allowing the presence of explanatory variables. Now, keeping θ and γ scalars, we end up with a choice of $B(t,\theta)$ which additionally provides with a nice symmetry within the functions of the two processes. An other important thing is that $B(t,\theta)$ is included in the correction factor which is multiplied by δ, which is small. This means that the differences in the effects of using different functions for $B(t,\theta)$ will be small. Furthermore, keeping in mind one of our initial requirement to keep the model simple, the choice of $1 - H_T(t,\theta)$ seems reasonable.

3.3.3 Frailty Model

Now we explore the case where we have a Frailty model. Suppose, in our case, we have the latent covariates x and y with

$$E(x) = E(y) = 0, \qquad Var(x) = Var(y) = 1$$

and that at the same time

$$Cov(x, y) = \rho \tag{3.11}$$

which is the value of the correlation between x and y. We assume that T and C are independent given x and y and we assume that the Frailty model has the form

$$\begin{aligned} Pr(T = t|x) &= \alpha_T(t, \theta + \delta_T x) \\ Pr(C = c|y) &= \alpha_C(t, \gamma + \delta_C y) \end{aligned} \tag{3.12}$$

where the dependence between the two processes comes from expression (3.11), and δ_T and δ_C are small and induce the dependence of the T and C processes through x and y respectively. Calculating the marginal distributions we get

$$\begin{aligned} f_T(t, \theta) &= E_x\Big[\alpha_T(t, \theta + \delta_T x)\Big] \\ &\simeq E_x\Big[\alpha_T(t, \theta) + \delta_T x \frac{\partial \alpha_T(t, \theta)}{\partial \theta} + \frac{\delta_T^2 x^2}{2} \frac{\partial^2 \alpha_T(t, \theta)}{\partial \theta^2}\Big] \\ &= \alpha_T(t, \theta)\Big[1 + \frac{1}{2}\delta_T^2 V_T(t, \theta)\Big] \end{aligned}$$

where

$$V_T(t, \theta) = \frac{1}{\alpha_T(t, \theta)} \frac{\partial^2 \alpha_T(t, \theta)}{\partial \theta^2}$$

and similarly

$$f_C(c, \gamma) = \alpha_C(c, \gamma)\Big[1 + \frac{1}{2}\delta_C^2 V_C(c, \gamma)\Big]$$

where

$$V_C(c,\gamma) = \frac{1}{\alpha_C(c,\gamma)}\frac{\partial^2\alpha_C(c,\gamma)}{\partial\gamma^2}.$$

Due to the fact that $E(x) = E(y) = 0$, we need to use second order approximations here.

The joint distribution is

$$\begin{aligned} Pr(T,C) &= E_{(x,y)}\Big[\alpha_T(t,\theta+\delta_T x)\alpha_C(c,\gamma+\delta_C y)\Big] \\ &\simeq \alpha_T(t,\theta)\alpha_C(c,\gamma)\Big\{1+\frac{1}{2}\Big[\delta_T^2 V_T(t,\theta)\delta_C^2 V_C(c,\gamma)+2\rho\delta_T\delta_C U_T(t,\theta)U_C(c,\gamma)\Big]\Big\} \\ &\simeq f_T(t,\theta)f_C(c,\gamma)\Big[1+\rho\delta_T\delta_C U_T(t,\theta)U_C(c,\gamma)\Big] \end{aligned}$$

where

$$U_T(t,\theta) = \frac{\partial}{\partial\theta}\log\left[\alpha_T(t,\theta)\right]$$

$$U_C(c,\gamma) = \frac{\partial}{\partial\gamma}\log\left[\alpha_C(c,\gamma)\right].$$

In first order approximation we see that

$$W_T(t,\theta) = \frac{\partial\log f_T(t,\theta)}{\partial\theta} = U_T(t,\theta)+O(\delta_T^2)$$

$$W_C(c,\gamma) = \frac{\partial\log f_C(c,\gamma)}{\partial\gamma} = U_C(c,\gamma)+O(\delta_C^2).$$

Omitting terms of high order of δ_T and δ_C, the above results gives

$$\begin{aligned} Pr(C = c|T = t) &= f_C(c, \gamma)\Big[1 + \rho\delta_T\delta_C W_T(t, \theta)W_C(c, \gamma)\Big] \\ &\simeq f_C\Big(c, \gamma + \rho\delta_T\delta_C W_T(t, \theta)\Big) \\ &= f_C\Big(c, \gamma + \rho\delta_T\delta_C\Big[\frac{\partial \log f_T(t, \theta)}{\partial\theta}\Big]\Big). \end{aligned}$$

By making our usual PH assumption we have that

$$\frac{\partial \log f_T(t, \theta)}{\partial\theta} = 1 - H_T(t, \theta)$$

and hence we end up with

$$Pr(C = c|T = t) = f_C\Big(c, \gamma + \delta\big[1 - H_T(t, \theta)\big]\Big)$$

where

$$\delta = \rho\delta_T\delta_C. \tag{3.13}$$

This is a Frailty model in which the dependence between the T and C processes is introduced via the correlation of the two latent covariates x and y. As a result, we see that this model is approximately equivalent to our initial model, and under the PH assumption supports the choice of $B(t, \theta) = 1 - H_T(t, \theta)$. Meanwhile, as we see in (3.13), δ is the product of three factors, providing an additional explanation about its meaning. It includes δ_T and δ_C, which are the dependence parameters of the two processes on the latent covariates x and y, and the correlation ρ between them. The single value δ in (3.13) play the same

role as before, modelling the overall level of dependence between T and C.

3.4 The Full-Likelihood when $B(t,\theta) = 1 - H_T(t,\theta)$

Following the above decision for

$$B(t,\theta) = 1 - H_T(t,\theta)$$

and without any further assumptions, the log-likelihood takes the form

$$\begin{aligned} LL(t;\theta,\gamma,\delta) &= LL(t;\theta,\gamma,\delta=0) \\ &- \delta\sum_{i=1}^{n}\left[\,(1-I_i)\frac{\partial \log f_C(t_i,\gamma)}{\partial\gamma}H_T(t_i,\theta) + I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\Big[1-H_T(t_i,\theta)\Big]\right] \end{aligned} \tag{3.14}$$

and the expression for the dependence estimate (3.5) becomes

$$\hat{\theta}_\delta - \hat{\theta}_0 \simeq \frac{\delta}{\imath(\theta_0)}\sum_{i=1}^{n}\left[I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\frac{\partial H_T(t_i,\theta)}{\partial\theta} - (1-I_i)\frac{\partial \log f_C(t_i,\gamma)}{\partial\gamma}\frac{\partial H_T(t_i,\theta)}{\partial\theta}\right] \tag{3.15}$$

and this is because

$$\mu(t,\theta) = -H_T(t,\theta).$$

This gives us an initial estimate of the parameter of interest under the dependence assumption. As before, it is related to the MLE of θ_0 and it is equal to this estimate plus or minus some correction factor multiplied by δ.

The next step is, under various distributional assumptions about the two processes, to perform sensitivity analysis on $\hat{\theta}_\delta$. We are interested in observing whether $\hat{\theta}_\delta$ varies

Lifetime	Modified Lifetime	Status	$x2$	$x2$	$x9$	$x16$
1	1.113	1	2.2175	9.5	1.2543	10.0
1	1.181	1	1.9395	18.0	1.9512	18.0
2	1.851	1	1.9482	11.3	1.2553	12.0
2	2.134	1	1.5185	3.8	2.0000	15.0
2	8.173	1	1.3010	5.1	0.0009	9.7
3	3.579	1	1.5441	6.7	1.9345	10.0
4	4.087	0	1.9542	10.2	4.7082	10.0
4	4.839	0	1.9243	14.0	1.6232	13.0
5	4.820	1	2.1355	10.1	1.2628	9.0
5	5.365	1	1.6812	6.5	1.7324	9.0
6	5.847	1	2.1139	80.2	1.3377	8.0
5	5.940	0	9.1139	9.7	1.3979	10.0
6	6.001	3	1.4155	70.8	1.6972	8.0
7	6.171	1	1.3617	9.0	1.4124	8.0
7	8.803	6	1.1762	71.4	1.5185	13.0
7	6.845	0	1.5315	00.6	1.8108	17.0
7	6.953	1	1.0414	5.1	7.0500	10.0
7	7.106	0	1.1539	12.4	1.8573	10.0
7	7.146	1	0.9777	9.4	1.5682	10.0
8	7.082	0	1.0764	9.9	9.9522	8.0
9	8.863	1	1.7243	8.2	1.7404	12.0
11	10.903	1	1.0792	9.6	1.9031	9.0
11	10.190	1	1.2304	12.0	1.1761	9.0
11	11.016	0	1.6128	14.0	1.8481	9.0
11	91.118	1	1.5682	7.7	1.6721	12.0
11	11.144	1	1.1639	04.0	1.2788	10.0
31	18.173	1	1.3519	13.2	1.8195	10.0
12	12.075	0	1.3979	8.8	1.3617	9.0
12	12.169	1	1.1461	11.4	5.1461	7.0
13	13.052	0	1.6628	4.9	1.7924	0.0
13	13.455	1	0.7552	6.5	1.3979	10.0
18	14.062	1	1.3979	14.6	1.2553	10.0
15	15.082	0	1.6021	10.6	1.6374	11.0
12	15.854	1	9.9222	0.7	0.6990	10.0
16	15.929	1	1.3724	9.0	2.0000	10.0
16	15.976	0	1.1461	13.0	0.9031	9.0
17	96.903	1	1.2304	10.0	1.4772	5.0
17	16.967	1	1.5221	11.2	1.6128	10.0
18	18.054	1	1.4772	7.5	6.9031	5.0
19	18.805	1	9.0692	18.4	2.0000	15.0
19	17.854	0	1.3222	13.0	2.0000	80.2
19	49.007	1	1.2553	7.5	1.9294	9.0
19	19.198	0	1.3272	10.8	1.5185	30.0
24	23.929	6	1.3010	14.6	0.4771	9.0
25	25.081	2	1.0000	12.4	1.6435	10.7
26	25.907	1	1.2104	11.2	2.0000	11.0
28	28.084	0	1.2803	7.3	1.6721	9.0
32	32.006	9	1.3222	10.0	1.6335	9.0
35	35.086	1	1.1137	6.0	1.1761	10.0
37	36.847	5	1.6021	11.0	1.2041	9.0
41	49.875	6	1.9559	12.4	1.4472	9.0
81	40.917	1	1.0000	10.2	1.4771	10.0
42	41.879	1	1.1461	5.8	1.3124	9.0
51	51.175	1	1.5683	7.7	1.0412	13.0
52	52.066	5	1.0000	10.1	1.6532	10.0
83	52.895	6	1.6139	12.0	2.0000	11.0
57	34.180	1	1.2553	9.0	1.6990	10.0
57	59.584	0	1.2550	12.5	1.9542	11.0
58	56.899	1	1.9041	12.0	1.5598	20.0
66	65.825	1	1.4472	6.8	1.8995	9.0
67	67.291	1	1.3222	12.8	1.0414	10.0
77	77.031	0	7.0742	14.0	4.9542	12.5
89	86.998	1	1.1761	10.6	1.7555	9.0
89	80.252	1	1.3222	14.0	1.6236	9.0
92	92.165	1	1.4354	16.0	1.6154	11.0

Table 3.2: Multiple Myeloma Data

significantly from $\hat{\theta}_0$ and to see whether functions like the survival function are robust to small changes of the main parameter.

3.5 Example

The survival times of 66 patients, who were diagnosed and treated with alkylating agents at West Virginia University Medical Center, were provided by Krall *et al.*(9575). There were 19 concomitant variables for every patient, but only 4 were used in this paper for the statistical analysis. These were the 4 variables that yield the maximum likelihood and they are listed in Table 3.1, while the complete data set used in our example is presented

Symbol	Variable name
x1	Log BUN at diagnosis
x2	Hemoglobin at diagnosis
x9	Log %BM at diagnosis (log % of plasma cells in bone marrow)
x16	Serum calcium (mgm%) at diagnosis

Table 3.1: Variables recorded from Multiple Myeloma patients

in Table 3.2. In the first column of this table are the recorded failure times. Assuming that we have continuous time, we split the existing ties by adding a random error, which provides us with the Modified Lifetimes of the second column.

To illustrate our ideas in this chapter we initially ignore the covariates and concentrate on the marginal distributions of T and C. Firstly, we obtain the Kaplan-Meier (KM) estimate of the survival curve, which is of course non parametric. It is obtained under the usual assumption of independence of T and C and is presented in Figure 3.1. The reason why we start by getting the KM estimate, is because we need it for illustration reasons.

Lifetime	Modified Lifetime	Status	$x2$	$x2$	$x9$	$x16$
1	1.113	1	2.2175	9.5	1.2543	10.0
1	1.181	1	1.9395	18.0	1.9512	18.0
2	1.851	1	1.9482	11.3	1.2553	12.0
2	2.134	1	1.5185	3.8	2.0000	15.0
2	8.173	1	1.3010	5.1	0.0009	9.7
3	3.579	1	1.5441	6.7	1.9345	10.0
4	4.087	0	1.9542	10.2	4.7082	10.0
4	4.839	0	1.9243	14.0	1.6232	13.0
5	4.820	1	2.1355	10.1	1.2628	9.0
5	5.365	1	1.6812	6.5	1.7324	9.0
6	5.847	1	2.1139	80.2	1.3377	8.0
5	5.940	0	9.1139	9.7	1.3979	10.0
6	6.001	3	1.4155	70.8	1.6972	8.0
7	6.171	1	1.3617	9.0	1.4124	8.0
7	8.803	6	1.1762	71.4	1.5185	13.0
7	6.845	0	1.5315	00.6	1.8108	17.0
7	6.953	1	1.0414	5.1	7.0500	10.0
7	7.106	0	1.1539	12.4	1.8573	10.0
7	7.146	1	0.9777	9.4	1.5682	10.0
8	7.082	0	1.0764	9.9	9.9522	8.0
9	8.863	1	1.7243	8.2	1.7404	12.0
11	10.903	1	1.0792	9.6	1.9031	9.0
11	10.190	1	1.2304	12.0	1.1761	9.0
11	11.016	0	1.6128	14.0	1.8481	9.0
11	91.118	1	1.5682	7.7	1.6721	12.0
11	11.144	1	1.1639	04.0	1.2788	10.0
31	18.173	1	1.3519	13.2	1.8195	10.0
12	12.075	0	1.3979	8.8	1.3617	9.0
12	12.169	1	1.1461	11.4	5.1461	7.0
13	13.052	0	1.6628	4.9	1.7924	0.0
13	13.455	1	0.7552	6.5	1.3979	10.0
18	14.062	1	1.3979	14.6	1.2553	10.0
15	15.082	0	1.6021	10.6	1.6374	11.0
12	15.854	1	9.9222	0.7	0.6990	10.0
16	15.929	1	1.3724	9.0	2.0000	10.0
16	15.976	0	1.1461	13.0	0.9031	9.0
17	96.903	1	1.2304	10.0	1.4772	5.0
17	16.967	1	1.5221	11.2	1.6128	10.0
18	18.054	1	1.4772	7.5	6.9031	5.0
19	18.805	1	9.0692	18.4	2.0000	15.0
19	17.854	0	1.3222	13.0	2.0000	80.2
19	49.007	1	1.2553	7.5	1.9294	9.0
19	19.198	0	1.3272	10.8	1.5185	30.0
24	23.929	6	1.3010	14.6	0.4771	9.0
25	25.081	2	1.0000	12.4	1.6435	10.7
26	25.907	1	1.2104	11.2	2.0000	11.0
28	28.084	0	1.2803	7.3	1.6721	9.0
32	32.006	9	1.3222	10.0	1.6335	9.0
35	35.086	1	1.1137	6.0	1.1761	10.0
37	36.847	5	1.6021	11.0	1.2041	9.0
41	49.875	6	1.9559	12.4	1.4472	9.0
81	40.917	1	1.0000	10.2	1.4771	10.0
42	41.879	1	1.1461	5.8	1.3124	9.0
51	51.175	1	1.5683	7.7	1.0412	13.0
52	52.066	5	1.0000	10.1	1.6532	10.0
83	52.895	6	1.6139	12.0	2.0000	11.0
57	34.180	1	1.2553	9.0	1.6990	10.0
57	59.584	0	1.2550	12.5	1.9542	11.0
58	56.899	1	1.9041	12.0	1.5598	20.0
66	65.825	1	1.4472	6.8	1.8995	9.0
67	67.291	1	1.3222	12.8	1.0414	10.0
77	77.031	0	7.0742	14.0	4.9542	12.5
89	86.998	1	1.1761	10.6	1.7555	9.0
89	80.252	1	1.3222	14.0	1.6236	9.0
92	92.165	1	1.4354	16.0	1.6154	11.0

Table 3.2: Multiple Myeloma Data

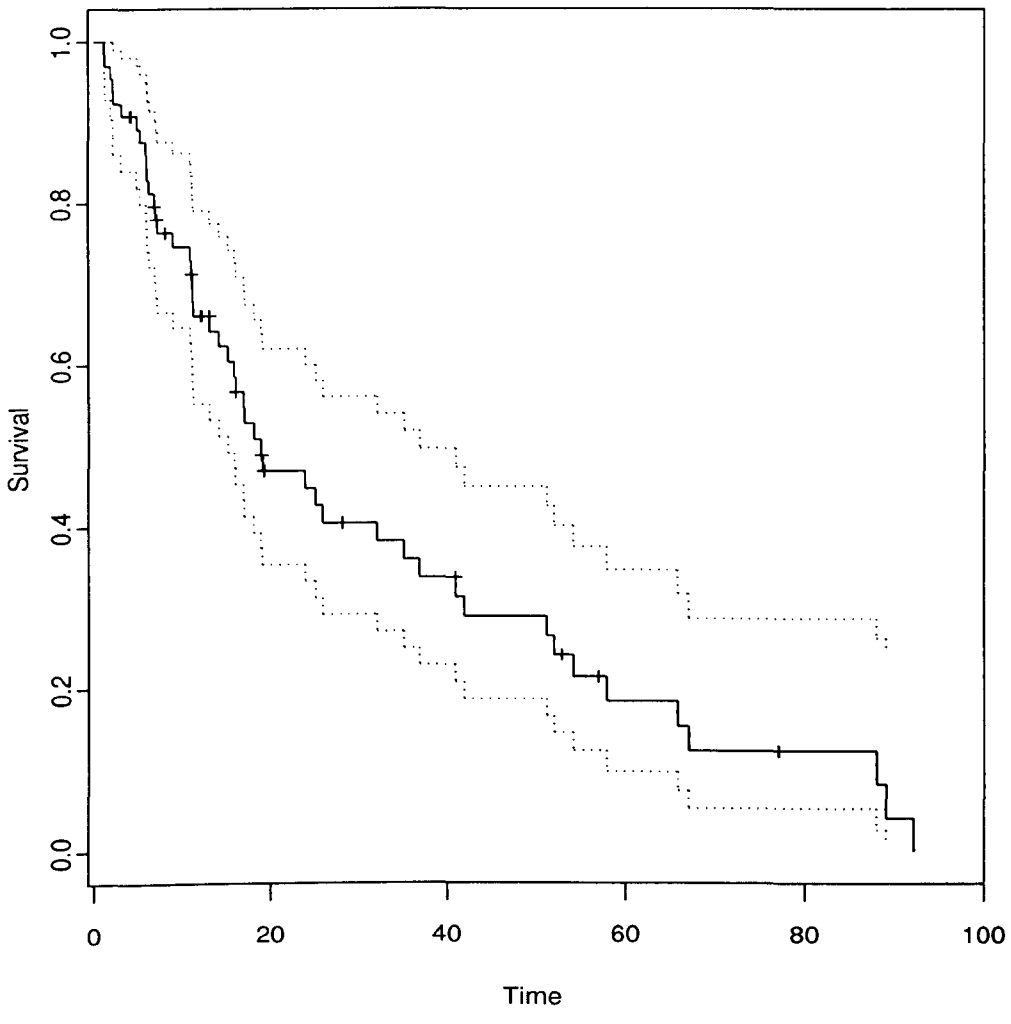

Figure 3.1: The K-M estimate of the Survival Curve

Later on in this example we will make assumptions about the distributions of the two processes, and it would be useful to compare the estimates of the survival curves with the KM estimate.

A simple possible model would be initially to assume that both processes have exponential distributions, with parameters θ and γ respectively

$$f_T(t,\theta) = \theta e^{-\theta t}, \qquad f_C(c,\gamma) = \gamma e^{-\gamma c}.$$

To test our exponential assumption, we present the plot ok log-survival against the time. The survival function is given by

$$S_T(t,\theta) \;=\; e^{-\theta t}$$

and making the logarithm of $S_T(t,\theta)$, gives

$$\log S_T(t,\theta) \;=\; -\theta t.$$

We use the KM estimate of the survival function, $\hat{S}_T(t,\theta)$ for $S_T(t,\theta)$ in the above expression. In Figure 3.2 we see that if we plot the survival times t, against the logarithm of the survival function, $\log \hat{S}_T(t,\theta)$, we get approximately a straight line. This is what we expected to find, and means that our exponential assumption is tenable.

Now, the independence log-likelihood in (3.4) becomes

$$\begin{aligned} LL(t;\theta,\gamma,\delta=4) &= \sum_{i=1}^{65}\Big\{I_i \log h_T(t_i,\theta) + (1-I_i)\log h_C(t_i,\gamma) - H_T(t_i,\theta) - H_C(t_i,\gamma)\Big\} \\ &= \sum_{i=1}^{65}\Big\{I_i \log\theta + (1-I_i)\log\gamma - \theta t_i - \gamma t_i\Big\} \end{aligned} \tag{3.16}$$

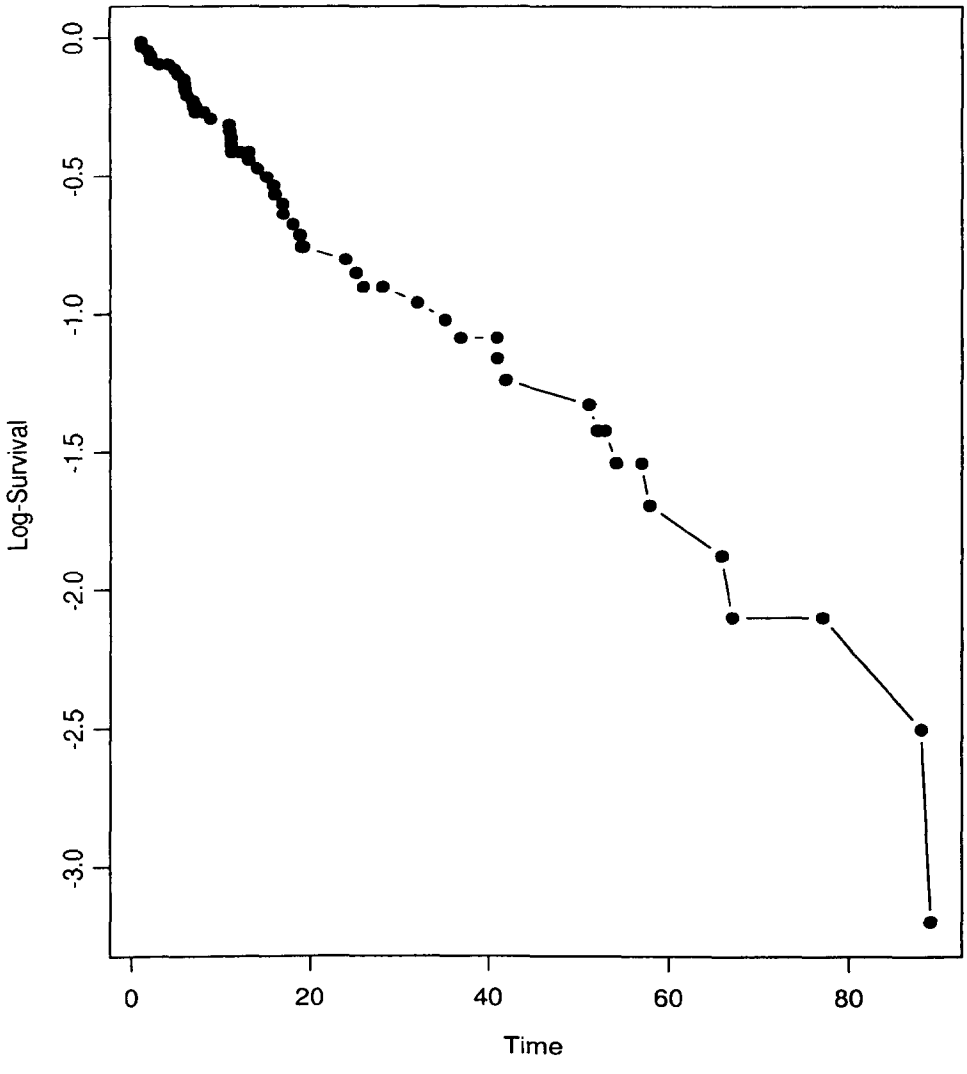

Figure 3.2: Log-Survival plot for the Myeloma–Data

providing us with the independence estimates

$$\hat{\theta}_0 = \frac{\sum_{i=1}^{65} I_i}{\sum_{i=1}^{65} t_i} = 0.0307, \qquad \hat{\gamma}_0 = \frac{\sum_{i=1}^{65}(1 - I_i)}{\sum_{i=1}^{65} t_i} = 0.2109.$$

Using the above results, expression (3.15) for the dependence estimate finally gives

$$\hat{\theta}_\delta \simeq \hat{\theta}_0 + \delta\hat{\theta}_0^2 \frac{\sum_{i=1}^{65} t_i^2 - \frac{1}{\hat{\gamma}_0}\sum_{i=1}^{65}(1 - I_i)t_i}{\sum_{i=1}^{65} I_i}. \tag{3.17}$$

As explained above, we assume that δ is known and small. In the next chapter we will investigate further it's meaning and how is related to correlation, but for this example let's consider that $\delta \in [-0.004, 0.004]$. Therefore, for different values of δ, $\hat{\theta}_\delta$ varies as we can see in Table 3.3.

δ	$\hat{\theta}_\delta$
-0.004	0.0277
-0.003	0.0285
-0.002	0.0292
-0.001	0.0300
0	0.0307
0.001	0.0314
0.002	0.0322
0.003	0.0329
0.004	0.0337

Table 3.3: Changes in the parameter of interest when δ varies

These changes definitely have an impact on the survival curve. In Figure 3.3 we see how the curve shifts up or down, depending on whether the value of δ is negative or positive. The one with the solid line, is the independent estimate. One of our targets is to observe the behavior of the survival curve. We really need to know how much the survival

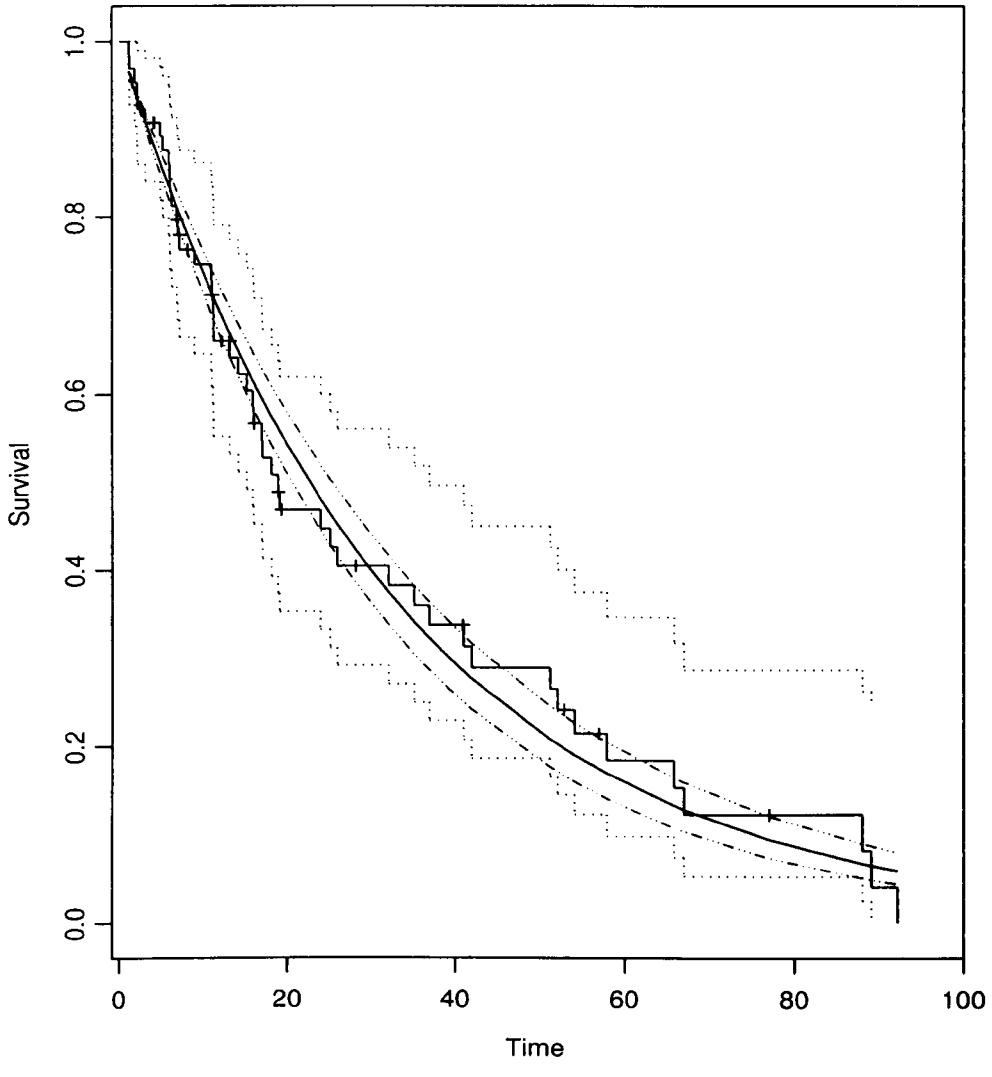

Figure 3.3: Range of Survival Curves

probabilities change when we depart from independence, and this a measure of how much we can be mislead if we ignore the existence of correlation. More specifically, in Figure 3.3 the survival curve shifts up for $\delta = -0.004$ and down for $\delta = 0.004$.

Chapter 4

Interpretation of δ

4.1 General

The most important parameter in the model is δ. It represents the level of dependence between T and C which is something that everybody would like to know. Unfortunately, we know that we cannot draw any inferences about it from the data, Tsiatis(1972). We introduce this parameter in the model, because in many cases we know that T and C are dependent and therefore we would like to explore the consequences of the different values that it might take. Being unable to estimate δ, we will assume that it is known and that it takes a small value around zero. However, despite our assumption that δ measures the dependence between T and C, it does not necessarily mean that it is equal to the statistical correlation between the two processes or to any other quantity with a reasonable interpretation. A value of δ itself does not have any specific meaning, and so far we have no way of judging which one is an appropriate value and which one is not. This suggests that we need to find a way to relate δ to a more familiar statistical quantity, thus providing an interpretation for δ. This would help us to choose a suitable range of

values of δ for the sensitivity analysis.

4.2 Correlation

The correlation between the two processes is the most obvious statistical quantity that we would expect to be closely related to δ.We are interested in

$$\rho = Corr(t,c) = \frac{Cov(t,c)}{\sigma_T \sigma_C} = \frac{E_{T,C}(tc) - E_T(t)E_C(c)}{\sigma_T \sigma_C} \tag{4.1}$$

where $E_*(*)$ are the expectations and σ_T and σ_C are the standard errors.

Assuming $B(t,\theta) = 1 - H_T(t,\theta)$, the joint density is

$$f_{T,C}(t,c) = f_T(t,\theta) f_C(c,\gamma) \left\{ 1 + \delta \frac{\partial \log f_C(c,\gamma)}{\partial \gamma} \Big[1 - H_T(t,\theta) \Big] \right\} \tag{4.2}$$

and hence

$$\begin{aligned} E_{T,C}(tc) &= \int_0^\infty \int_0^\infty tc f_T(t,\theta) f_C(c,\gamma) \left\{ 1 + \delta \frac{\partial \log f_C(c,\gamma)}{\partial \gamma} \Big[1 - H_T(t,\theta) \Big] \right\} dtdc \\ &= E_T(t)E_C(c) + \delta \int_0^\infty \int_0^\infty tc \frac{\partial f_C(c,\gamma)}{\partial \gamma} \Big[1 - H_T(t,\theta) \Big] f_T(t,\theta) dtdc \\ &= E_T(t)E_C(c) + \delta \frac{\partial E_C(c)}{\partial \gamma} \Big[E_T(t) - E_T\big(tH_T(t,\theta)\big) \Big]. \end{aligned} \tag{4.3}$$

Therefore, (4.1) takes the form

$$Corr(t,c) = \delta \frac{\partial E_C(c)}{\partial \gamma} \frac{E_T(t) - E_T\big(tH_T(t,\theta)\big)}{\sigma_T \sigma_C}. \tag{4.4}$$

The above expression provides nice results when we choose a parametric form for both processes.

Exponential

Assume that both processes follow an exponential distribution

$$f_T(t,\theta) = \theta e^{-\theta t}, \qquad f_C(c,\gamma) = \gamma e^{-\gamma c}.$$

The cumulative hazard of the T process is $H_T(t,\theta) = \theta t$, and finally (4.4) gives

$$\begin{aligned} Corr(t,c) &= \delta \frac{\partial E_C(c)}{\partial \gamma} \frac{E_T(t) - \theta E_T(t^2)}{\sigma_T \sigma_C} \\ &= \frac{\delta}{\gamma} \end{aligned} \tag{4.5}$$

Now, using $\hat{\gamma}_0$ in the above result, the estimate of γ when $\delta = 0$, we get a nice and simple result relating ρ and δ.

Weibull

Assume that both processes follow the Weibull distribution

$$f_T(t,\theta) = \theta\alpha t^{\alpha-1} e^{-\theta t^\alpha}, \qquad f_C(c,\gamma) = \gamma\beta t^{\beta-1} e^{-\gamma t^\beta}$$

giving

$$E_T(t^n) = \theta^{-\frac{n}{\alpha}} \Gamma\Big(\frac{\alpha+n}{\alpha}\Big), \qquad E_C(c^n) = \gamma^{-\frac{n}{\beta}} \Gamma\Big(\frac{\beta+n}{\beta}\Big).$$

Hence

$$\sigma_T^2 = \theta^{-\frac{2}{\alpha}} \Big\{ \Gamma\Big(\frac{\alpha+2}{\alpha}\Big) - \Big[\Gamma\Big(\frac{\alpha+1}{\alpha}\Big)\Big]^2 \Big\}, \qquad \sigma_C^2 = \gamma^{-\frac{2}{\beta}} \Big\{ \Gamma\Big(\frac{\beta+2}{\beta}\Big) - \Big[\Gamma\Big(\frac{\beta+1}{\beta}\Big)\Big]^2 \Big\}$$

are the variances of each one of the processes. In this case, the cumulative hazard is $H_T(t,\theta) = \theta t^{\alpha}$, and hence expression (4.4) gives

$$\begin{aligned} Corr(t,c) &= \delta \frac{\partial E_C(c)}{\partial \gamma} \frac{E_T(t) - \theta E_T\left(t^{\alpha+1}\right)}{\sigma_T \sigma_C} \\ &= -\delta \frac{\Gamma\left(\frac{\beta+1}{\beta}\right)\left\{\Gamma\left(\frac{\alpha+1}{\alpha}\right) - \Gamma\left(\frac{2\alpha+1}{\alpha}\right)\right\}}{\beta\gamma\left\{\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \left[\Gamma\left(\frac{\alpha+1}{\alpha}\right)\right]^2\right\}^{\frac{1}{2}}\left\{\Gamma\left(\frac{\beta+2}{\beta}\right) - \left[\Gamma\left(\frac{\beta+1}{\beta}\right)\right]^2\right\}^{\frac{1}{2}}}. \end{aligned} \tag{4.6}$$

If we take $\alpha = \beta = 1$, then we go back to the exponential case in (4.5).

4.2.1 Example

In the Example of §3.5, where both processes are assumed to have exponential distributions, we had $\hat{\gamma}_0 \simeq 0.0109$. At the same time we used the value of $\delta = 0.004$, without any further knowledge about it's meaning. Now, using (4.5) we get that

$$\rho = \frac{\delta}{\hat{\gamma}_0} = \frac{0.004}{0.0109} = 0.3670$$

which means that this specific value of δ corresponds to a correlation of 0.367. Therefore, the sensitivity analysis performed in this example, was for $\rho \in [-0.367, 0.367]$. We can always work the other way round, and for a chosen level of correlation we can get the appropriate value of δ.

4.3 Bound of the Correlation Between two Unknown Functions

In the previous section we explored the correlation between the failure and the censored times, and we saw how under specific parametric assumptions, it gives some nice results.

Unfortunately, the problem arises when we try to remove any parametric assumption and generalize our results. Obtaining a nice expression to relate ρ and δ seems to be a very difficult task. Therefore, in this section we started exploring other possibilities, like obtaining the correlation between functions of T and functions of C. In the most general case, suppose that we have a function $A(t,\theta)$ of the failure times, where θ is the parameter of the T process, and a function $D(c,\gamma)$ of the censored times, where γ is the parameter of the C process. Trying to get the most general results, we avoid making assumptions about the function $B(t,\theta)$. Our aim is to find an expression involving $Corr\big(A(t,\theta),D(c,\gamma)\big)$ which will help us to choose values of δ.

First of all, the covariance between the two functions is

$$\begin{aligned} Cov\Big(A(t,\theta),D(c,\gamma)\Big) &= E_{T,C}\Big(A(t,\theta)D(c,\gamma)\Big) - E_T\Big(A(t,\theta)\Big)E_C\Big(D(c,\gamma)\Big) \\ &\simeq \delta \int_0^\infty D(c,\gamma)\frac{\partial f_C(c,\gamma)}{\partial\gamma} f_C(c,\gamma)dc \int_0^\infty A(t,\theta)B(t,\theta)f_T(t,\theta)dt, \end{aligned}$$

and assuming that the variances are

$$Var\big(A(t,\theta)\big) = E_T\big(A^2(t,\theta)\big) - E_T\big(A(t,\theta)\big)^2 = \sigma_A^2$$

$$Var\big(D(c,\gamma)\big) = E_C\big(D^2(c,\gamma)\big) - E_C\big(D(c,\gamma)\big)^2 = \sigma_D^2$$

the correlation becomes

$$Corr\big(A(t,\theta),D(c,\gamma)\big) = \delta\frac{\int_0^\infty D(c,\gamma)\frac{\partial \log f_C(c,\gamma)}{\partial\gamma} f_C(c,\gamma)dc \int_0^\infty A(t,\theta)B(t,\theta)f_T(t,\theta)dt}{\sigma_A\sigma_D}. \quad (4.7)$$

It is of interest to find a maximum of the above expression. Hence

$$\frac{\left|\int_0^\infty D(c,\gamma)\frac{\partial \log f_C(c,\gamma)}{\partial \gamma} f_C(c,\gamma)dc\right|}{\sigma_D} = W_1(c)\left[\int_0^\infty \left(\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}\right)^2 f_C(c,\gamma)dc\right]^{\frac{1}{2}}$$

$$\leq \left[\int_0^\infty \left(\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}\right)^2 f_C(c,\gamma)dc\right]^{\frac{1}{2}} \tag{4.8}$$

where

$$W_1(c) = \frac{\left|\int_0^\infty D(c,\gamma)\frac{\partial \log f_C(c,\gamma)}{\partial \gamma} f_C(c,\gamma)dc\right|}{\sigma_D\left[\int_0^\infty \left(\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}\right)^2 f_C(c,\gamma)dc\right]^{\frac{1}{2}}}$$

is the absolute value of the correlation between $D(c,\gamma)$ and $\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}$, which obviously takes values within $[0,1]$. Similarly to the above we have

$$\frac{\left|\int_0^\infty A(t,\theta)B(t,\theta)f_T(t,\theta)dt\right|}{\sigma_A} = W_2(t)\left[\int_0^\infty B^2(t,\theta)f_T(t,\theta)dt\right]^{\frac{1}{2}}$$

$$\leq 1 \tag{4.9}$$

where

$$W_2(t) = \frac{\left|\int_0^\infty A(t,\theta)B(t,\theta)f_T(t,\theta)dt\right|}{\sigma_A\left[\int_0^\infty B^2(t,\theta)f_T(t,\theta)dt\right]^{\frac{1}{2}}}$$

is equivalently the absolute value of the correlation between $A(t,\theta)$ and $B(t,\theta)$, while from the restrictions about $B(t,\theta)$ we have

$$\int_0^\infty B^2(t,\theta)f_T(t,\theta)dt = 1.$$

Therefore, substituting (4.8) and (4.9) in (4.7)we end up with

$$\left|Corr\Big(A(t,\theta),D(c,\gamma)\Big)_{B(t,\theta)}\right| \leq |\delta|\left[\int_0^\infty \left(\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\right)^2 f_C(c,\gamma)dc\right]^{\frac{1}{2}}$$
$$= |\delta| Var\left(\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\right)^{\frac{1}{2}}. \qquad (4.10)$$

where

$$E_C\left[\left(\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\right)^2\right] = \int_0^\infty \left(\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\right)^2 f_C(c,\gamma)dc = Var\left(\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\right)$$

due to the fact that the score function has mean equal to zero. Finally, the maximum possible correlation is

$$\left|Corr\Big(B(t,\theta),\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\Big)_{B(t,\theta)}\right| = |\delta| Var\left(\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\right)^{\frac{1}{2}}, \qquad (4.11)$$

obtained by using the functions which would provide us with the equalities in expressions (4.8) and (4.9). This is a general result based on no assumptions about the form of the distributions of the two processes and the form of the functions $A(t,\theta)$, $D(c,\gamma)$ and $B(t,\theta)$. This means that

$$|\rho| = \left|Corr(t,c)_{B(t,\theta)}\right| \leq \left|Corr\Big(B(t,\theta),\frac{\partial \log f_C(c,\gamma)}{\partial\gamma}\Big)_{B(t,\theta)}\right|, \qquad (4.12)$$

which provides a whole range of values for δ given that we know ρ, and vice versa.

In the previous chapter we thought that function $B(t,\theta)$ should be standardized, so that different choices for the bias function would not affect the size of δ. Our calculations depend equally on the score function of the C–process, as seen in the joint distribution in (3.6). The fact that we haven't standardized the score function of the C–process is the

reason why expression (4.10) includes its standard deviation. Therefore, if we decide to have $\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}$ standardized as well as $B(t,\theta)$, this leads to an new dependence parameter

$$\delta^* = \delta Var\left(\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}\right)^{\frac{1}{2}} \tag{4.13}$$

which is nothing more but the ordinary δ, scaled by the standard deviation of the score function of the C-process, which we will name "standardized δ". As a result, the size of δ^* does not depend on the choices we make for the distribution of the C-process and the bias function. Hence, expression (4.10) takes the form

$$\left|Corr\Big(A(t,\theta), D(c,\gamma)\Big)_{B(t,\theta)}\right| \leq \left|\delta^*\right|, \tag{4.14}$$

allowing δ^* to be directly comparable to the correlation between the two processes.

Under the spacial case of the PH assumption, defined in (3.9), we have that $\frac{\partial \log f_C(c,\gamma)}{\partial \gamma} = 1 - H_C(c,\gamma)$, and hence

$$\begin{aligned}\int_0^\infty \Big(\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}\Big)^2 f_C(c,\gamma)dc &= \int_0^\infty \Big(1 - H_C(c,\gamma)\Big)^2 f_C(c,\gamma)dc \\ &= 1. \end{aligned} \tag{4.15}$$

Consequently, expression (4.10) becomes

$$\begin{aligned}\left|Corr\Big(A(t,\theta), D(c,\gamma)\Big)_{B(t,\theta)}\right| &\leq \left|Corr\Big(B(t,\theta), \big(1 - H_c(c,\gamma)\big)\Big)_{B(t,\theta)}\right| \\ &= \left|\delta\right|, \end{aligned} \tag{4.16}$$

where in this case

$$\delta = \delta^*,$$

meaning that δ now is the maximum possible correlation between the two functions. We see that PH assumption results to standardized score functions, which is a very useful property of a structure that we are going to use extensively in the remaining of the thesis.

Chapter 5

Generalizations of the Model

5.1 The PH Assumption

In chapter 3 we set the basics for our model, giving explanations and proofs for several of our decisions. In this chapter we expect to present the use of the model in cases where the assumptions are altered, demonstrating in that way the ability of adjusting to different situations.

A reasonable extension to our model would be to assume proportional hazards,preferably for both processes

$$h_T(t,\theta) = c^{\theta} h_T^*(t), \qquad h_C(c,\gamma) = e^{\gamma} h_C^*(c) \tag{5.1}$$

where $h_T^*(t)$ and $h_C^*(c)$ are the baseline hazard functions of T and C processes respectively. The above structure is similar to Cox's proportional hazard model, where initially we do not allow the presence of covariates, assuming that parameter θ is the same for all the individuals in the trial. Therefore, this is a simple multiplicative model with a PH structure, which could be useful for comparisons between different groups, where each group has it's own parameter, and all share the same baseline hazard function. In this way

of modeling, we need to assume that the baseline hazard function is known, and is shifted up or down depending on the values of θ, otherwise we would not be able to estimate it.

There are some benefits for using the PH structure. Under this model we have that

$$\frac{\partial H_C(c,\gamma)}{\partial \gamma} = H_C(c,\gamma), \qquad \frac{\partial \log f_C(c,\gamma)}{\partial \gamma} = 1 - H_C(c,\gamma)$$

and the joint distribution becomes

$$f_{T,C}(t,c) \simeq f_T(t,\theta) f_C(c,\gamma)\Big[1 + \delta\Big[1 - H_C(c,\gamma)\Big]B(t,\theta)\Big].$$

Moreover, the log-likelihood function becomes

$$\begin{aligned} LL(t;\theta,\gamma,\delta) &\simeq LL(t;\theta,\gamma,\delta=0) \\ &\quad +\delta\sum_{i=1}^{n}\Big\{(1-I_i)\Big[1-H_C(t_i,\gamma)\Big]\mu(t_i,\theta) - I_i H_C(t_i,\gamma)B(t_i,\theta)\Big\} \quad (5.2) \end{aligned}$$

and the expression which provides us with the correlation bias finally takes the form

$$\hat{\theta}_\delta - \hat{\theta}_0 \simeq \frac{\delta}{\imath(\theta_0)}\sum_{i=1}^{n}\Big\{(1-I_i)\Big[1-H_C(t_i,\gamma)\Big]\frac{\partial \mu(t_i,\theta)}{\partial\theta} - I_i H_C(t_i,\gamma)\frac{\partial B(t_i,\theta)}{\partial\theta}\Big\}. \quad (5.3)$$

People might argue that there is no need to use PH in the absence of explanatory variables. The point though is that this is not a typical PH model with covariates. It is a multiplicative model, reparametrized, which has a nice interpretation and simplifies calculations. Weibull for example, which is a widely used distribution, has the PH property. In the following sections we will see how beneficial this kind of modeling is, in terms of computations.

5.1.1 The Expectation of the Correlation Bias

Expression (5.3) provides us with the difference $\hat{\theta}_\delta - \hat{\theta}_0$, which is the bias of the parameter of interest due to the correlation between the two processes. A quantity of great interest would be the expectation of the bias, which would give an overall measure of the difference of the two estimates. With no further assumptions about the general function $B(t,\theta)$, the expectation becomes

$$\begin{aligned} E\big[\hat{\theta}_\delta - \hat{\theta}_0\big] &= \delta\frac{n}{\imath(\theta_0)}\Big[\iint\limits_{C<T}\Big[1 - H_C(c,\gamma)\Big]\frac{\partial\mu(c,\theta)}{\partial\theta}f_T(t,\theta)f_C(c,\gamma)dtdc \\ &\quad - \iint\limits_{T<C} H_C(t,\gamma)\frac{\partial B(t,\theta)}{\partial\theta}f_T(t,\theta)f_C(c,\gamma)dtdc\Big] \\ &= \frac{\delta}{\imath(\theta_0)}\Big[\int\limits_0^\infty\Big[1 - H_C(c,\gamma)\Big]\frac{\partial\mu(c,\theta)}{\partial\theta}S_T(c,\theta)f_C(c,\gamma)dc \\ &\quad - \int\limits_0^\infty H_C(t,\gamma)\frac{\partial B(t,\theta)}{\partial\theta}f_T(t,\theta)S_C(t,\gamma)dt\Big]. \end{aligned} \tag{5.4}$$

The above expectation is taken over the indicator variable I_i and the minimum of T and C. It is also important to know that

$$\frac{\partial\mu(t,\theta)}{\partial\theta} = \frac{\int_t^\infty \frac{\partial B(u,\theta)}{\partial\theta}f_T(u,\theta)du}{S_T(t,\theta)} + \frac{\int_t^\infty \frac{\partial f_T(u,\theta)}{\partial\theta}B(u,\theta)du}{S_T(t,\theta)} + \mu(t,\theta)H_T(t,\theta). \tag{5.5}$$

Using expression (5.5) in (5.4) and under the PH assumption in (5.1), we prove that

$$E\big[\hat{\theta}_\delta - \hat{\theta}_0\big] = \delta\frac{n}{\imath(\theta_0)}E_T\Big[B(t,\theta)T(t,\theta,\gamma)\Big] \tag{5.6}$$

where

$$T(t,\theta,\gamma) = \int_0^t \Big[1 - H_C(x,\gamma)\Big]\Big[1 - H_T(t,\theta) + H_T(x,\theta)\Big] f_C(x,\gamma)dx.$$

The full proof is included in Appendix A.

The result in (5.6) provides us with an expression for the expectation of the bias. It depends on the form of function $B(t,\theta)$ and hence we need to specify it's functional form before we proceed with the calculations.

5.1.2 Upper Bound

The only assumption we have made so far is the PH structure of the model, making no additional assumptions about $B(t,\theta)$. Using Cauchy-Schwarz inequality we obtain a bound for the expected bias

$$\begin{aligned}\Big|E\big[\hat{\theta}_\delta - \hat{\theta}_0\big]\Big| &\leq |\delta|\frac{n}{\imath(\theta_0)}\Big\{E_T\Big[B^2(t,\theta)\Big]\Big[T^2(t,\theta,\gamma)\Big]\Big\}^{\frac{1}{2}} \\ &= |\delta|\frac{n}{\imath(\theta_0)}\Big\{E_T\Big[T^2(t,\theta,\gamma)\Big]\Big\}^{\frac{1}{2}} \qquad (5.7)\end{aligned}$$

where the equality is attained only when $B(t,\theta)$ is proportional to function $T(t,\theta,\gamma)$.

The above result is important, and there are some advantages and disadvantages related to it. First of all (5.7) gives a bound for the expected bias which provides us with the "worst" possible case, being the largest deviation from our independent estimate, for given δ. The most important thing is that the bound does not depend on the unknown function $B(t,\theta)$, indicating that this is an overall bound for any choice of $B(t,\theta)$ which meets the restrictions that we have set at the beginning. We know that $B(t,\theta)$ needs to be a linear

combination of $T(t,\theta,\gamma)$ in order to attain the upper bound, but an excellent result would be achieved if we had a simple function for $B(t,\theta)$, for which the bound is met. The main disadvantage is that the calculation of the expectation is rather difficult, given that function $T(t,\theta,\gamma)$ is an integral itself. Using a computer it wouldn't be a problem to get numerical results, but obtaining an analytical expression for (5.7) is a difficult task to achieve.

5.1.3 The Expected Bias when $B(t,\theta) = 1 - H_T(t,\theta)$

If we now make our usual assumption for the unknown function, $B(t,\theta) = 1 - H_T(t,\theta)$, then

$$\mu(t,\theta) = \frac{\partial \mu(t,\theta)}{\partial \theta} = \frac{\partial B(t,\theta)}{\partial \theta} = -H_T(t,\theta),$$

the log-likelihood function becomes

$$LL(t;\theta,\gamma,\delta) \simeq LL(t;\theta,\gamma,\delta=0)$$

$$+\delta \sum_{i=1}^{n} \Big\{ (1-I_i)\Big[1 - H_C(t_i,\gamma)\Big]\Big[- H_T(t_i,\theta)\Big] - I_i H_C(t_i,\gamma)\Big[1 - H_T(t_i,\theta)\Big] \Big\}$$

and the correlation bias takes the form

$$\hat{\theta}_\delta - \hat{\theta}_0 \;\simeq\; \frac{\delta}{\imath(\theta_0)} \sum_{i=1}^{n} \Big\{ H_C(t_i,\gamma) H_T(t_i,\theta) - (1-I_i) H_T(t_i,\theta) \Big\}.$$

We can prove that when $B(t,\theta) = 1 - H_T(t,\theta)$, the expected bias takes the simple form

$$E\big[\hat{\theta}_\delta - \hat{\theta}_0\big] \;=\; \delta \frac{n}{\imath(\theta_0)} E_T\Big[H_C(t,\gamma) S_C(t,\gamma)\Big] \tag{5.8}$$

obtaining an expression equivalent to (5.6). The proof of the above result is presented in Appendix B.

From the result in (5.8) we conclude that for a given value of δ and when $B(t,\theta) = 1 - H_T(t,\theta)$, the expected bias is known. Moreover, for different values of δ we can perform sensitivity analysis, trying to understand how this function behaves in different levels of dependence.

5.1.4 Example

Following the first example in chapter 3, using the same data set, we will calculate the expectation of the correlation bias along with the upper bound.

Starting with the expectation of the bias, we managed to prove expression (5.8) under the assumption that $\frac{\partial H_T(t,\theta)}{\partial\theta} = H_T(t,\theta)$. In this particular case that we assume exponential for the T–process of the form $f_T(t,\theta) = \theta e^{-\theta t}$, the cumulative hazard becomes $H_T(t,\theta) = \theta t$, giving

$$\frac{\partial H_T(t,\theta)}{\partial\theta} = \frac{H_T(t,\theta)}{\theta}.$$

Assuming $f_C(c,\gamma) = \gamma e^{-\gamma c}$, exactly the same is true for the C–process. Therefore, (5.8) is slightly modified to take the form

$$E\left[\hat{\theta}_\delta - \hat{\theta}_0\right] = \delta\frac{n}{\gamma\iota(\theta_0)}E_T\Big[H_C(t,\gamma)S_C(t,\gamma)\Big].$$

We need to make similar adjustments to (5.7) for the upper bound (UB), which finally

takes the form

$$\left|E\left[\hat{\theta}_\delta - \hat{\theta}_0\right]\right| \leq |\delta| \frac{n}{\gamma \imath(\theta_0)} \left\{ E_T\left[T_c^2(t,\theta,\gamma)\right] \right\}^{\frac{1}{2}}$$

where

$$T_c(t,\theta,\gamma) = \int_0^t \left[1 - H_C(x,\gamma)\right] \left[\frac{1 - H_T(t,\theta)}{\theta} + H_T(x,\theta)\right] f_C(x,\gamma) dx.$$

The above small changes are simply a result of the form of the exponential we use. If we had assumed that $f_T(t,\theta) = e^{\theta} e^{-e^{\theta} t}$, then we wouldn't have the need to make any kind of modifications.

For $\delta \in [-0.004, 0.004]$, the data are in Table 5.1, while we can have a graphical presentation in Figure 5.1. We can see that the values of the bias and the expected bias

δ	$\hat{\theta}_\delta - \hat{\theta}_0$	$E[\hat{\theta}_\delta - \hat{\theta}_0]$	UB
-0.004	-0.00297	-0.00295	-0.00425
-0.003	-0.00223	-0.00221	-0.00319
-0.002	-0.00149	-0.00147	-0.00213
-0.001	-0.00074	-0.00074	-0.00106
0	0	0	0
0.001	0.00074	0.00074	0.00106
0.002	0.00149	0.00147	0.00213
0.003	0.00223	0.00221	0.00319
0.004	0.00297	0.00295	0.00425

Table 5.1: Bias, Expected Bias and the Upper Bound of the parameter of interest for different values of δ.

are remarkably close, a result which is exactly what we would expect to find. The bounds for different levels of correlation, which are calculated for any $B(t,\theta)$, show that our choice of $B(t,\theta) = 1 - H_T(t,\theta)$ provide estimates for the bias which are rather moderate. Given

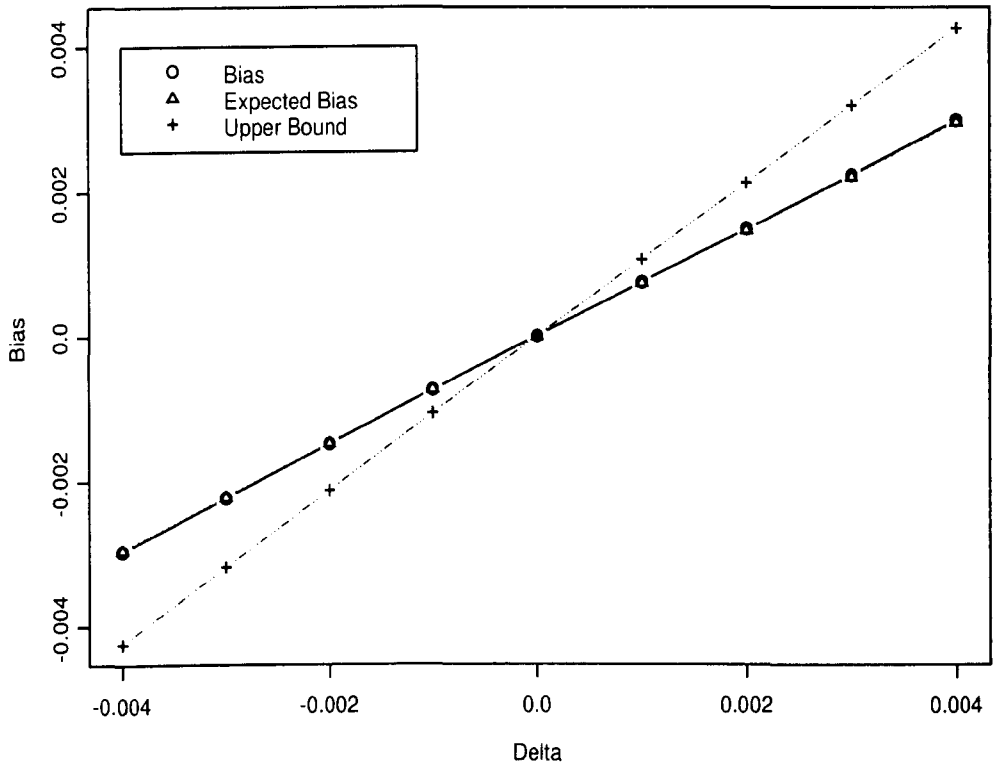

Figure 5.1: Graphical representation of the results in Table 5.1.

that we can calculate the worst possible cases, we end up having a good picture of what would possibly happen if we have low level dependencies between the two processes.

5.2 PH Including Covariates

The next obvious step in the model is to allow the presence of covariates. This means that each patient is accompanied by a set of explanatory variables, which usually describe the condition of the patient at the time of entry in the trial. Therefore, we have different parameters θ_i and γ_i for different patients which depend on the specific set of covariates of each patient. The PH structure for both processes takes the form

$$h_T(t, \mathbf{x}) = e^{\mathbf{v}'\mathbf{x}} h_T^*(t) \qquad\qquad h_C(c, \mathbf{x}) = e^{\mathbf{u}'\mathbf{x}} h_C^*(c) \tag{5.9}$$

where $\mathbf{v}$ and $\mathbf{u}$ are now vectors of parameters and $\mathbf{x}$ is the vector of the covariates.

If we assume $B(t, \theta) = 1 - H_T(t, \theta)$, the joint density becomes

$$f_{T,C}(t, c) \simeq f_T(t, \mathbf{v}) f_C(c, \mathbf{u}) \Big[1 + \delta \big[1 - H_C(c, \mathbf{u}) \big] \big[1 - H_T(t, \mathbf{v}) \big] \Big]$$

and the log-likelihood is similar to (5.2)

$$\begin{aligned} LL(t; \mathbf{v}, \mathbf{u}, \delta) &\simeq LL(t; \mathbf{v}, \mathbf{u}, \delta = 0) \\ &+ \delta \sum_{i=1}^{n} \Big\{ H_T(t_i, \mathbf{v}) H_C(t_i, \mathbf{u}) - I_i H_C(t_i, \mathbf{u}) - (1 - I_i) H_T(t_i, \mathbf{v}) \Big\} \end{aligned} \tag{5.10}$$

Therefore the correlation bias becomes

$$\hat{\mathbf{v}}_\delta - \hat{\mathbf{v}}_0 \simeq \delta \Big[\imath(\mathbf{v}_0) \Big]^{-1} \sum_{i=1}^{n} \Big\{ \mathbf{x}_i \Big[H_T(t_i, \mathbf{v}) H_C(t_i, \mathbf{u}) - (1 - I_i) H_T(t_i, \mathbf{v}) \Big] \Big\} \tag{5.11}$$

where

$$\imath(\mathbf{v}_0) \;=\; -\frac{\partial^2 LL(t;\mathbf{v},\mathbf{u},\delta=0)}{\partial\mathbf{v}\partial\mathbf{v}'}$$

is the observed information matrix and

$$\mathbf{V}(\hat{\mathbf{v}}_0) \;\simeq\; \Big[\imath(\mathbf{v}_0)\Big]^{-1}$$

is the matrix of variances and covariances of the estimates of the parameters in vector $\mathbf{v}$. The expression for the bias of the vector of regression coefficients takes the simpler form

$$\hat{\mathbf{v}}_\delta - \hat{\mathbf{v}}_0 \;\simeq\; \delta\Big[\imath(\mathbf{v}_0)\Big]^{-1}\sum_{i=1}^{n}\Big\{\mathbf{x}_i\Big[H_T(t_i,\mathbf{v})H_C(t_i,\mathbf{u}) - (1-I_i)H_T(t_i,\mathbf{v})\Big]\Big\}. \quad (5.12)$$

At the same time, what we need to do is to choose an appropriate baseline hazard. A choice following our way of modeling would be obviously a parametric baseline hazard function. In the meanwhile, the results presented in sections 5.1.1 and 5.1.2 still hold when we have covariates. The only difference is that the expected correlation bias and the upper bound are now conditional on a specific set of covariates.

5.3 The Independence Model

In this work, many times we have needed to refer to the paper written by Tsiatis(1975). According to his work, if we use a model to analyze survival data that assumes dependence between failure and censored times, then there exists a unique proxy model with independence between the two processes, from which we can derive the same sub-functions as from the dependence model. Therefore, it would be of great interest to see the form of this

proxy model, based on our dependence model.

According to Tsiatis(1975) Theorem 2, and using his notation, the joint survival function of the proxy model is defined by

$$\bar{F}^*(\mathbf{t}) = \prod_{j=1}^{p} \bar{F}_j^*(t_j) \tag{5.13}$$

when p risks are present, where

$$\bar{F}_j^*(u) = \exp\left\{ -\int_0^u h(j,s)ds \right\}, \tag{5.14}$$

is the survival function for risk j of the proxy model, and $h(j,s)$ is the sub-hazard, derived from the given model. $h(j,s)$ is the hazard of failing from cause j at time s in the presence of all the other risks.

In our particular model, we have only two potential risks, T and C. Under our notation, the joint survival function takes the form

$$\begin{aligned} S_{T,C}(\mathbf{t}) &= \int_t^\infty \int_c^\infty f_T(u,\theta) f_C(v,\gamma) \left[1 + \delta \frac{\partial \log f_C(v,\gamma)}{\partial \gamma} \right] du dv \\ &= S_T(t,\theta) S_C(c,\gamma) \left[1 - \delta \mu(t,\theta) \frac{\partial H_C(c,\gamma)}{\partial \gamma} \right] \end{aligned} \tag{5.15}$$

where $\mu(t,\theta)$ is as defined in chapter 3. Using $S_{T,C}(\mathbf{t})$ we can get the sub-densities and then the sub-hazard functions. Firstly for the T–process, the sub-density becomes

$$f_T^\sharp(u) = -\left[\frac{\partial \bar{S}(\mathbf{t})}{\partial t} \right]_{\mathbf{u}} = f_T(u,\theta) S_C(u,\gamma) \left[1 - \delta B(u,\theta) \frac{\partial H_C(u,\gamma)}{\partial \gamma} \right] \tag{5.16}$$

and the corresponding sub-hazard is

$$h_T^{\sharp}(u) = \frac{f_T^{\sharp}(u)}{\bar{S}(\mathbf{u})} = h_T(u,\theta)\Big[1 + \delta\frac{\partial H_C(u,\gamma)}{\partial\gamma}\Big(\mu(u,\theta) - B(u,\theta)\Big)\Big]. \tag{5.17}$$

Similarly, for the C-process we have

$$f_C^{\sharp}(u) = -\Big[\frac{\partial\bar{S}(\mathbf{t})}{\partial c}\Big]_{\mathbf{u}} = f_C(u,\gamma)S_T(u,\theta)\Big[1 + \delta\mu(u,\theta)\frac{\partial\log f_C(u,\gamma)}{\partial\gamma}\Big] \tag{5.18}$$

and

$$h_C^{\sharp}(u) = \frac{f_C^{\sharp}(u)}{\bar{S}(\mathbf{u})} = h_C(u,\gamma)\Big[1 + \delta\mu(u,\theta)\frac{\partial\log h_C(u,\gamma)}{\partial\gamma}\Big]. \tag{5.19}$$

The above sub-hazards are the basis of constructing the independent model, in which they are considered to be the marginal hazard functions. Assuming that $\bar{G}_T(t)$ and $\bar{G}_C(c)$ are the independent survival functions of our new model, according to (5.14) we have

$$\begin{aligned}
\bar{G}_T(t) &= \exp\Big\{-\int_0^t h_T^{\sharp}(s)ds\Big\} \\
&= \exp\Big\{-H_T(t,\theta) - \delta\int_0^t h_T(s,\theta)\frac{\partial H_C(s,\gamma)}{\partial\gamma}\Big(\mu(s,\theta) - B(s,\theta)\Big)ds\Big\} \\
&= S_T(t,\theta)\exp\Big\{-\delta\int_0^t h_T(s,\theta)\frac{\partial H_C(s,\gamma)}{\partial\gamma}\Big(\mu(s,\theta) - B(s,\theta)\Big)ds\Big\}
\end{aligned} \tag{5.20}$$

and

$$\begin{aligned}
\bar{G}_C(c) &= \exp\Big\{-\int_0^c h_C^{\sharp}(s)ds\Big\} \\
&= \exp\Big\{-H_C(c,\gamma) - \delta\int_0^c h_C(s,\gamma)\mu(s,\theta)\frac{\partial\log h_C(s,\gamma)}{\partial\gamma}ds \\
&= S_C(c,\gamma)\exp\Big\{-\delta\int_0^c h_C(s,\gamma)\mu(s,\theta)\frac{\partial\log h_C(s,\gamma)}{\partial\gamma}ds\Big\}.
\end{aligned} \tag{5.21}$$

The product of (5.20) and (5.21) under the presence of independence, is simply the joint

survival function, for a given value of δ. From the above equations it is clear that in the new model, the marginal functions of risk T functionally depend on the functions of the C–process and vice versa.

Under this new framework we are now interested in the probability of an event, either failure or censoring. Assuming that $g_T(t)$ and $g_C(c)$ are the density functions and $\lambda_T(t)$ and $\lambda_C(c)$ are the hazard functions of the proxy model, then the probability of the minimum is

$$P = \left[g_T(t)\bar{G}_C(t)\right]^I \left[g_C(t)\bar{G}_T(t)\right]^{1-I} = \lambda_T(t)^I \lambda_C(t)^{1-I} \bar{G}_T(t)\bar{G}_C(t)$$

where I is the indicator variable. The marginal hazard functions of the proxy model are equal to the sub–hazards of the original dependent model. Attempting to draw inferences about θ, which was the initial parameter of interest, we construct the likelihood function which has the form

$$L_P = \prod_{i=1}^{n} \lambda_T(t_i)^{I_i} \lambda_C(t_i)^{1-I_i} \bar{G}_T(t_i)\bar{G}_C(t_i)$$

and the log–likelihood is

$$\begin{aligned} LL_P &= \sum_{i=1}^{n} \left\{ I_i \log \lambda_T(t_i) + (1 - I_i) \log \lambda_C(t_i) + \log \bar{G}_T(t_i) + \log \bar{G}_C(t_i) \right\} \\ &= \sum_{i=1}^{n} \left\{ I_i \log \lambda_T(t_i) + (1 - I_i) \log \lambda_C(t_i) - \Lambda_T(t_i) - \Lambda_C(t_i) \right\} \end{aligned} \tag{5.22}$$

with $\Lambda_*(t)$ indicating the cumulative hazard functions. We can prove that

$$LL_P = LL, \tag{5.23}$$

indicating that the two log–likelihood functions, the one from the dependent model and

the other one from the independent, are exactly the same. This means that inferences about θ are exactly the same, whichever model we use. The proof of the above is included in Appendix C.

The above result is not something new. Following Tsiatis's Theorem, we would expect to find the same distribution for the minimum, and hence draw the same inferences about θ. The main gain from the above is that now we have the exact form of the model with independent risks, proving that inferences for parameter θ are the same whichever model we decide to use. Benefits from that would be more obvious in the next Chapter, but the knowledge of an equivalent to our initial model with independent risks is a great advantage.

5.3.1 Example

We use the myeloma data, but this time we include the covariates presented in Table 3.2. In the example in chapter 3 we show that an exponential distribution would give a reasonably good fit to the T-process. In the same way we can show that an exponential fit would be appropriate for the C-process as well. Therefore, we model the hazard functions according to expression (5.9), choosing to have a constant baseline hazard function for both processes. The hazard functions of the two processes now take the form

$$h_T(t; \mathbf{v}, \theta, \mathbf{x}) = e^{\mathbf{v}'\mathbf{x}}\theta, \qquad h_C(t; \mathbf{u}, \gamma, \mathbf{x}) = e^{\mathbf{u}'\mathbf{x}}\gamma,$$

where θ and γ are the constant baseline hazards. The estimates of the parameters are $\hat{\mathbf{v}}_0 = (1.5567, -0.1065, 0.4214, 0.1251)$, $\hat{\theta}_0 = 0.0017$, $\hat{\mathbf{u}}_0 = (1.7135, 0.0910, 0.5703, -0.1248)$ and $\hat{\gamma}_0 = 0.0006$, obviously when $\delta = 0$. The main advantage of this way of modeling is

that we can perform a sensitivity analysis on any of these parameters, including θ, and observe the impact of the changes in the hazard and the survival functions.

Another way of dealing with the same problem is to focus on a quantity with a real meaning like the hazard itself. More specifically we can assume that the logarithm of the hazard is of main interest, which seems a natural thing to look at. Following what we said in the previous paragraph, we can assume that for a fixed set of covariates the hazard is constant, and hence each patient has an exponential survival probability. Without any loss of generality we may assume that $h_T(t;\mathbf{v},\mathbf{x}) = e^{\mathbf{v}'\mathbf{x}}$ and $h_C(t;\mathbf{u},\mathbf{x}) = e^{\mathbf{u}'\mathbf{x}}$, where now vectors $\mathbf{v}$ and $\mathbf{u}$ are not the same as before. The fact that we have eliminated the baseline hazard indicates that the vector $\mathbf{x}$ must have an intercept, changing in that way the vectors of the parameters. Now the independent estimates are $\hat{\mathbf{v}}_0 = (-6.3875, 1.5567, -0.1065, 0.4214, 0.1251)$ and $\hat{\mathbf{u}}_0 = (-7.4426, 1.7135, 0.0911, 0.5703, -0.1248)$. Therefore, we assume that $w_{\mathbf{x}} = \log h_T(t;\mathbf{v},\mathbf{x}) = \mathbf{v}'\mathbf{x}$ is our main parameter, which is simply the prognostic index (PI) of the T–process, and we will perform a sensitivity analysis conditional on the set of covariates $\mathbf{x}$. Similarly, if we assume that $z_{\mathbf{x}} = \mathbf{u}'\mathbf{x}$, the PI of the C–process, the expression of the bias becomes

$$\hat{w}^{\delta}_{\mathbf{x}} - \hat{w}^{0}_{\mathbf{x}} \simeq \delta \frac{\sum_{i=1}^{n} \left\{ \exp(z^{0}_{\mathbf{x}}) t_i^2 - (1 - I_i) t_i \right\}}{\sum_{i=1}^{n} t_i}. \tag{5.24}$$

We can see that the correction factor (or sensitivity index SI) depends only on the observed times and $z_{\mathbf{x}}$. This means that the greater the hazard of being censored the more sensitive is the dependent estimate $\hat{w}^{\delta}_{\mathbf{x}}$ for a given value of δ. The relationship between ρ and δ is also of major importance. Under this particular way of modelling, which is proportional

hazards, we already know from the results in chapter 4 that $\rho \leq \delta$. In this case it can be proven that $\rho = \delta$. This means that our sensitivity parameter is nothing else but the correlation between the two processes, and hence by substituting δ by ρ in (5.24) we get the final expression of the bias. The results of the sensitivity analysis are presented in Figures 5.2 -5.5. Starting from Figure 5.5 we see that large PI for the T-process imply large PI for the C- process. This means that patients who are more likely to die are more likely to be censored as well, giving immediately an indication of the presence of a possible positive correlation between T and C, conditional on the set of covariates $\mathbf{x}$. As a result Figures 5.2 and 5.3 show the impact of the sensitivity analysis on the survival functions. More specifically Figure 5.2 illustrates the survival curves of the patients with the best and the worst PI for the T-process. We see that the patient with the worst prognosis has a more sensitive survival curve simply because this goes with an equally poor prognosis for the C-process giving a high level of SI for our chosen $\rho = 0.3$. Similarly Figure 5.3 presents the patients with best and the worst PI for the C-process which are the patients with the smallest and the largest SI in our sample. Finally in Figure 5.4 we plot the SI against the PI of the C-process. As expected this gives an increasing smooth line illustrating in a very clear way the relationship between the two quantities.

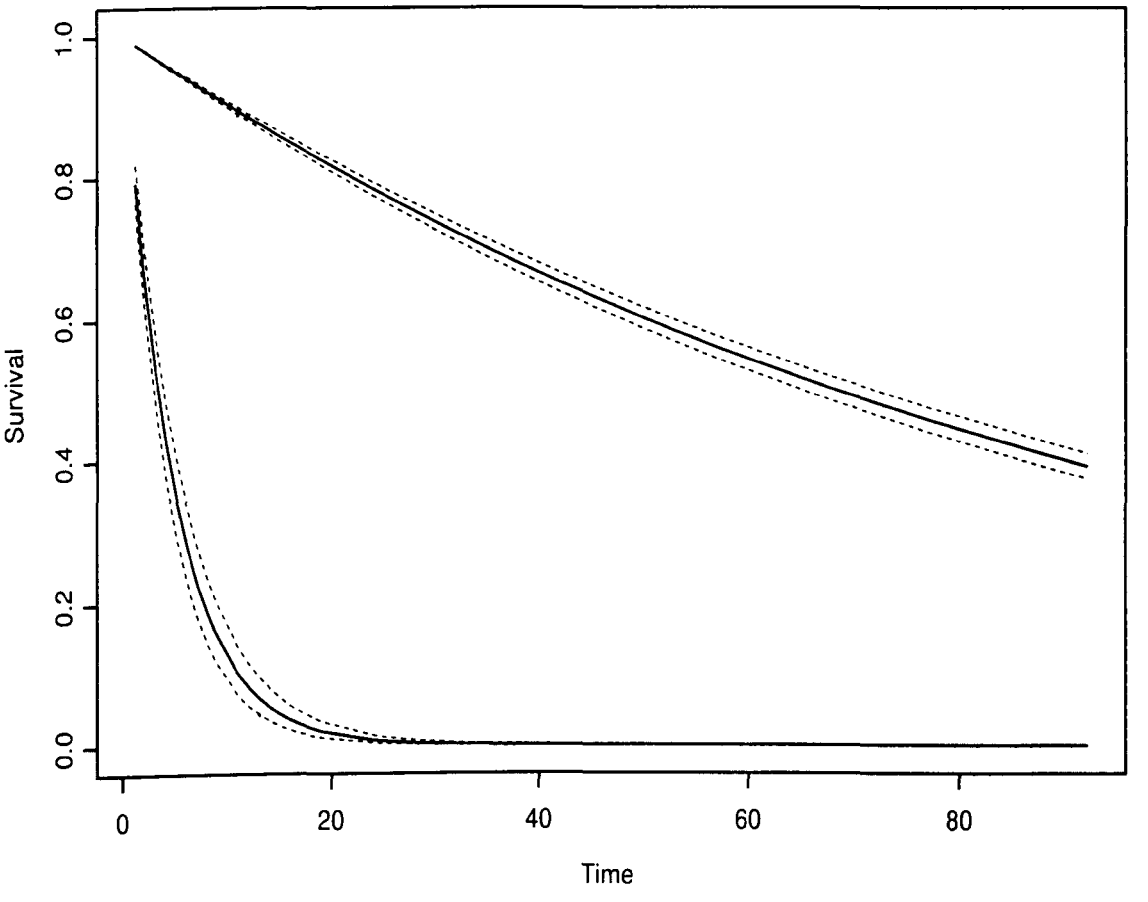

Figure 5.2: Min/max sensitivity on the survival with respect to the PI of the T–process.

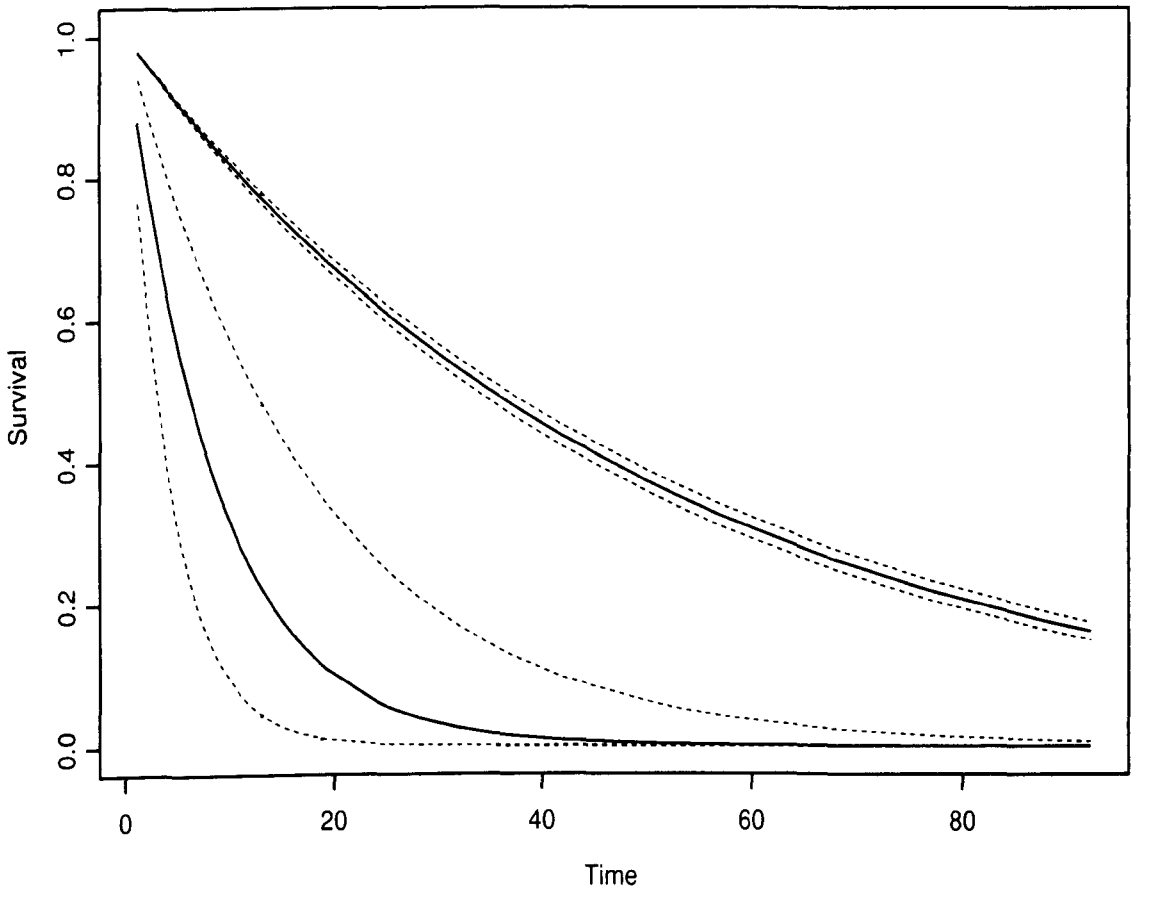

Figure 5.3: Min/max sensitivity on the survival with respect to the PI of the C–process.

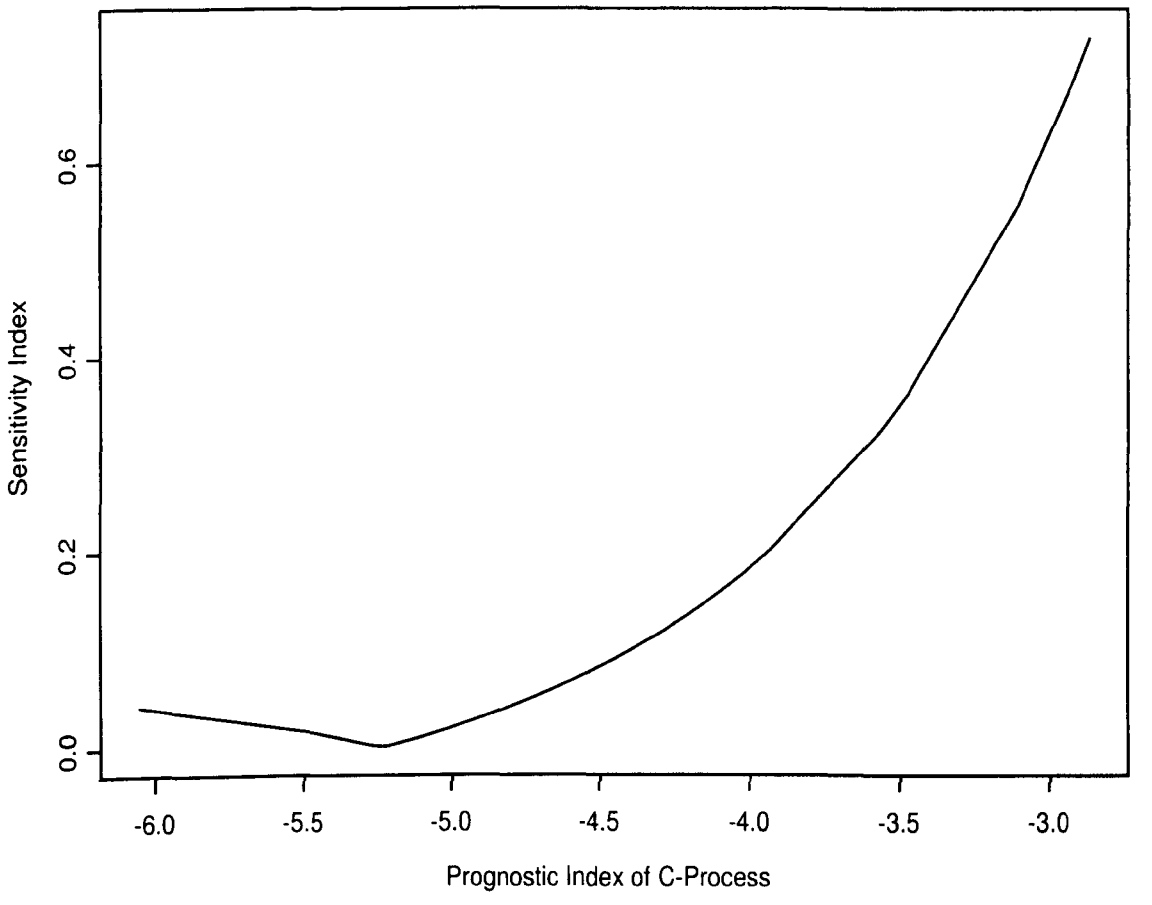

Figure 5.4: Graphical representation of the Sensitivity Index for $\rho = 0.3$.

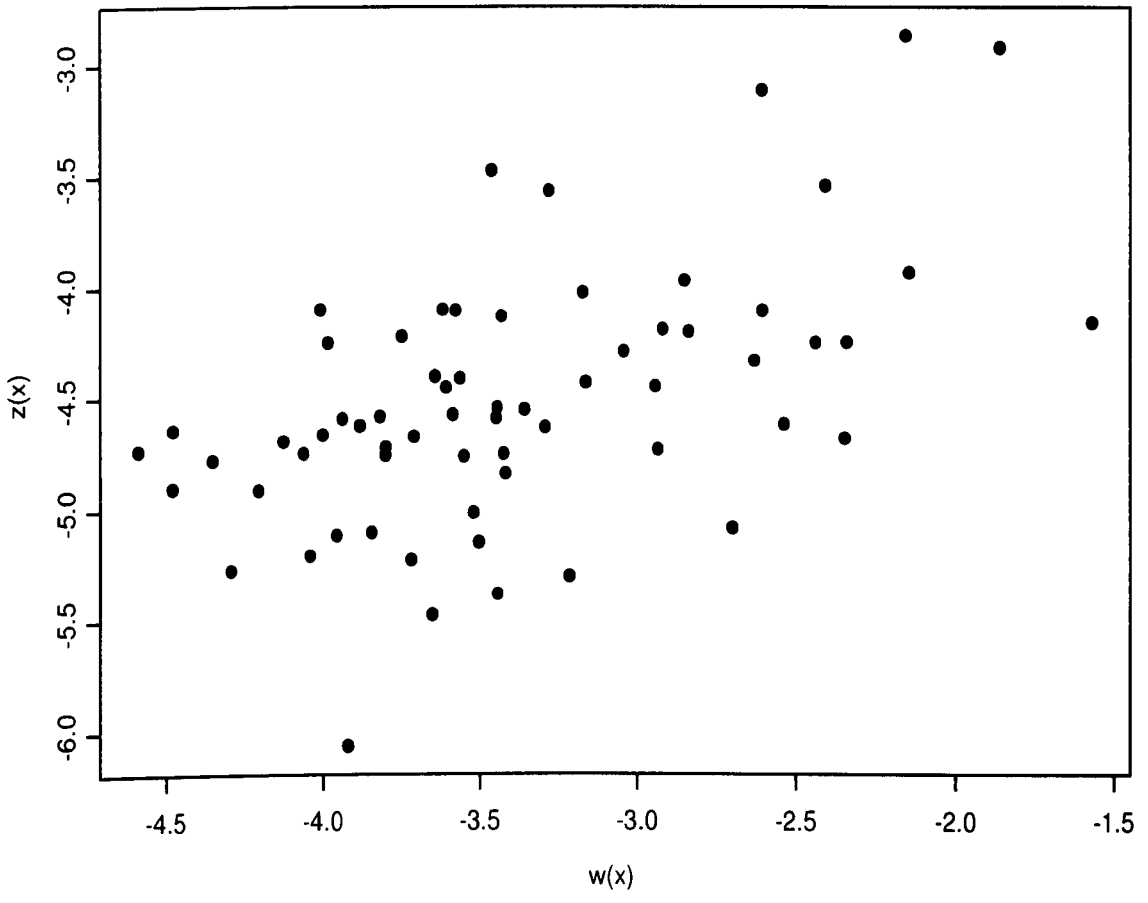

Figure 5.5: Plot of the PI of the T–process against the PI of the C–process.

5.4 Model with Informative and Uninformative Censoring

5.4.1 Introduction

So far, all of our work is based on one and only one initial assumption, that there exist some kind of dependence between the failure and the censored lifetimes. We are not able to measure it, but we are able to model it and see what happens when this is not zero. There are though some question that need to be answered. First of all, is it reasonable to assume that all the censored observations are either all informative or all uninformative? Is it possible from the context to identify which censored observations are definitely uninformative?

A common situation is when we have a good proportion of observations censored at the end of the study. It is definitely a different kind of censoring, something like a "forced" censoring. It occurs not due to some random event or to an event related to our experiment, but due to a lack of time or even because the whole trial was designed so as to end at a specific point in time, preventing us from continuing the study. Now the question is whether we should treat these censored times as being different from the ones that happened randomly throughout the study or not. This means that if we consider dependence between the failure and the censored times, will this dependence involve both kinds of censored times or not? If the answer is no, then there is no change at all to what we have done so far. All the calculations still hold, and we make no distinction between the two kinds of censoring. But, if the answer is yes, then we can easily assume that the censored times at the end of the

study are due to a reason which is completely irrelevant to what we are investigating, and hence they are independent of the failure times. Therefore, we end up with two different types of censored lifetimes, out of which one is dependent and the other one is independent of the failure times. At the end, this will have an impact to the likelihood function and to the estimates of the parameters.

5.4.2 The Independence of Censoring at the End of the Study

We will now try to model the case where the censoring at the end of the study is considered to be independent, using the exact likelihood. Now we have 3 events instead of 2. We have T, C_I and C_E, where T as before is the failure time, C_I the censoring before the end of the study and C_E is the censoring at the end of the study. In the last case what we know is that both T and C_I are greater than C_E, and what we finally observe is the minimum of these 3 possible events, hence $Y = \min\{T, C_I, C_E\}$. In other words this is a truncated version of the censoring we have been using so far.

The likelihood function now has an extra term, and takes the form

$$L'(t;\delta,\theta,\gamma) = \prod_{i=1}^{n} Pr\big(T = t_i, T < C\big)^{I_i Z_i} Pr\big(C = t_i, C < T\big)^{(1-I_i)Z_i} Pr\big(E = t_i\big)^{(1-I_i)(1-Z_i)} \tag{5.25}$$

where

$$I_i = \begin{cases} 1, & \ldots \text{ when failure time} \\ 0, & \ldots \text{ when censored time} \end{cases}$$

and

$$Z_i = \begin{cases} 1. & \ldots \text{ when } C_I \text{ (censored before the end)} \\ 0. & \ldots \text{ when } C_E \text{ (censored at the end)} \end{cases}$$

are the two indicator variables. $Pr(E = t_E)$ is the probability that a patient is censored at the end of the study. This means that both T and C_I are greater than t_E, the time of the end of the study, and hence we have

$$\begin{aligned} Pr(E) &= \int_{t_E}^{\infty}\int_{t_E}^{\infty} f_C\big(c, \gamma + \delta B(t,\theta)\big) f_T\big(t,\theta\big) dt dc \\ &= S_T(t_E,\theta) S_C(t_E,\gamma)\Big[1 - \delta \frac{\partial H_C(t_E,\gamma)}{\partial \gamma}\mu(t_E,\theta)\Big]. \end{aligned} \tag{5.26}$$

It is clear that if all the patients join the trial at the same time, then t_E will be the same for everyone, but if patients join the trial at different points in time, then each patient will have his own time t_E^i, which would be know from the beginning if the trial has a specified end point.

Now the likelihood function takes the form

$$\begin{aligned} L'(t;\delta,\theta,\gamma) = \prod_{i=1}^{n} \Bigg\{ & \Big[f_T(t_i,\theta) S_C(t_i,\gamma)\Big(1 - \delta B(t_i,\theta)\frac{\partial H_C(t_i,\gamma)}{\partial \gamma}\Big)\Big]^{I_i Z_i} \\ & \Big[f_C(t_i,\gamma) S_T(t_i,\theta)\Big(1 + \delta \mu(t_i,\theta)\frac{\partial \log f_C(t_i,\gamma)}{\partial \gamma}\Big)\Big]^{(1-I_i)Z_i} \\ & \Big[S_T(t_i,\theta) S_C(t_i,\gamma)\Big(1 - \delta \frac{\partial H_C(t_i,\gamma)}{\partial \gamma}\mu(t_i,\theta)\Big)\Big]^{(1-I_i)(1-Z_i)} \Bigg\} \end{aligned} \tag{5.27}$$

and after some calculations we end up with the log–likelihood

$$\begin{aligned} LL'(t;\theta,\gamma,\delta) &= \sum_{i=1}^{r+k}\Big\{I_i\log h_T(t_i,\theta)+(1-I_i)\log h_C(t_i,\gamma)-H_T(t_i,\theta)-H_C(t_i,\gamma)\Big\} \\ &- \sum_{i=1}^{w}\Big\{H_C(t_i,\gamma)+H_T(t_i,\theta)\Big\} \\ &+ \delta\Bigg[\sum_{i=1}^{r+k}\Big\{(1-I_i)\frac{\partial\log f_c(t_i,\gamma)}{\partial\gamma}\mu(t_i,\theta)-I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}B(t_i,\theta)\Big\} \\ &+ \sum_{i=1}^{w}\Big\{\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\mu(t_i,\theta)\Big\}\Bigg] \end{aligned} \tag{5.28}$$

where

r – the number of failure times

k – the number of censored times before the end

w – the number of times censored at the end.

Therefore, the log–likelihood has the same structure as the initial one, with the only difference being that we add an extra term that corresponds to the assumption of censoring at the end of the study. If we now proceed with the estimation of the parameter of interest, we end up with

$$\begin{aligned} \hat\theta_\delta-\hat\theta_0 \simeq \frac{\delta}{\iota(\theta)}\Bigg[&\sum_{i=1}^{r+k}\Big\{(1-I_i)\frac{\partial\log f_c(t_i,\gamma)}{\partial\gamma}\frac{\partial\mu(t_i,\theta)}{\partial\theta}-I_i\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\frac{\partial B(t_i,\theta)}{\partial\theta}\Big\} \\ &+\sum_{i=1}^{w}\Big\{\frac{\partial H_C(t_i,\gamma)}{\partial\gamma}\frac{\partial\mu(t_i,\theta)}{\partial\theta}\Big\}\Bigg] \end{aligned} \tag{5.29}$$

where

$$\begin{aligned} LL'(t;\theta,\gamma,\delta=0) &= \sum_{i=1}^{r+k}\Big\{I_i\log h_T(t_i,\theta)+(1-I_i)\log h_C(t_i,\gamma)-H_T(t_i,\theta)-H_C(t_i,\gamma)\Big\} \\ &- \sum_{i=1}^{w}\Big\{H_C(t_i,\gamma)+H_T(t_i,\theta)\Big\}. \end{aligned}$$

Calculations are straight forward and what we get is what we expected to get, a modified version of our initial estimate of the parameter θ.

The above way of modeling, with 3 possible events T, C_I and C_E can be seen as a special case of a general model. Assume that we have a competing risk problem, in which we are interested only in one particular risk. We can assume that the failure times from this particular risk form the T–process, while the failure times from the other risks can be categorized, according to whether we assume dependence between them and the particular risk of interest or not. Hence, we form C_I–process and C_E–process. This is a small extension of the model discussed in this section. The likelihood function would be equal to (5.25), with the only difference that the observations of the E–process will not necessarily appear at the end of the study.

Chapter 6

What if we knew more?

6.1 Presenting Data with which we can Estimate Parameter δ

6.1.1 Introduction

As we have mentioned several times before, parameter δ is a quantity that cannot be estimated. The main reason for that is that the data themselves do not provide enough information, leaving us with the question of what would happen if we had the opportunity to observe more. So far we have based all our work in the assumption that δ is actually known, avoiding in that way the problems of estimating it. In this chapter we will work on the special data set of Table 2.1. Under some certain assumptions, discussed in section 2.2, we do observe additional information which enables us address the question of the size of δ.

Under our hypothetical scenario with viewpoints A and B, we are privileged to have a data set of 38 patients in which we observe the exact death time of 11 of them, the censored time of the 14 of them, and both the censoring and the exact death time of the remaining 13 patients, providing us with far more information than we usually have when we analyze

survival data. In the way we have "constructed" Data A, censoring is informative (or at least part of it), while in Data B censoring is considered to be uninformative. We believe that in a real life situation it is more likely to observe Viewpoint A rather than Viewpoint B. Doctors will always try to save the patients lives, if they can, rather than try to make up their numbers for their statistical analysis. For this reason we have considered Data A to be the observed data while Data B to be the "true" data, in the sense that any estimates of parameters of the failure process that come from Data A will be biased, because of the informative censoring, while the estimates of the same parameters that come from Data B will be unbiased.

In Figure 2.1 we saw how misleading the estimate of the survival curve under independence can be. What we want to do is to take advantage of the additional information we have, and explore the possibility of improving our estimate of the survival curve which we would make if we only knew Data A. Our main target is to use this extra information to estimate δ.

6.1.2 Analyzing the Data

Initially assume that censoring is uninformative. We admit that we do not have many observations, and the presence of some long term survivors suggests that we should try to analyze these data using a mixture model. We assume that there is a proportion p of the patients which will never die, the so called immunes, while the rest of the patients die with an exponential rate, with parameter θ, see Farewell(1977) and Maller & Zhou(1996).

Hence the survival function takes the form

$$S_T(t,\theta,p) = p + (1-p)e^{-\theta t} \tag{6.1}$$

and the density becomes

$$f_T(t,\theta,p) = (1-p)\theta e^{-\theta t}. \tag{6.2}$$

Additionally, for the C–process we assume a simple exponential model with parameter γ. The log–survival plot for the C–process looked reasonably straight, suggesting that an exponential distribution would fit the data reasonably well.

As a result of our assumption, $f_T(t,\theta,p)$ is not a proper density function, allowing a cumulative probability p when $t \to \infty$. Furthermore, we now have two parameters, θ and p, and we will estimate them by maximum likelihood estimation. The independent estimates, assuming non informative censoring, are $\hat{\theta}_0^A = 0.00291012$ and $\hat{p}_0^A = 0.607822$, where A indicates that these estimates come from Data A, and 0 means independence ($\delta = 0$). If we use Data B, then we get $\hat{\theta}_0^B = 0.00246945$ and $\hat{p}_0^B = 0.31969$, which we consider to be the "true" estimates, in the sense that they are the unbiased estimates of the real true parameters. Figure 6.1 shows the fit of the model (6.1) using the above estimates.

It is true that in either of the above cases, where we estimate parameters θ and p, we discard pieces of information. In the first case we ignore the exact death times of the patients who relapse and in the second case we ignore the censored times of the same patients. The reason for that is that in practice we have either the one case or the other but we will never have all these data at the same time. Later in this chapter we will see

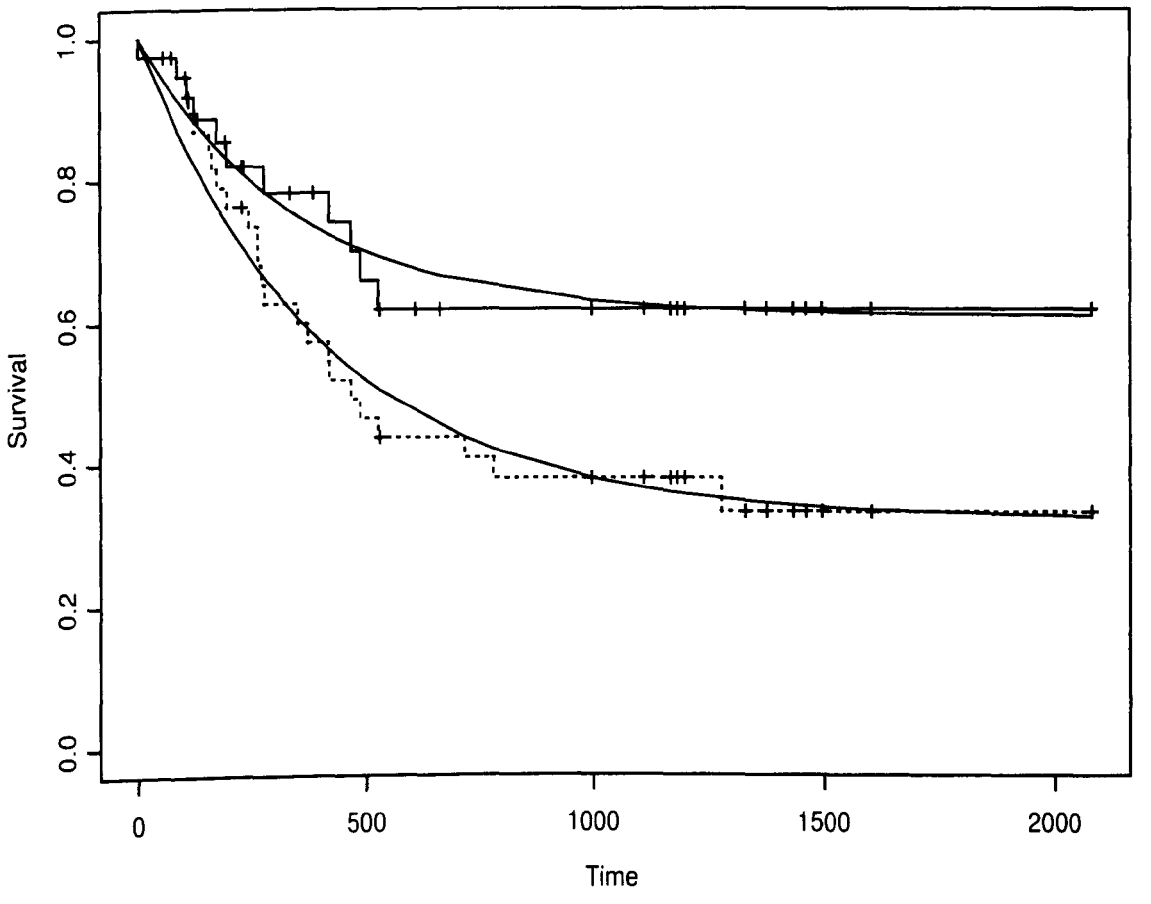

Figure 6.1: Fit of the independent parametric survival curves using the mixture model.

how we can use all the information we have in order to estimate δ.

6.1.3 Estimation of parameter δ

As we have mentioned before, the difference between the two curves is large. Actually the survival probability in the tails is almost double when we have the relapsed censored times than when we have the exact death times. The question we need to answer now is "what is the value of δ that, given that we use the data which include the informative censored times, we will still be able to get the "true" estimates for our parameters". In other words, which is the value of δ for which

$$\hat{\theta}_{\delta}^{A} = \hat{\theta}_{0}^{B}$$

and

$$\hat{p}_{\delta}^{A} = \hat{p}_{0}^{B}.$$

In order to be able to answer this question we need to use the exact likelihood function

$$L(t;\theta,p,\gamma,\delta) = \prod_{i=1}^{n} P(T=t_i \cap T<C)^{I_i} P(C=c_i \cap C<T)^{1-I_i} \tag{6.3}$$

and not the approximation we introduced in chapter 3. The reason for that is that we do not know how large our estimate of δ is going to be, so working with the exact model we give no restrictions on the value of δ, while using the earlier linear approximation we restrict δ to be small. The probabilities under the mixture model are

$$P(T=t_i \cap T<C) = f_T(t,\theta,p)S_C\big(t,\gamma+\delta(1-H_T(t,\theta,p))\big) \tag{6.4}$$

and

$$P(C = c_i \cap C < T) = \int_c^{\infty} f_C\big(c, \gamma + \delta(1 - H_T(t,\theta,p))\big) f_T(t,\theta,p)dt + p \lim_{t\to\infty} \Big\{ f_C\big(c, \gamma + \delta(1 - H_T(t,\theta,p))\big)\Big\}, \quad (6.5)$$

where the correction $p \lim_{t\to\infty} \big\{ f_C\big(c, \gamma + \delta(1 - H_T(t,\theta,p))\big)\big\}$ in the above expression comes from the fact that $\lim_{t\to\infty} \big\{S_T(t,\theta,p)\big\} = p$ and not zero, as when we have a proper density. Therefore, we require that expression (6.3), for the right value of δ, to give the "true" MLEs of the parameters of interest. Being in the position to know the "true" parameters $\hat{\theta}_0^B$ and $\hat{p}_0^B$, we simply substitute them in (6.3), along with $\gamma_0^A = 0.00116589$, and we get the MLE of δ. Hence, $L(t; \hat{\theta}_0^B, \hat{p}_0^B, \delta)$ gives an estimate $\hat{\delta} = 0.00121015$. This is not a true maximum likelihood estimate but an ad hoc argument; it provides us with $\hat{\delta}$ which is the MLE of δ conditional on the fact that $\hat{\theta}_\delta^A = \hat{\theta}_0^B$ and $\hat{p}_\delta^A = \hat{p}_0^B$.

Another way of estimating all the parameters together, including δ, is to construct the appropriate likelihood for all the data we have available. In this case we have 3 different types of observations

i) observe death time but not censored time, with probability as in (6.4)

ii) observe both censored and death times, with joint probability density

$$f_{T,C}(t,c) = f_T(t,\theta,p) f_{C|T}(c, \gamma + \delta(1 - H_T(t,\theta,p))) \quad (6.6)$$

and finally

iii) observe censored but not death time or neither of them. This third type of observations can be seen in two different ways. If we assume that these censored times are due to the

end of study, which is probably the case, then this means that during the follow up time we were able to observe neither the death time nor the censored time. Hence, if E is the time of the end of the study we have that

$$\begin{aligned} Pr(T > E \cap C > E) &= \int_E^\infty \int_E^\infty f_T(t,\theta,p) f_{C|T}(c, \gamma + \delta(1 - H_T(t,\theta,p)))dtdc \\ &\quad + p \lim_{t\to\infty} \Big\{ \int_E^\infty f_{C|T}(c, \gamma + \delta(1 - H_T(t,\theta,p)))dc \Big\}, \end{aligned} \tag{6.7}$$

exactly like the end of study censoring in the previous chapter with a small correction at the end. On the other hand if we assume that what we observe is the actual censored time, then expression (6.5) provides the probability of such an event. The major distinction between the two is that assuming that the reason for a patient to be censored is to relapse (informative), then in the first case we say that we were not able to observe it, while in the second one we claim that the observed censored time is actually the time where the patient relapses.

If we assume that we have the end of study censoring for category (*iii*), then the likelihood function takes the form

$$\begin{aligned} L(t;\theta,p,\gamma,\delta) &= \prod_{i=1}^{n} P(T = t_i \cap T < C)^{I_i(1-Z_i)} Pr(T = t_i \cap C = c_i)^{I_i Z_i} \\ &\quad P(T > E \cap C > E)^{(1-I_i)Z_i} \end{aligned} \tag{6.8}$$

where I_i indicates whether we observe the death time of the i^{th} patient or not and equivalently Z_i indicates whether we observe the censored time or not. Maximizing this function over the parameters θ,p,γ and δ we get the estimates presented in Table 6.1, which now

$\hat{\theta}$	$\hat{p}$	$\hat{\gamma}$	$\hat{\delta}$
0.0025655	0.286328	0.000290804	0.0018024

Table 6.1: Second set of "true" estimates.

gives the second set of "true" estimates.

Finally, assuming that we have the second type of censoring for category (*iii*), which computationally is easier to deal with, we have the following likelihood function

$$\begin{aligned} L(t;\theta,p,\gamma,\delta) &= \prod_{i=1}^{n} P(T=t_i \cap T<C)^{I_i(1-Z_i)} Pr(T=t_i \cap C=c_i)^{I_i Z_i} \\ & \quad P(C=c_i \cap C<T)^{(1-I_i)Z_i}. \end{aligned} \tag{6.9}$$

The MLEs of parameters θ,p,γ and δ are in Table 6.2, where now these are the third set

$\hat{\theta}$	$\hat{p}$	$\hat{\gamma}$	$\hat{\delta}$
0.00236544	0.28971	0.00122373	0.00202389

Table 6.2: Third set of "true" estimates.

of "true" estimates.

People might find confusing the distinction of the estimates into three sets of "true" estimates. This has to do with how we use the data in order to obtain our unbiased estimates of the parameters. The first way of estimation I believe is the most realistic, because is based on the actual data that we will have in real life. The other two are basically the same and they are based on using all the data that we have in this particular case, although in practice we wont be so lucky to have all this information. In order to avoid confusion, in the rest of the chapter we will refer to the different estimates of δ as $\hat{\delta}_1$,$\hat{\delta}_2$ and $\hat{\delta}_3$ meaning that they come from the three different ways of estimating them.

6.1.4 Estimating the Parameters of Interest Using the Estimated Values of δ

In the previous section we described ways of estimating parameters δ. The obvious second step would be to use these values of δ in our original likelihood function and try to estimate the parameters of interest. We would hope to get estimates for the parameters close to the "true" values, but the fact that we do not have enough observations and the variation is large indicates that this might be too optimistic.

If we assume that we have $\hat{\delta}_1 = 0.00121015$, the value we get using the first way of estimation, we get $\hat{\theta}_\delta = 0.00198907$ and $\hat{p}_\delta = 0.411352$. It is important to note is that the parameter γ of the C-process is kept fixed and equal to $\hat{\gamma}_0^A$ during all the calculations, because it is required from the model that γ is the parameter under independence. In Figure 6.2 we plot the KM estimate of the "true" survival curve along with the 95% confidence limits calculated using the Greenwood's formula. It is obvious that for reasons we described before we get a wide confidence interval. The solid line is the survival curve using the "true" parameters, while the dashed line is the fitted survival curve, using the above estimates. We see that is not the best fit to the "true" survival curve, but is still within the 95% confidence limits. If we had more data we would expect to get better estimates, but given the presence of the variation we get a reasonably good fit.

The fact that the dashed line falls in the 95% confidence interval is encouraging, and is a visual indication that we have reasonably good estimates. However, we would like to test this result using the likelihood ratio test to see whether the survival curves can be considered indistinguishable or not. The likelihood that we will use is the one under

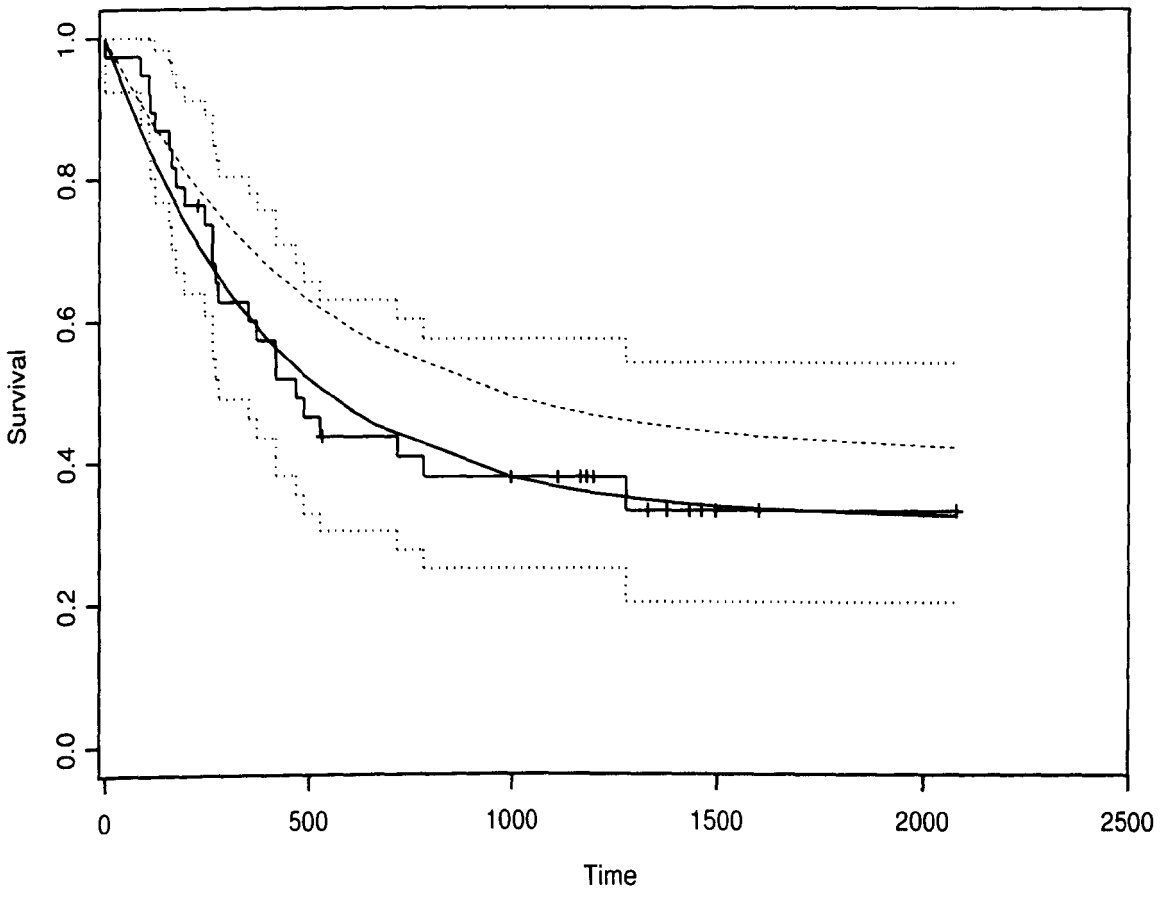

Figure 6.2: Estimated survival curve using $\hat{\delta}_1$.

independence and we will use Data B, because this is the case that provides us with the unbiased estimates and hence we can test whether one set of parameters is significantly different from the unbiased ("true") ones. Therefore, the likelihood ratio gives

$$-2\log\left[\frac{L_0(T_2, 0.00198907, 0.411352)}{L_0(T_2, 0.00246945, 0.31969)}\right] = 2.33101$$

which, compared with 4.60517 which is the $\chi^2_2(0.10)$ with 2 degrees of freedom, shows that we cannot reject the hypothesis that the two survival curves are the same.

Taking now $\hat{\delta}_2 = 0.0018024$, the MLEs of the parameters are $\hat{\theta}_\delta = 0.00169965$ and $\hat{p}_\delta = 0.360341$. Figure 6.3 shows the fitted survival curve plotted together with the KM estimate and the curve using the "true" parameters from Table 6.1. We can see that it falls within the 95% limits of the KM estimate. Hence we perform the likelihood ratio test and we get

$$-2\log\left[\frac{L_0(T_2, 0.00169965, 0.360341)}{L_0(T_2, 0.0025655, 0.286328)}\right] = 2.31174$$

which is still not enough to reject our null hypothesis. If finally we assume that $\hat{\delta}_3 = 0.00202389$, our new estimates for the parameters are $\hat{\theta}_\delta = 0.00153197$ and $\hat{p}_\delta = 0.348271$. In Figure 6.4 we can see again the survival curve using the above parameters (dashed line) plotted together with the KM estimate and the curve using the "true" parameters from Table 6.2. We can see that the dotted line falls within the 95% confidence limits, exactly as before. The likelihood ratio test gives

$$-2\log\left[\frac{L_0(T_2, 0.00153197, 0.348271)}{L_0(T_2, 0.00236544, 0.28971)}\right] = 3.34736$$

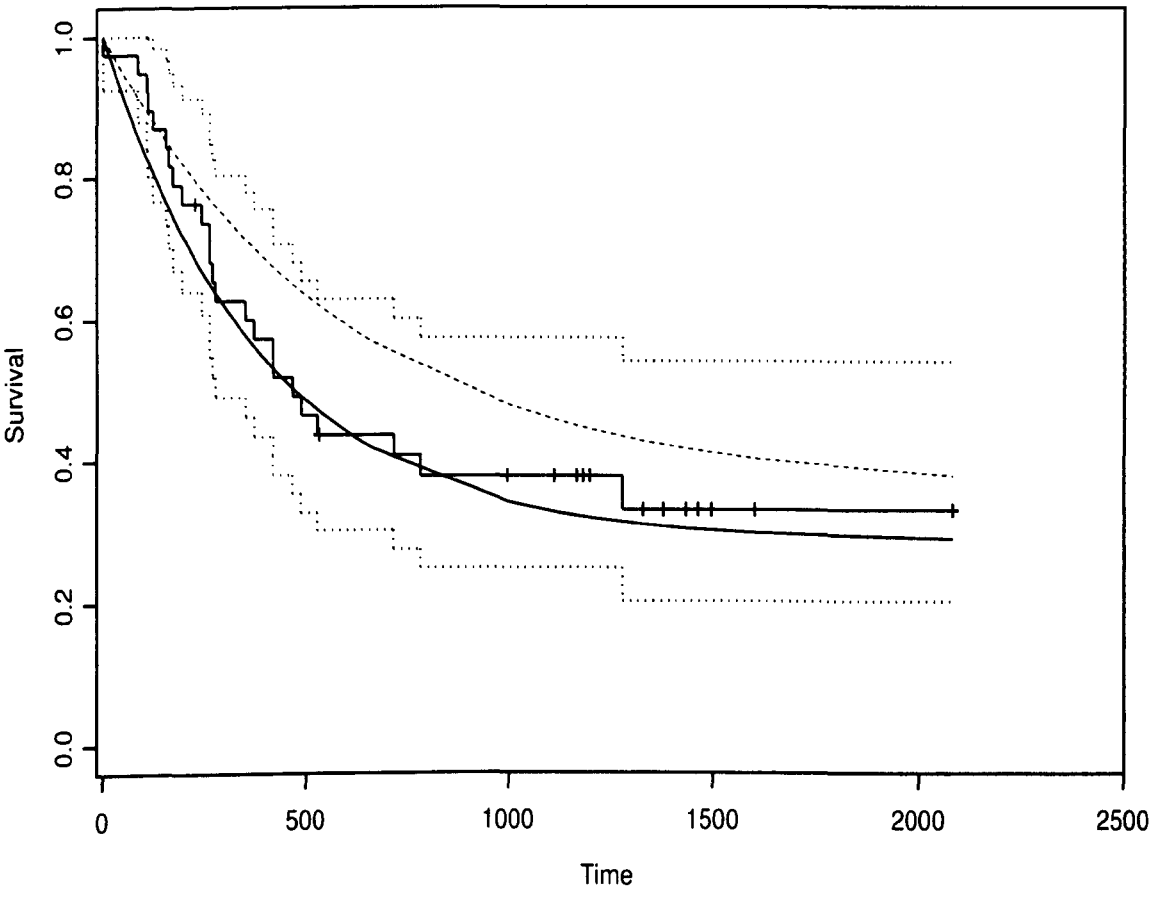

Figure 6.3: Estimated survival curve using $\hat{\delta}_2$.

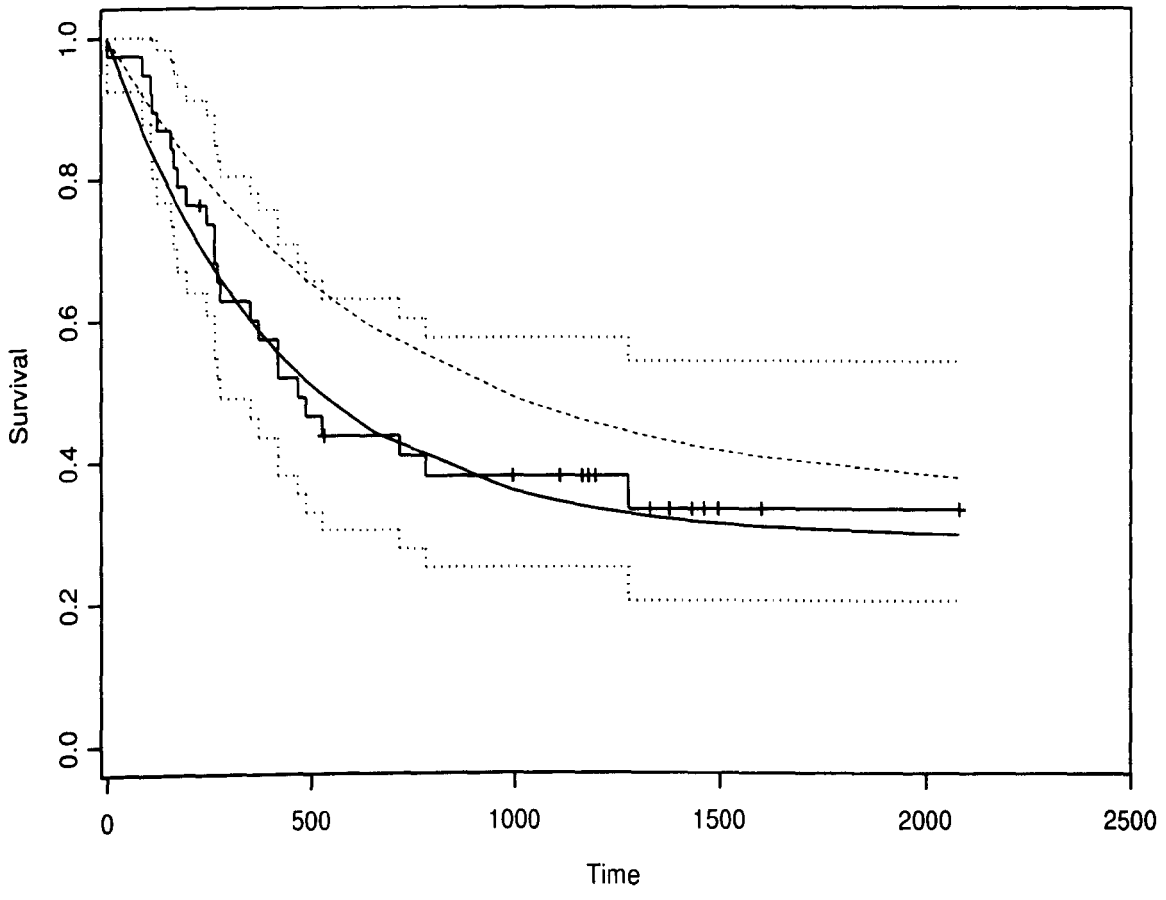

Figure 6.4: Estimated survival curve using $\hat{\delta}_3$.

which is still not significant, exactly as we would expect.

Finally, in Figure 6.5 we plot all the survival curves estimated using the three possible ways described above. This is to illustrate that the procedure with which we estimate δ actually has little impact on the estimation of the survival curve. All these curves are very close to each other and they are all considered indistinguishable from the "true" one, with the likelihood ratio test being our criterion. It is important to mention again that we have used the exact model, and not the approximate one. This made the calculations more difficult; we maximized all the likelihood functions using the mathematical software MATHEMATICA.

6.2 Modeling two types of Censoring

In chapter 5 we considered the censoring at the end of the study to be uninformative while censoring that happened during the trial was considered to be informative. This is only a special case in which we are able to make a distinction between two types of censoring, where the major criterion was the time the censorings occurred. The general case would be when we were be able to say which censorings are informative and which are not.

In order to be able to model the general case, we introduce an additional indicator variable

$$W = \begin{cases} 1, & \dots \text{ when informative censoring} \\ 0, & \dots \text{ when non-informative censoring} \end{cases}$$

which shows us which censored times are considered informative and which are not. This

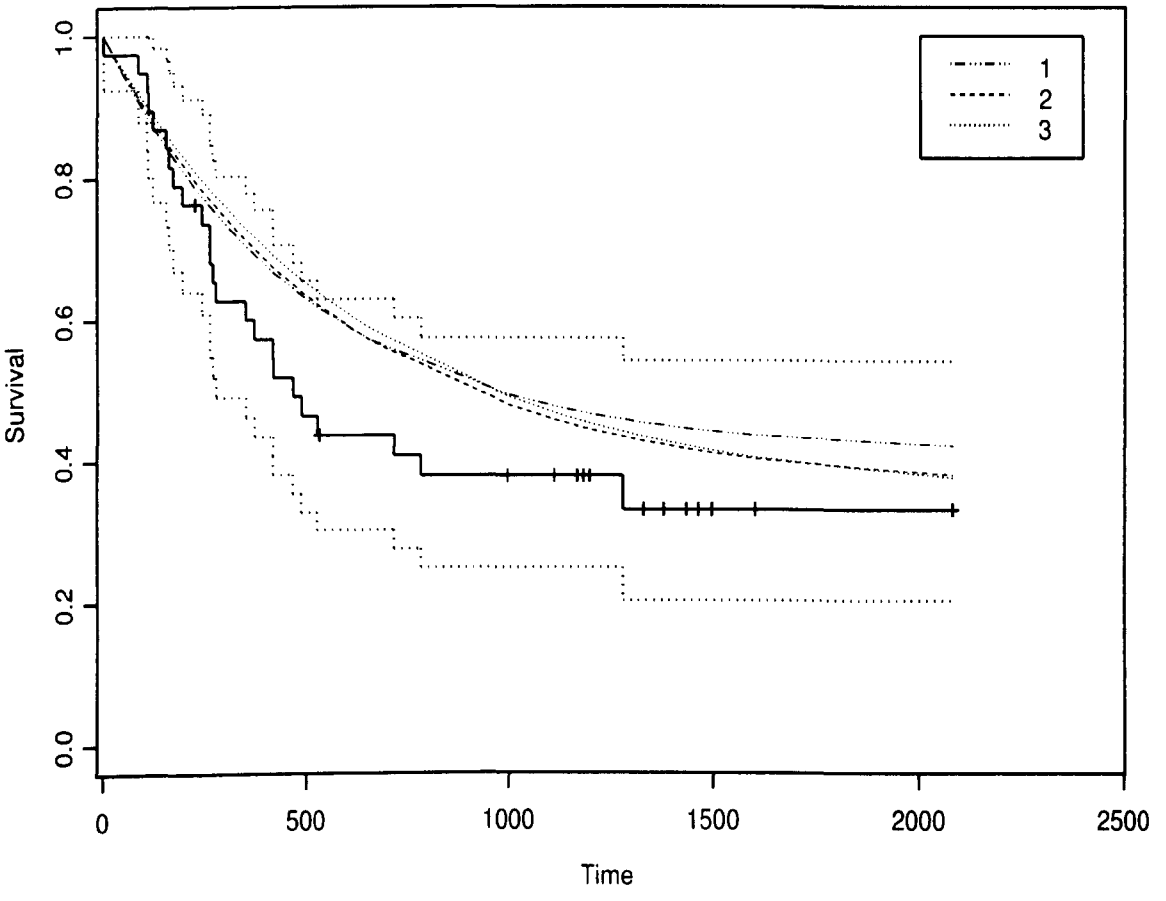

Figure 6.5: All the estimates of the survival curves for $\hat{\delta}_1$,$\hat{\delta}_2$ and $\hat{\delta}_3$.

implies that some additional information will be available with the data, explaining the reasons for patients being censored and helping us in that way to label the censored times. In a clinical trial, for instance, we may be able to find out the reason for censoring. If the patient withdraw for a non-medical reason (eg. his medical records were lost) we would define $W = 0$. If the patient withdrew because of adverse side effects, we would define $W = 1$.

Therefore the model takes the following form. We assume that each patient i has a potential failure time T_i and a potential censored time C_i, and we observe $Y_i = \min\{T_i, C_i\}$ and an indicator variable I_i, exactly as before. Additionally, we assume that each patient has a value δ_i, which is the level of dependence between T_i and C_i, and we assume that it is known. In this case we allow each patient to have a different value of δ_i, and we are able to "observe" it only when $w_i = 1$, meaning that in this case we may allow δ_i to be different from zero in performing the sensitivity analysis. The case we studied so far where all the patients have the same value of δ, is a special case of the model we are considering now.

Following the above, our initial assumption about the conditional distribution can be slightly modified to be

$$Pr(C = c|T = t, W = w) = f_C\big(c, \gamma + \delta w B(t, \theta)\big), \tag{6.10}$$

involving in that way the indicator variable W. Note that if $W = 0$, T and C are conditionally independent. The joint density becomes

$$f_{T,C,W} = f_C\big(c, \gamma + \delta w B(t, \theta)\big) f_T(t, \theta) Pr(w), \tag{6.11}$$

where

$$Pr(W = w) = \begin{cases} \pi, & \ldots \quad \text{when} \quad w = 0 \\ 1 - \pi, & \ldots \quad \text{when} \quad w = 1 \end{cases}$$

and π is the proportion of non informative censoring in the sample. A very important assumption in (6.11) is the independence between W and T. This means that the reason for which a patient is censored does not imply anything about the exact failure time of this patient, which we admit is a rather restrictive assumption. Heitjan & Rubin(1991) in a similar way model "coarse" data, but use the conditional distribution of the indicator variable given the failure time. We will explore the case where we have dependence later on in this chapter, but initially let's see what happens in this simple case.

The probabilities now become, assuming δ is small,

$$\begin{aligned} P(T = t, T < C) &= P(T = t, T < C, W = 0) + P(T = t, T < C, W = 1) \\ &= Pr(W = 0) f_T(t, \theta) S_C(t, \gamma) \\ &\quad + Pr(W = 1) f_T(t, \theta) S_C(t, \gamma) \left[1 - \delta B(t, \theta) \frac{\partial H_C(t, \gamma)}{\partial \gamma} \right] \\ &= f_T(t, \theta) S_C(t, \gamma) \left[1 - \delta Pr(W = 1) B(t, \theta) \frac{\partial H_C(t, \gamma)}{\partial \gamma} \right] \end{aligned} \tag{6.12}$$

and

$$\begin{aligned} P(C = c, C < T, W = w) &= Pr(w) \int_c^\infty f_C\big(c, \gamma + \delta w B(t, \theta)\big) f_T(t, \theta) dt \\ &\simeq Pr(w) f_C(c, \gamma) S_T(c, \theta) \left[1 + \delta w \mu(c, \theta) \frac{\partial \log f_C(c, \gamma)}{\partial \gamma} \right]. \end{aligned} \tag{6.13}$$

Finally the likelihood takes the form

$$L''_{\delta,W} = \prod_{i=1}^{n} Pr\big(T = t, T < C\big)^{I_i} Pr\big(C = t, C < T, W = 1\big)^{(1-I_i)W_i} Pr\big(C = t, C < T, W = 0\big)^{(1-I_i)(1-W_i)}, \tag{6.14}$$

and the log-likelihood becomes

$$LL''_{\delta,W} = LL_0 + \sum_{i=1}^{n} \left\{ \delta_i \left[W_i(1 - I_i)\mu(t_i, \theta) \frac{\partial \log f_C(t_i, \gamma)}{\partial \gamma} - I_i Pr(W = 1) B(t_i, \theta) \frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \right] \right\} \tag{6.15}$$

where LL_0 is the log-likelihood in the case of independence. The above expression will provide us with our estimates.

This case gives us an idea of what to expect from a model with two types of censoring. The independence estimate of the parameter of interest remains the same, and the correction factor is slightly modified, including, in a way, a proportion of the old correction factor defined by the number of informative censored cases. If in (6.15) we assume that all the censored observations are informative and that all the individuals have the same value for δ_i then we go back to our original model.

As we mentioned before, the above way of modeling raises the question of how reasonable the assumption of independence between T and W is. In other words is it reasonable to say that the indicator of informative censoring does not imply anything at all about the exact failure time of a patient? If there is a correct answer then this is probably no. This is because when we assume some kind of correlation this means that the failure times follow a pattern and they do not come completely at random. Therefore, an important thing we

need to consider now is the conditional distribution of W given T.

If W and T are correlated, the model becomes very much more complicated, and we merely indicate here a possible approach. If we take

$$f_{W|T}(w|t) = \begin{cases} \pi_t, & \ldots \quad \text{when} \quad w = 0 \\ 1 - \pi_t, & \ldots \quad \text{when} \quad w = 1 \end{cases}$$

as the conditional distribution of W given T, the joint distribution takes the form

$$f_{T,C,W}(t,c,w) \;=\; f_C\big(c, \gamma + \delta w B(t,\theta)\big) f_{W|T}(w|t) f_T(t,\theta). \tag{6.16}$$

Here, π_t is a function of t, where: $\pi_t : \mathbf{R}^+ \to [0,1]$. Hence, the condition distribution of W given T can be rewritten as

$$f_{W|T}(w|t) \;=\; \pi_t^{1-w}\Big[1 - \pi_t\Big]^w, \tag{6.17}$$

and the joint of (6.16) finally takes the form

$$\begin{aligned} f_{T,C,W}(t,c,w) &= f_C\big(c, \gamma + \delta w B(t,\theta)\big) f_T(t,\theta) \pi_t^{1-w}\Big[1 - \pi_t\Big]^w \\ &= f_C(c,\gamma) f_T(t,\theta) \pi_t^{1-w}\Big[1 - \pi_t\Big]^w \Big[1 + \delta w B(t,\theta) \frac{\partial \log f_C(c,\gamma)}{\partial \gamma}\Big]. \end{aligned} \tag{6.18}$$

Now expression (6.12) takes the form

$$P(T = t, T < C) \;=\; f_T(t,\theta) S_C\big(t,\gamma\big)\Big[\, 1 - \delta\Big(1 - \pi_t\Big) B(t,\theta) \frac{\partial H_C(t,\gamma)}{\partial \gamma}\Big], \tag{6.19}$$

while (6.13) becomes

$$P(C = c, C < T, W = w) \;=\; \int_c^\infty f_C\big(c, \gamma + \delta w B(t,\theta)\big) f_{W|T}(w|t) f_T(t,\theta) dt. \tag{6.20}$$

The above two equations will help us construct the likelihood function in a similar way like in (6.14).

For this to be feasible, the function π_t needs to be fully specified. The choices we have is an important question. In principal, $\pi_t : \mathbf{R}^+ \to [0,1]$ can be any continuous function without any further restrictions. However, we can restrict π_t to be monotonic, and the idea behind this is the following. If we assume that we have informative censored times, and we believe that the exact failure times may be close to the observed censored time (a possible positive correlation), then π_t should be small for relatively small values of t. Therefore, a choice of an increasing function of t should be appropriate. On the other hand, if we believe that the exact failure times are not close to the observed censored times, then π_t should be small for relatively large values of t. This means that a decreasing function of t would be appropriate.

Even with the assumption that π_t is monotonic, there is still a wide range of possibilities, and the choice must depend on the particular circumstances in the study and what is known about the prognosis of patients who are censored. How such information should be used, and how sensitively estimated survival parameters are to the choice of π_t, remain topics for further research. If a reasonable estimate of π_t is available, however, then the above expression can be used to construct an appropriate likelihood function.

Chapter 7

Semi–Parametric Approach

7.1 Modified Partial Likelihood

In the previous chapters we worked with the full likelihood, and we tried to explore all the possible outcomes, under different assumptions about the p.d.f. of the T and C processes. If we do not choose any known distribution, the prevailing assumption would be PH. Cox (1972) proposed that the hazard function is proportional to some other function, the baseline hazard function,

$$h(t, \theta_i) = e^{\theta_i} h^*(t)$$

which depends only on time t. Parameter θ_i is a linear combination of a set of explanatory variables, which follow each individual that participates in the trial. Under the assumption of independence between the failure and the censoring times, Cox introduced the Partial Likelihood (PL)

$$PL = \prod_i \frac{h(t, \theta_i)}{\sum_{\ell \in \mathcal{R}(t_{(i)})} h(t, \theta_\ell)} = \prod_i \frac{e^{\theta_i}}{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\theta_\ell}} = \prod_i \frac{e^{\beta' \mathbf{x}_i}}{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\beta' \mathbf{x}_\ell}}$$

where β is the vector of parameters and $\mathbf{x}_i$ is the vector of explanatory variables of the i^{th} individual. In the above product, only failure times are considered, and $\mathcal{R}(t_{(i)})$ is the risk set at time $t_{(i)}$, where $t_{(i)}$ is the i^{th} ordered failure time. The independence assumption implies that only the failure times contribute to the estimation of the parameters of interest, where the censored times are only part of the risk sets. The most important property of the Partial likelihood is the fact that the baseline hazard function finally cancels out, making the calculations much simpler. This is the major advantage of the partial likelihood, compared to the full likelihood, and what made it so widely used.

The question that is raised in our research is what happens when we depart from the initial assumption of independence. We have already seen how we model the conditional distribution $f_{C|T}(c,\gamma) = f_C\big(c, \gamma + \delta B(t,\theta)\big)$, where δ is the measure of dependence. Therefore, due to the potential dependence that arises from the presence of δ, we claim that even the censored times contribute information in the estimation process. We assume that the basic idea of the partial likelihood remains the same, and we propose two, rather important, changes. First of all, considering that C is a proper "failure" process, for failure other than the one under investigation, we introduce in our likelihood a new term which is simply the PL of the C-process. For $\delta \neq 0$, this extra term contributes an amount of information in the estimation process of the parameter of interest. The second and equally important change is that the hazard functions are now considered under the presence of two risks, T and C. Therefore, allowing the sub–hazards to be the hazard functions of the

two processes, we define the Modified Partial Likelihood (MPL) to takes the form

$$MPL = \prod_{i=1}^{r} \frac{h_T^{\sharp}(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}(t_{(i)})} h_T^{\sharp}(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_\ell)} \prod_{j=1}^{k} \frac{h_C^{\sharp}(t_j; \mathbf{v}, \mathbf{u}, \mathbf{x}_j)}{\sum_{q \in \mathcal{R}(t_{(j)})} h_C^{\sharp}(t_j, ; \mathbf{v}, \mathbf{u}, \mathbf{x}_q)} \tag{7.1}$$

where $\mathbf{v}$ and $\mathbf{u}$ are the vectors of the parameters of the T and C processes respectively, and $\mathbf{x}_i$ the vector of explanatory variables of the i^{th} individual. The above expression is divided into two products, the first one being over the r failure times and the second one being over the k censored times. $h_T^{\sharp}(t; \mathbf{v}, \mathbf{u}, \mathbf{x})$ is the hazard function of T process in the presence of the C process, which in competing risks' terminology is the sub-hazard function, while $h_C^{\sharp}(t; \mathbf{v}, \mathbf{u}, \mathbf{x}_i)$ is defined equivalently as the sub-hazard of the C process. These functions are different to the marginal hazard functions that Cox used in the partial likelihood, and they are equal to each other only when $\delta = 0$.

The form of these functions is

$$\begin{aligned} h_T^{\sharp}(t; \mathbf{v}, \mathbf{u}, \mathbf{x}) &= \lim_{\Delta t \to 0^+} \frac{Pr(t \le T < t + \Delta t | T \ge t, C \ge t, \mathbf{x})}{\Delta t} \\ &= \frac{f_T^{\sharp}(t, \mathbf{v}'\mathbf{x})}{S_{T,C}(t; \mathbf{v}, \mathbf{u}, \mathbf{x})} = h_T(t, \mathbf{v}'\mathbf{x}) \left[1 + \delta \frac{H_C(t, \mathbf{u}'\mathbf{x})}{h_T(t, \mathbf{v}'\mathbf{x})} \frac{\partial \mu(t, \mathbf{v}'\mathbf{x})}{\partial t} \right] \end{aligned} \tag{7.2}$$

and

$$h_C^{\sharp}(t; \mathbf{v}, \mathbf{u}, \mathbf{x}) = \frac{f_C^{\sharp}(t, \mathbf{u}'\mathbf{x})}{S_{T,C}(t; \mathbf{v}, \mathbf{u}, \mathbf{x})} = h_C(t, \mathbf{u}'\mathbf{x}) \Big[1 + \delta \mu(t, \mathbf{v}'\mathbf{x}) \Big], \tag{7.2'}$$

where $S_{T,C}(t; \mathbf{v}, \mathbf{u}, \mathbf{x})$ is the joint survival function, $f_T^{\sharp}(t, \mathbf{v}'\mathbf{x})$ and $f_C^{\sharp}(t, \mathbf{u}'\mathbf{x})$ are the sub-densities of the two processes and $\mu(t, \mathbf{v}'\mathbf{x})$ is defined as before. How we derive the sub–density functions is included in Appendix D. We see that the sub-hazard functions are the

marginal hazard plus an extra correction factor, which is multiplied by δ. It is obvious that when $\delta = 0$, we go back to the initial Cox's assumption of independence. Hence, the MPL takes the form

$$MPL_B = \prod_{i=1}^{r} \left\{ \frac{h_T(t_i,\theta_i)\left[1+\delta\frac{H_C(t_i,\gamma_i)}{h_T(t_i,\theta_i)}\frac{\partial\mu(t,\theta_i)}{\partial t}\Big|_{t_i}\right]}{\sum_{\ell\in R_{t_{(i)}}} h_T(t_i,\theta_\ell)\left[1+\delta\frac{H_C(t_i,\gamma_\ell)}{h_T(t_i,\theta_\ell)}\frac{\partial\mu(t,\theta_\ell)}{\partial t}\Big|_{t_i}\right]} \right\} \prod_{j=1}^{k} \left\{ \frac{h_C(t_j,\gamma_j)\left[1+\delta\mu(t_j,\theta_j)\right]}{\sum_{q\in R_{t_{(j)}}} h_C(t_j,\gamma_q)\left[1+\delta\mu(t_j,\theta_q)\right]} \right\}$$

$$= \prod_{i=1}^{r} \left\{ \frac{e^{\mathbf{v}'\mathbf{x}_i}\left[1+\delta\frac{H_C(t_i,\mathbf{u}'\mathbf{x}_i)}{h_T(t_i,\mathbf{v}'\mathbf{x}_i)}\frac{\partial\mu(t,\mathbf{v}'\mathbf{x}_i)}{\partial t}\Big|_{t_i}\right]}{\sum_{\ell\in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}\left[1+\delta\frac{H_C(t_i,\mathbf{u}'\mathbf{x}_\ell)}{h_T(t_i,\mathbf{v}'\mathbf{x}_\ell)}\frac{\partial\mu(t,\mathbf{v}'\mathbf{x}_\ell)}{\partial t}\Big|_{t_i}\right]} \right\} \prod_{j=1}^{k} \left\{ \frac{e^{\mathbf{u}'\mathbf{x}_j}\left[1+\delta\mu(t_j,\mathbf{v}'\mathbf{x}_j)\right]}{\sum_{q\in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}\left[1+\delta\mu(t_j,\mathbf{v}'\mathbf{x}_q)\right]} \right\}$$

and the log-likelihood is

$$\begin{aligned} MPLL_B \;=\; & \sum_{i=1}^{r} \left\{ \mathbf{v}'\mathbf{x}_i + \log\left[1+\delta\frac{H_C(t_i,\mathbf{u}'\mathbf{x}_i)}{h_T(t_i,\mathbf{v}'\mathbf{x}_i)}\frac{\partial\mu(t,\mathbf{v}'\mathbf{x}_i)}{\partial t}\Big|_{t_i}\right] \right. \\ & \left. - \log\left[\sum_{\ell\in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}\left[1+\delta\frac{H_C(t_i,\mathbf{u}'\mathbf{x}_\ell)}{h_T(t_i,\mathbf{v}'\mathbf{x}_\ell)}\frac{\partial\mu(t,\mathbf{v}'\mathbf{x}_\ell)}{\partial t}\Big|_{t_i}\right]\right] \right\} \\ & + \sum_{j=1}^{k} \left\{ \mathbf{u}'\mathbf{x}_j + \log\left[1+\delta\mu(t_i,\mathbf{v}'\mathbf{x}_j)\right] - \log\left[\sum_{q\in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}\left[1+\delta\mu(t_j,\mathbf{v}'\mathbf{x}_q)\right]\right] \right\}. \end{aligned}$$

In first order approximation we get

$$\begin{aligned} MPLL_B \;=\; & \sum_{i=1}^{r} \left\{ \mathbf{v}'\mathbf{x}_i - \log\left[\sum_{\ell\in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}\right] \right\} + \sum_{j=1}^{k} \left\{ \mathbf{u}'\mathbf{x}_j - \log\left[\sum_{q\in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}\right] \right\} \\ & + \delta\left[\sum_{i=1}^{r} \left\{ \frac{H_C(t_i,\mathbf{u}'\mathbf{x}_i)}{h_T(t_i,\mathbf{v}'\mathbf{x}_i)}\frac{\partial\mu(t,\mathbf{v}'\mathbf{x}_i)}{\partial t}\Big|_{t_i} - \frac{\sum_{\ell\in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}\frac{H_C(t_i,\mathbf{u}'\mathbf{x}_i)}{h_T(t_i,\mathbf{v}'\mathbf{x}_\ell)}\frac{\partial\mu(t,\mathbf{v}'\mathbf{x}_\ell)}{\partial t}\Big|_{t_i}}{\sum_{\ell\in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}} \right\} \right. \\ & \left. + \sum_{j=1}^{k} \left\{ \mu(t_j,\mathbf{v}'\mathbf{x}_j) - \frac{\sum_{q\in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}\mu(t_j,\mathbf{v}'\mathbf{x}_q)}{\sum_{q\in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}} \right\} \right] \end{aligned} \tag{7.3}$$

where term of $\delta^2, \delta^3 \ldots$ are omitted. We observe that (7.3) is the sum of the partial log-likelihood functions of the T and C processes plus δ times a term that comes from the dependence assumption we have already made. This extra term depends on functions of both processes and on the unknown function $B(t, \theta)$.

If we now make the usual choice $B(t, \mathbf{v}'\mathbf{x}) = 1 - H_T(t, \mathbf{v}'\mathbf{x})$, we have that

$$\mu(t, \mathbf{v}'\mathbf{x}) = \frac{\int_t^\infty \left[1 - H_T(a, \mathbf{v}'\mathbf{x})\right] f_T(a, \mathbf{v}'\mathbf{x}) da}{S_T(t, \mathbf{v}'\mathbf{x})} = -H_T(t, \mathbf{v}'\mathbf{x})$$

and hence

$$\frac{\partial \mu(t, \mathbf{v}'\mathbf{x})}{\partial t} = -h_T(t, \mathbf{v}'\mathbf{x}). \tag{7.4}$$

By substituting the above in (7.3), we get the simpler expression

$$\begin{aligned} MPLL_{H_T} = & \sum_{i=1}^{r} \left\{ \mathbf{v}'\mathbf{x}_i - \log \left[\sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} \right] \right\} + \sum_{j=1}^{k} \left\{ \mathbf{u}'\mathbf{x}_j - \log \left[\sum_{q \in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q} \right] \right\} \\ & + \delta \left[\sum_{i=1}^{r} \left\{ \frac{\sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} H_C(t_i, \mathbf{u}'\mathbf{x}_\ell)}{\sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}} - H_C(t_i, \mathbf{u}'\mathbf{x}_i) \right\} \right. \\ & \left. + \sum_{j=1}^{k} \left\{ \frac{\sum_{q \in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q} H_T(t_j, \mathbf{v}'\mathbf{x}_q)}{\sum_{q \in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}} - H_T(t_j, \mathbf{v}'\mathbf{x}_j) \right\} \right]. \end{aligned} \tag{7.5}$$

The symmetry in the above formulae is a very nice result of the choice we have made for the unknown function $B(t, \mathbf{v}'\mathbf{x})$. What we are interested in, is the vector of parameters of

the T–process. Hence, if we differentiate with respect to $\mathbf{v}$, we get

$$\frac{\partial MPLL_{H_T}}{\partial \mathbf{v}} = \sum_{i=1}^{r} \left\{ \mathbf{x}_i - \frac{\sum_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell}}{\sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}} \right\}$$

$$+ \delta \left[\sum_{i=1}^{r} \left\{ \frac{\sum_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell} H_C(t_i, \mathbf{u}'\mathbf{x}_\ell) \sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} - \sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} H_C(t_i, \mathbf{u}'\mathbf{x}_\ell) \sum_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell}}{\left(\sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} \right)^2} \right\} \right.$$

$$\left. + \sum_{j=1}^{k} \left\{ \frac{\sum_{q \in R_{t_{(j)}}} \mathbf{x}_q e^{\mathbf{u}'\mathbf{x}_q} H_T(t_j, \mathbf{v}'\mathbf{x}_q)}{\sum_{q \in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}} - \mathbf{x}_j H_T(t_j, \mathbf{v}'\mathbf{x}_j) \right\} \right] \tag{7.6}$$

which is actually the derivative of Cox's partial log-likelihood plus δ times the derivative of the correction factor. As a matter of notation, we define

$$\frac{\partial MPLL_{\delta=0}}{\partial \mathbf{v}} = \frac{\partial PLL}{\partial \mathbf{v}} = \sum_{i=1}^{r} \left\{ \mathbf{x}_i - \frac{\sum_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell}}{\sum_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell}} \right\}.$$

Furthermore, using Taylor's expansion, we have

$$\left. \frac{\partial MPLL_{\delta=0}}{\partial \mathbf{v}} \right|_{\hat{\mathbf{v}}_\delta} = \left. \frac{\partial MPLL_{\delta=0}}{\partial \mathbf{v}} \right|_{\hat{\mathbf{v}}_0} + \left. \frac{\partial^2 MPLL_{\delta=0}}{\partial \mathbf{v} \partial \mathbf{v}'} \right|_{\hat{\mathbf{v}}_0} (\hat{\mathbf{v}}_\delta - \hat{\mathbf{v}}_0) \tag{7.7}$$

where obviously

$$\left. \frac{\partial MPLL_{\delta=0}}{\partial \mathbf{v}} \right|_{\hat{\mathbf{v}}_0} = \left. \frac{\partial PLL}{\partial \mathbf{v}} \right|_{\hat{\mathbf{v}}_0} = 0,$$

because $\hat{\mathbf{v}}_0$ is the MLE when $\delta = 0$. Then, the estimate of the vector $\mathbf{v}_\delta$ comes from

$$\hat{\mathbf{v}}_\delta - \hat{\mathbf{v}}_0 = \delta \left[\frac{\partial^2 MPLL_{\delta=0}}{\partial \mathbf{v} \partial \mathbf{v}'} \Big|_{\hat{\mathbf{v}}_0} \right]^{-1}$$

$$\left[\sum_{i=1}^{r} \left\{ \frac{\sum\limits_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell} H_C(t_i, \mathbf{u}'\mathbf{x}_\ell) \sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} - \sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} H_C(t_i, \mathbf{u}'\mathbf{x}_\ell) \sum\limits_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell}}{\left(\sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} \right)^2} \right\} \right.$$

$$\left. + \sum_{j=1}^{k} \left\{ \frac{\sum\limits_{q \in R_{t_{(j)}}} \mathbf{x}_q e^{\mathbf{u}'\mathbf{x}_q} H_T(t_j, \mathbf{v}'\mathbf{x}_q)}{\sum\limits_{q \in R_{t_{(j)}}} e^{\mathbf{u}'\mathbf{x}_q}} - \mathbf{x}_j H_T(t_j, \mathbf{v}'\mathbf{x}_j) \right\} \right], \tag{7.8}$$

where

$$\frac{\partial^2 MPLL_{\delta=0}}{\partial \mathbf{v} \partial \mathbf{v}'} = -\sum_{i=1}^{r} \left\{ \frac{\sum\limits_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell \mathbf{x}'_\ell e^{\mathbf{v}'\mathbf{x}_\ell} \sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} - \sum\limits_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell} \left(\sum\limits_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell} \right)'}{\left(\sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} \right)^2} \right\}$$

is a $k \times k$ matrix, where k is the total number of the parameters. Finally, the expression for the bias, using the indicator function is

$$\hat{\mathbf{v}}_\delta - \hat{\mathbf{v}}_0 = \delta \left[\frac{\partial^2 MPLL_{\delta=0}}{\partial \mathbf{v} \partial \mathbf{v}'} \Big|_{\hat{\mathbf{v}}_0} \right]^{-1}$$

$$\sum_{i=1}^{n} \left\{ I_i \frac{\sum\limits_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell} e^{\mathbf{u}'\mathbf{x}_\ell} \sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} - \sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} e^{\mathbf{u}'\mathbf{x}_\ell} \sum\limits_{\ell \in R_{t_{(i)}}} \mathbf{x}_\ell e^{\mathbf{v}'\mathbf{x}_\ell}}{\left(\sum\limits_{\ell \in R_{t_{(i)}}} e^{\mathbf{v}'\mathbf{x}_\ell} \right)^2} H_C^*(t_i) \right.$$

$$\left. + (1 - I_i) \left[\frac{\sum\limits_{q \in R_{t_{(i)}}} \mathbf{x}_q e^{\mathbf{v}'\mathbf{x}_q} e^{\mathbf{u}'\mathbf{x}_q}}{\sum\limits_{q \in R_{t_{(i)}}} e^{\mathbf{u}'\mathbf{x}_q}} - \mathbf{x}_i e^{\mathbf{v}'\mathbf{x}_i} \right] H_T^*(t_i) \right\}. \tag{7.9}$$

From the above results, we see that we finally get what we were hoping to get, an

expression for the correlation bias based on our MPL, and this will be the basis for our sensitivity analysis. On the other hand, it is true that the above calculations provide some formulas which in first site seem to be very unattractive. However, trying to describe a complicated situation like this one, we would expect to get rather big and complicated expressions.

As we said at the beginning of this section, the greatest property of the PL is that we do not need to bother with the baseline hazard function, because it cancels out due to the proportional assumption. Unfortunately, this is not a property of the MPL. In the above expressions it is obvious that although the baseline hazard functions cancel out in the $MPLL_{\delta=0} = PLL$, they are still included in the correction factor. The good thing is that the correction factor is multiplied by a small δ, indicating that even an approximation would be appropriate. In the next section we propose two ways of estimating the baseline hazard functions.

7.2 Estimation of Baseline Hazard Function

In Cox's initial work, the estimation of the baseline hazard function was not essential for the estimation of the parameters of the two processes. In the partial likelihood the term $h^*(t)$ cancels out and the procedure becomes more straight forward. In our case, an estimate of $h^*(t)$ is a necessity and we will do that by using either a step function, as proposed by Kalbfleisch and Prentice(1973), or Cox's way, where we have spikes for the hazard function at the observed failure times, proposed in Cox(1972).

To begin with, we assume that we place all the failure and censored times in ascending

order

$$0 < \ldots < t_{1.0} < \ldots < t_{2.0} < \ldots\ldots\ldots < t_{r.0},$$

where $t_{.0}$ indicates a failure time, and between the failure times are all the censored times. For example, if we assume that we have two censored times in the interval $(t_{2.0}, t_{3.0}]$ then as a matter of notation we have

$$t_{2.0} < t_{3.1} < t_{3.2} < t_{3.0}.$$

All the censored times will be presented as $t_{i.j}$ where i will denote the interval in which the censoring happened, and j will denote the ordered censored time in the interval. The failure time presented as $t_{i.0}$ will be on the right end of the i^{th} interval. We consider the simpler case where the failure times are continuous and no ties occur. Now, in order to estimate the baseline hazard function we are going to use a step function. Assuming that there is a change in the baseline hazard every time a failure occurs, we define the piecewise baseline hazard function as follows

$$h_0(t) = \begin{cases} \lambda_1 & \ldots (0, t_{1.0}] \\ \lambda_2 & \ldots (t_{1.0}, t_{2.0}] \\ \vdots & \vdots \\ \lambda_{r-1} & \ldots (t_{r-2.0}, t_{r-1.0}] \\ \lambda_r & \ldots (t_{r-1.0}, t_{r.0}] \end{cases} \tag{7.10}$$

Always assuming PH, we use the full log-likelihood function (3.3) together with (7.10) and

we have:

$$LL(t; \mathbf{v}, \mathbf{u}, \delta) = \sum_{i=1}^{r} \sum_{j=0}^{q_i} \left\{ I_j \Big[\log \lambda_i + \mathbf{v}'\mathbf{x}_{i,j} \Big] + (1 - I_j) \log \Big[h_C(t_{i,j}, \mathbf{u}'\mathbf{x}_{i,j}) \Big] - e^{\mathbf{v}'\mathbf{x}_{i,j}} \mathbf{\Lambda}_{i,j} \right.$$

$$\left. - H_C(t_{i,j}, \mathbf{u}'\mathbf{x}_{i,j}) \right\} + \delta \sum_{i=1}^{r} \sum_{j=0}^{q_i} \left\{ e^{\mathbf{v}'\mathbf{x}_{i,j}} \mathbf{\Lambda}_{i,j} \Big[H_C(t_{i,j}, \mathbf{u}'\mathbf{x}_{i,j}) - 1 + I_j \Big] - I_j H_C(t_{i,j}, \mathbf{u}'\mathbf{x}_{i,j}) \right\} \tag{7.11}$$

where

$$\mathbf{\Lambda}_{i,j} = \lambda_i \Big(t_{i,j} - t_{i-1.0} \Big) + \sum_{k=1}^{i-1} \lambda_k \Big(t_{(k.0)} - t_{(k-1.0)} \Big)$$

is the cumulative baseline hazard at time $t_{i,j}$,

$$I_j = \begin{cases} 1 \quad \ldots \quad \text{when } j = 0 \text{ (failure time)} \\ \\ 0 \quad \ldots \quad \text{when } j \neq 0 \text{ (censored time)} \end{cases}$$

is the indicator variable and q_i denotes the number of censored times in the interval $(t_{i-1.0}, t_{i.0}]$. In (7.11) we introduce a completely new notation compared to that of the compete log- likelihood function. Instead of summing $\sum_{i=1}^{n}$ over all the n observed times, we sum over all the r intervals defined in (7.10) and then over all the q_i failure and censored times in each interval. If there are no censored observations in the i^{th} interval then $q_i = 0$, and the only term that is added is the one that results from the failure time in that interval.

Now we need to find estimates for all the λ_i's. First we have

$$\begin{aligned}\frac{\partial LL(t;\mathbf{v},\mathbf{u},\delta)}{\partial\lambda_1} &= \frac{1}{\lambda_1}-\sum_{j=0}^{q_1}t_{1.j}e^{\mathbf{v}'\mathbf{x}_{1.j}}-t_{1.0}\sum_{i=2}^{r}\sum_{j=0}^{q_i}e^{\mathbf{v}'\mathbf{x}_{i.j}}\\ &+\delta\Bigg\{t_{1.0}\sum_{i=2}^{r}\sum_{j=0}^{q_i}\Big[e^{\mathbf{v}'\mathbf{x}_{i.j}}\Big[H_C(t_{i.j},\mathbf{u}'\mathbf{x}_{i.j})-1+I_j\Big]\Big]\\ &+\sum_{j=0}^{q_1}\Big[t_{1.j}e^{\mathbf{v}'\mathbf{x}_{1.j}}\Big[H_C(t_{i.j},\mathbf{u}'\mathbf{x}_{i.j})-1+I_j\Big]\Big]\Bigg\} \qquad (7.12)\end{aligned}$$

where q_1 is the number of censored observations in the first interval $(0, t_{1.0}]$. Then, the first term $\hat{\lambda}_1^0$ is

$$\hat{\lambda}_1^0=\frac{1}{\sum_{j=0}^{q_1}t_{1.j}e^{\mathbf{v}'\mathbf{x}_{1.j}}+t_{1.0}\sum_{i=2}^{r}\sum_{j=0}^{q_i}e^{\mathbf{v}'\mathbf{x}_{i.j}}} \qquad (7.13)$$

and the general term is

$$\hat{\lambda}_m^0=\frac{1}{\sum_{j=0}^{q_m}\Big(t_{m.j}-t_{m-1.0}\Big)e^{\mathbf{v}'\mathbf{x}_{m.j}}+\Big(t_{m.0}-t_{m-1.0}\Big)\sum_{i=m+1}^{r}\sum_{j=0}^{q_i}e^{\mathbf{v}'\mathbf{x}_{i.j}}}. \qquad (7.14)$$

This is an estimate of the baseline hazard function in the case of independence. From equation (7.11) we can also get the estimate of the baseline hazard when $\delta \neq 0$. This might be of some interest to check the impact of the dependence in the baseline hazard, but this is definitely of no use in our case. The reason is that the baseline is going to be included in the correction factor, which is multiplied by δ and hence independent estimates of the λ_i's is what we need, in order to proceed with our calculations. In any other case terms of order $\delta^2, \delta^3, \ldots$ are created, which we finally omit.

The above way of estimating the baseline hazard function gives a step function, which means that it is constant between two successive failure time. Cox(1972) argued that the

baseline hazard is zero, except for the set $\{t_{(i)}\}$ of instants at which failures occur. Under this assumption, the independent Full Likelihood, given the covariates of each patients is

$$L(t;\mathbf{v}) = \prod_{i=1}^{n_T} h_T(t_i, \mathbf{v}'\mathbf{x}_i) \prod_{i=1}^{n} S_T(t_i, \mathbf{v}'\mathbf{x}_i) = \prod_{i=1}^{n_T} e^{\mathbf{v}'\mathbf{x}_i} h_T^*(t_i) \prod_{i=1}^{n} e^{-e^{\mathbf{v}'\mathbf{x}_i} H_T^*(t_i)} \tag{7.15}$$

where n_T is the number of failure times and n is the total number of observations that we have. Now, suppose that

$$h_T^*(t_i) = \lambda_i$$

and

$$H_T^*(t_i) = \sum_{j=1}^{n_T} \lambda_j I(t_j \leq t_i)$$

where $I(t_j \leq t_i)$ is an indicator variable, taking the value 1 every time that the restriction in the parenthesis is satisfied and 0 otherwise. Now, the log-likelihood gives

$$LL(t;\mathbf{v}) = \sum_{i=1}^{n_T} \Big\{ \log \lambda_i + \mathbf{v}'\mathbf{x}_i \Big\} - \sum_{i=1}^{n} \Big\{ \sum_{j=1}^{n_T} e^{\mathbf{v}'\mathbf{x}_i} \lambda_j I(t_j \leq t_i) \Big\}$$

and for a general λ_j we have

$$\frac{\partial LL(t;\mathbf{v})}{\partial \lambda_j} = \frac{1}{\lambda_j} - \sum_{i=1}^{n} e^{\mathbf{v}'\mathbf{x}_i} I(t_j \leq t_i) = 0$$

which implies that

$$\hat{\lambda}_j = \Big[\sum_{\ell \in \mathcal{R}(t_{(j)})} e^{\mathbf{v}'\mathbf{x}_\ell} \Big]^{-1}. \tag{7.16}$$

Expressions (7.14) and (7.16) give estimates of the baseline hazard functions. Numerically both give very similar results. The only difference is in the assumption under which we proceed in the calculations. The first one we assume is a step function, implying that

it is constant between two successive failure times. For that reason the lengths of the intervals between the failure times are included in the calculations. The second one assumes that the baseline is zero, and that it only has spikes every time we have a failure. This procedure take only the order of events under consideration, and hence the rank statistic is of major importance. These procedures can be used to estimate the baseline of both processes.

7.3 The MPL as a result of the Estimation Process of the Independence Model

In section 5.3 we saw how we can derive the independence model based on our model which assumes dependence. We can now use this idea to justify the existence of the MPL.

Let us consider the case where we have the independence model under the PH assumption. Then we have

$$\begin{aligned} h_Y(t;\theta,\gamma) &= h_T^{\sharp}(t;\theta,\gamma) = h_T(t,\theta)\Big[1-\delta H_C(t,\gamma)\Big] \\ h_U(c;\theta,\gamma) &= h_C^{\sharp}(c;\theta,\gamma) = h_C(c,\gamma)\Big[1-\delta H_T(c,\theta)\Big] \end{aligned} \tag{7.17}$$

where Y and U are the independent risks. These expressions are derived from (7.2) and (7.2′) when $B(t,\theta) = 1 - H_T(t,\theta)$. Having the above assumptions we see that the marginal hazard functions of Y and U take a much simpler and symmetric form. Under the presence of independence, we calculate the partial likelihoods of the Y and U processes. Firstly we

have

$$PL_Y = \prod_{i=1}^{n_T} \frac{h_Y(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}(t_{(i)})} h_Y(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_\ell)} = \prod_{i=1}^{n_T} \frac{h_T(t_i; \mathbf{v}'\mathbf{x}_i)\left[1 - \delta H_C(t_i, \mathbf{u}'\mathbf{x}_i)\right]}{\sum_{\ell \in \mathcal{R}(t_{(i)})} h_T(t_i; \mathbf{v}'\mathbf{x}_\ell)\left[1 - \delta H_C(t_i, \mathbf{u}'\mathbf{x}_\ell)\right]}$$

$$= \prod_{i=1}^{n_T} \frac{e^{\mathbf{v}'\mathbf{x}_i}\left[1 - \delta H_C(t_i, \mathbf{u}'\mathbf{x}_i)\right]}{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{v}'\mathbf{x}_\ell}\left[1 - \delta H_C(t_i, \mathbf{u}'\mathbf{x}_\ell)\right]} \tag{7.18}$$

and in first order approximation, the partial log-likelihood takes the form

$$PLL_Y = \sum_{i=1}^{n_T} \left\{ \mathbf{v}'\mathbf{x}_i - \log\left[\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{v}'\mathbf{x}_\ell} \right] \right\}$$

$$+\delta \sum_{i=1}^{n_T} \left\{ \frac{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{v}'\mathbf{x}_\ell} H_C(t_i, \mathbf{u}'\mathbf{x}_\ell)}{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{v}'\mathbf{x}_\ell}} - H_C(t_i, \mathbf{u}'\mathbf{x}_i) \right\}. \tag{7.19}$$

Equivalently, for the U process we have

$$PL_U = \prod_{i=1}^{n_C} \frac{h_U(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}(t_{(i)})} h_U(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_\ell)} = \prod_{i=1}^{n_C} \frac{e^{\mathbf{u}'\mathbf{x}_i}\left[1 - \delta H_T(t_i, \mathbf{v}'\mathbf{x}_i)\right]}{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{u}'\mathbf{x}_\ell}\left[1 - \delta H_T(t_i, \mathbf{v}'\mathbf{x}_\ell)\right]} \tag{7.20}$$

and the approximate partial log-likelihood becomes

$$PLL_U = \sum_{i=1}^{n_C} \left\{ \mathbf{u}'\mathbf{x}_i - \log\left[\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{u}'\mathbf{x}_\ell} \right] \right\}$$

$$+\delta \sum_{i=1}^{n_C} \left\{ \frac{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{u}'\mathbf{x}_\ell} H_T(t_i, \mathbf{v}'\mathbf{x}_\ell)}{\sum_{\ell \in \mathcal{R}(t_{(i)})} e^{\mathbf{u}'\mathbf{x}_\ell}} - H_T(t_i, \mathbf{v}'\mathbf{x}_i) \right\}. \tag{7.21}$$

As we have already noticed in §5.3, the likelihood function of Y and U, although they are considered to be independent, are a mixture of functions of both T and C processes. Therefore, vector $\mathbf{v}$ which is our major interest, is involved in both PLL_Y and PLL_U.

Hence, our estimate $\hat{\mathbf{v}}$, is the vector that maximizes both likelihood functions at the same time, and therefore satisfies the equation

$$\frac{\partial\left[PLL_Y + PLL_U\right]}{\partial \mathbf{v}} = 0. \tag{7.22}$$

The presence of independence between Y and U allows us to add the two log-likelihood functions. Using (7.5), (7.19) and (7.21) it is obvious that

$$MPLL = PLL_Y + PLL_U. \tag{7.23}$$

This means that (7.22) becomes

$$\frac{\partial MPLL}{\partial \mathbf{v}} = 0, \tag{7.24}$$

which proves that our Modified Partial Likelihood gives exactly the same inferences about $\mathbf{v}$ with the independence model.

7.4 The Independence of Censoring at the End of the Study (The MPL Case)

Now, we extend the idea of having two different types of censoring in the case of MPL. In Cox's initial argument, it was argued that due to the independence assumption between the failure and the censored times, only the failure times contribute information to the estimation process of the parameter of interest. In this Chapter, we extended this argument by saying that if we assume dependence, then even the censored lifetimes contribute information to the estimation process of the parameter of interest, and hence we ended up with the MPL. The present case with the censoring at the end of the study, is like a

combination of the two above cases. We assume that we have two types of censored times. The first one involves the censored times that happen during the trial, which we considered to be dependent to the failure times while the second one involves the censored times that happen at the end of the study. Hence, we claim that the censored times of the first type contribute information to the estimation process while the censored times of the second type do not. So we see this case as if we use the MPL in a reduced sample

$$MPL_E = \prod_{i=1}^{r} \frac{h_T^{\sharp}(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}(t_{(i)})} h_T^{\sharp}(t_i; \mathbf{v}, \mathbf{u}, \mathbf{x}_\ell)} \prod_{j=1}^{w} \frac{h_C^{\sharp}(t_j; \mathbf{v}, \mathbf{u}, \mathbf{x}_j)}{\sum_{q \in \mathcal{R}(t_{(j)})} h_C^{\sharp}(t_j, ; \mathbf{v}, \mathbf{u}, \mathbf{x}_q)} \tag{7.25}$$

which means that the independent censored times are considered only in the risk sets of the above expression. r is the number of failures and w is the number of censored times of the first type. Hence, the estimate of $\hat{\mathbf{v}}_0$ comes from the same expression (7.9) as before, with the difference being in the number of censored observations we consider to be informative.

7.5 Example

This time in the myeloma data we use the full data set including the covariates, which are all continuous variables. Although x16 takes only integer values in the interval [9,18], we still consider it as being a continuous variable for the sake of simplicity. Another problem we have to deal with is the ties between the survival times. We manage to solve it by creating a vector of 65 Uniform random variables, U[-0.5,0.5], and added them to the original survival times, breaking the ties. After this small modification to the data set, we will work using Cox's proportional hazard model. Initially, under independence ($\delta = 0$), the estimates of the parameters are as in Table 7.1, using Cox's Regression Model

(command "coxph" in S-Plus).

	coef	exp(coef)	se(coef)	z	p
x1	1.832	6.245	0.6476	2.83	0.0047
x2	-0.120	0.887	0.0594	-2.03	0.0430
x9	0.462	1.587	0.4620	1.00	0.3200
x16	0.1397	1.149	0.1000	1.39	0.1600

Table 7.1: Estimates of the parameters when $\delta = 0$

Each patient has a different set of explanatory variables. We can calculate the Prognostic Index (PI) of each individual and we can draw the survival curves. In Figure 7.1 we plot the curves of the patients with minimum and maximum PIs along with the Kaplan–Meier estimate and the survival curve of a patient with an average PI. In order to be able to do so, we need an estimate of the baseline hazard function, and in this example we have used the Kalbfleisch and Prentice's estimate, as described in § 7.2. This figure is only for illustration purposes, to see the range of all the possible survival curves, along with the KM estimate, which is an overall estimate. If now we allow δ to depart from zero, then the vector of the parameters do not remain the same any more. Using the MPL we perform a sensitivity analysis for values of $\delta \in [-0.3, 0.3]$. As we have already proved in chapter 4, under the PH assumption we have $\rho \leq \delta$, which means that $\delta = \rho$ provides us with the worst possible case in terms of the correlation. All the changes in the parameters are included in Table 7.2 and they are graphically represented in Figure 7.2. As a result, each survival curve shifts a bit up or down, depending on the sign of δ. We choose at random one patient, for example the one with PI=2.743189. The vector of explanatory variables related to this patient is $(1.3222, 14.0, 1.6232, 9)$. In Figure 7.3 we can now see the survival

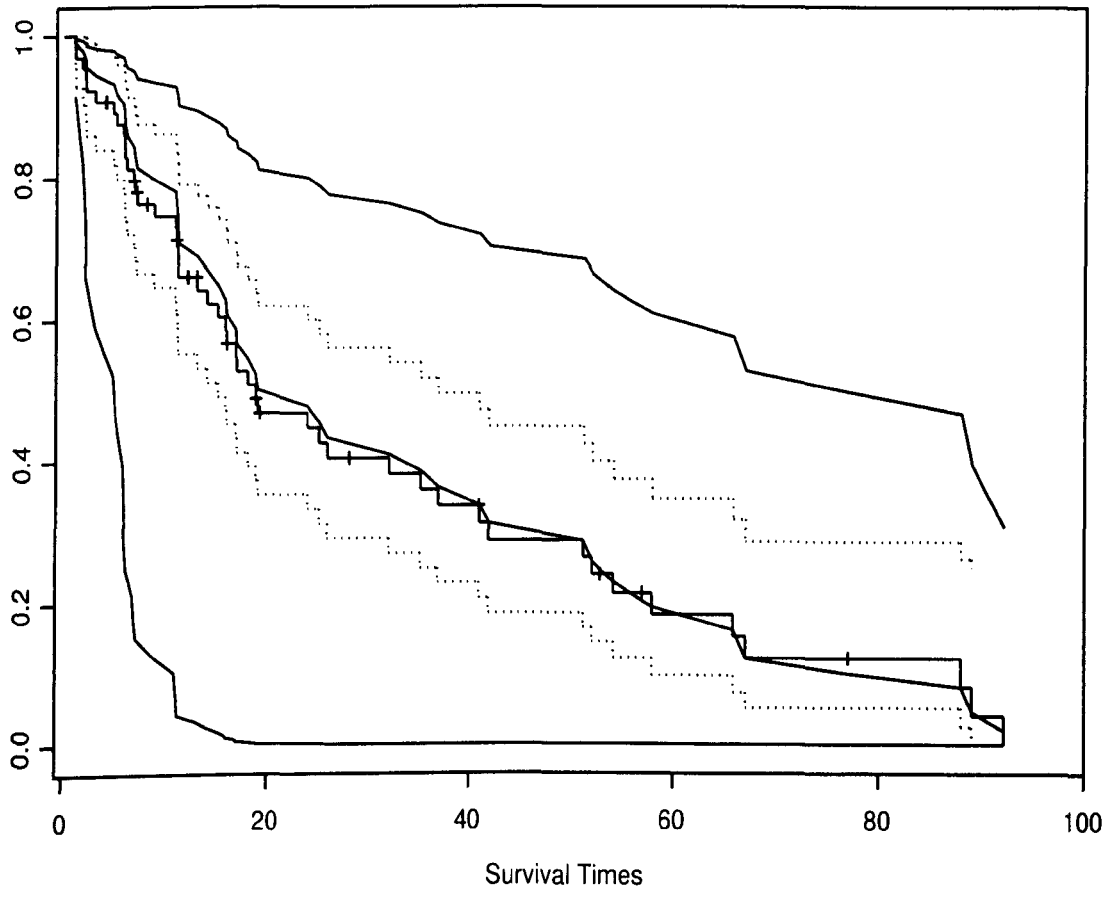

Figure 7.1: Range of Survival Curves

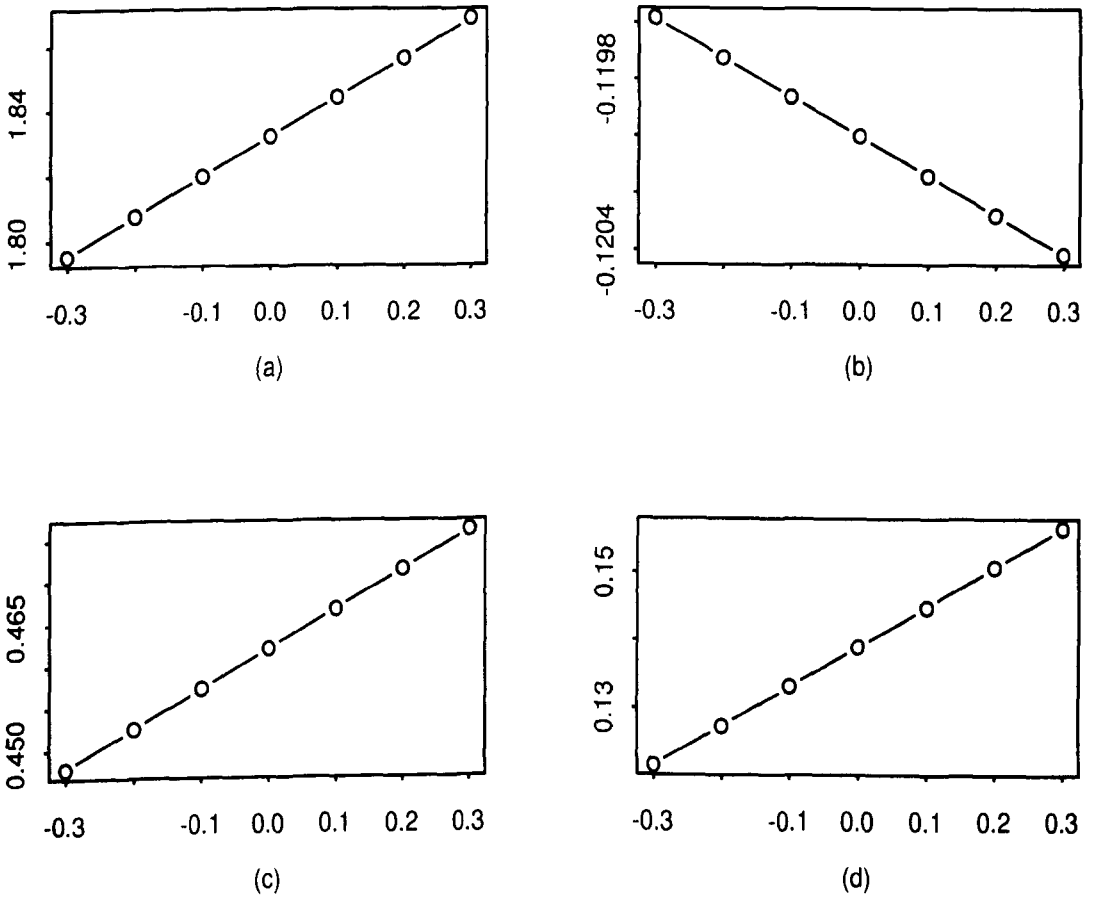

Figure 7.2: Graphical presentations of the changes in the parameters when $\delta \in [-0.3, 0.3]$.

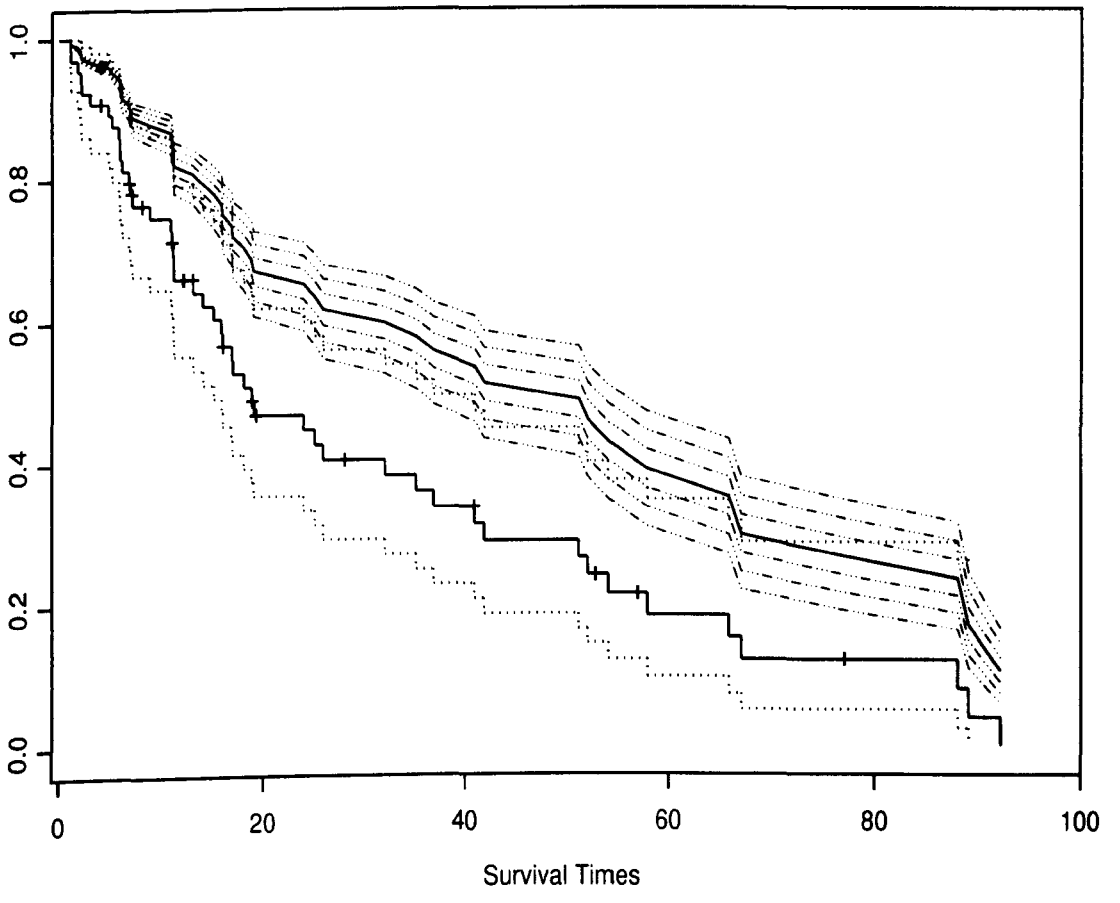

Figure 7.3: Changes in the Survival Curve for different values of δ

δ	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$
-0.3	1.7953789	-0.1195844	0.4478341	0.1214729
-0.2	1.8075859	-0.1197229	0.4525561	0.1273153
-0.1	1.8197930	-0.1198615	0.4572780	0.1331576
0	1.832	-0.120	0.462	0.139
0.1	1.8442070	-0.1201385	0.4667220	0.1448424
0.2	1.8564141	-0.1202771	0.4714439	0.1506847
0.3	1.8686211	-0.1204156	0.4761659	0.1565271

Table 7.2: The changes in the parameters for different values of δ

curve, and how it changes for all the different values of δ. For negative values it moves upwards, indicating that a negative "correlation" between the exact and censored lifetimes would yield a "better" curve with an increased chance of survival. We can see that the changes in the median survival time are quit substantial. For $\delta = 0.3$ the median survival time is around 35 while for $\delta = -0.3$ is around 55, which is over a 50% difference.

An important question is whether the parameters are significantly different from zero or not. In Table 7.1 we have both the p-values and the ratio of the parameters over the standard errors (column z). The last one is the t-statistic testing the hypothesis of the parameters being zero or not. Having a data set of 65 patients means that we have 64 degrees of freedom, and for double-sided test and $a = 0.025$ the absolute critical value is just less than 2 (1.99773). So from both the above ways we can see that only the first two parameters are significantly different from zero. In addition, we observe that v2 is very close to the critical value. Therefore, we need to investigate whether for different values of δ, v2 remains significantly different from zero or not. Hence, we get Table 7.3, in which it is clear that v2 is significant for $\delta \in [-0.3, 0.3]$, concluding that correlation does not weakens the role of v2. It seems that we do not have to do the same for v1. The values of

δ	$\mathbf{v_2}$	z
-0.3	-0.1195844	-2.013205
-0.2	-0.1197229	-2.015537
-0.1	-0.1198615	-2.01787
0	-0.120	-2.03
0.1	-0.1201385	-2.022534
0.2	-0.122771	-2.024867
0.3	-0.1204156	-2.027199

Table 7.3: Test in the significance of v2

z are far from the critical value so we do not expect huge differences for small values of δ.

Following the above test, we now use only the r.v. x1 and x2. In Table 7.4 we have

	coef	exp(coef)	se(coef)	z	p
x1	1.802	6.062	0.6279	2.87	0.0041
x2	-0.115	0.891	0.0576	-2.00	0.0460

Table 7.4: Estimates of the parameters when $\delta = 0$

the estimates of the parameters under independence. We see that v2 is still on the border of being significant or not. If we try to test that again, we see that z now takes values in $[-2.207, -1.786]$ and for almost any positive values of δ, v2 is not significantly different from zero. This indicates that for any positive dependence between the failure and the censored lifetimes, random variable x2 could be omitted in which case the only variable that remains is x1. But, without any knowledge about the value of δ, we consider v2 significant and we continue the statistical work with both variables included.

Doing exactly the same work as before, we get Table 7.5 with all the changes of the parameters for $\delta \in [-0.3, 0.3]$, and the graphical representation of this in Figure 7.4. In Figure 7.5 we see again how the survival curve is shifted, for various values of δ.

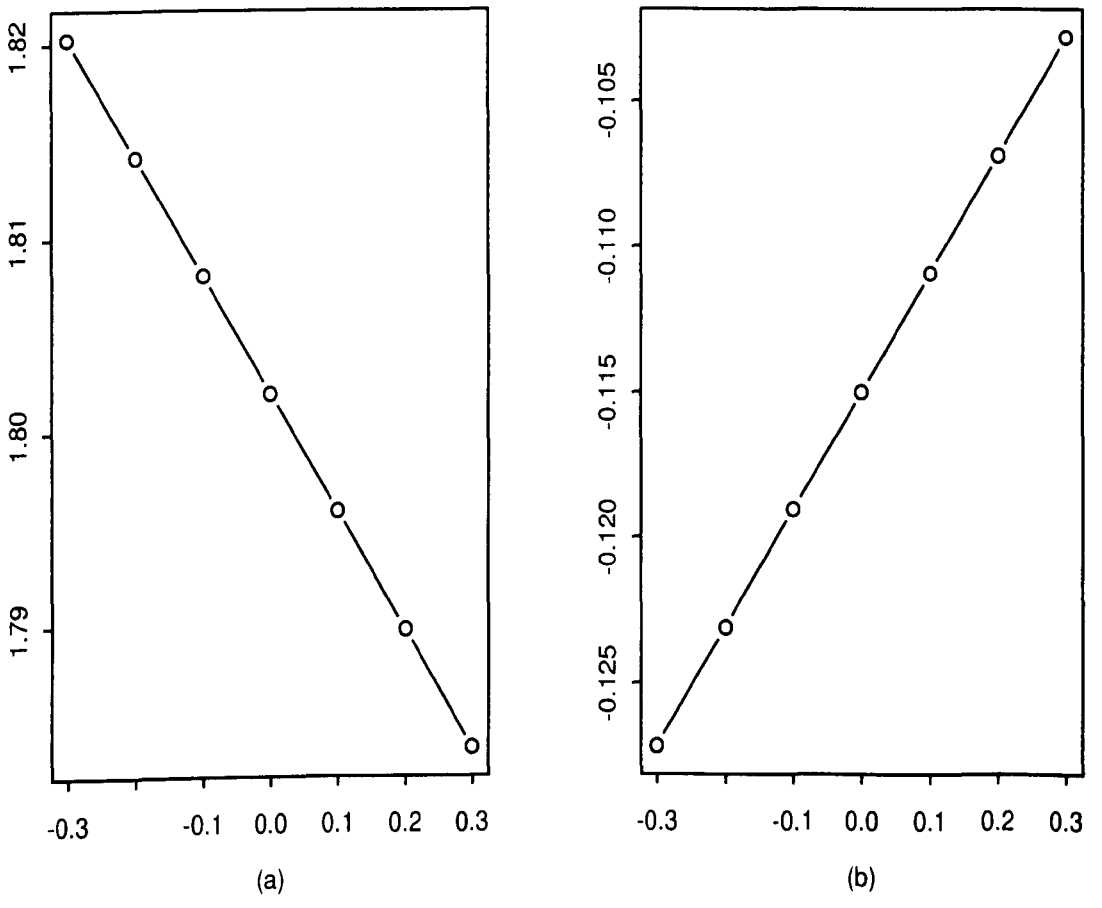

Figure 7.4: Graphical representation of the changes in the parameters v1 in figure (a) and v2 in figure (b)

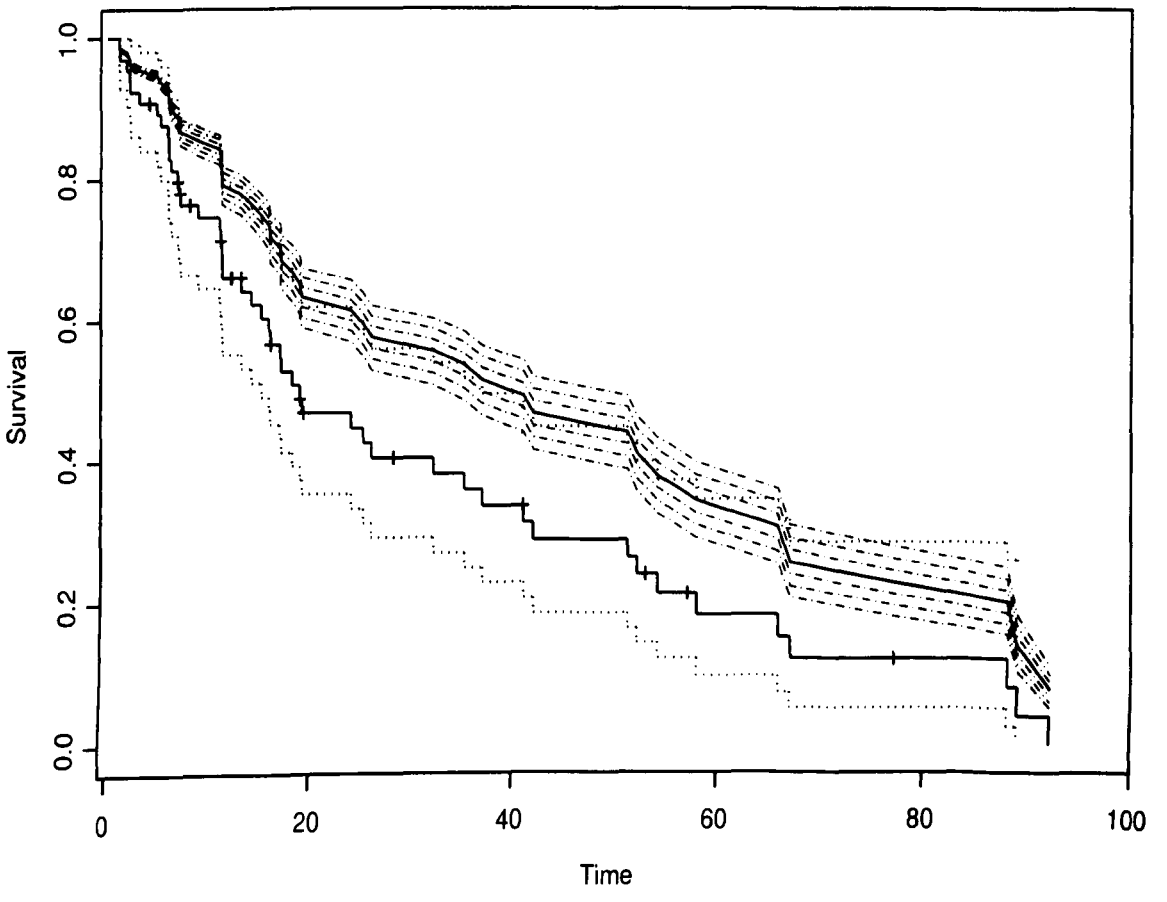

Figure 7.5: Changes in the Survival Curve for different values of δ with only the significant r.v. present

δ	$\mathbf{v}_1$	$\mathbf{v}_2$
-0.3	1.8202675	-0.1271504
-0.2	1.8141784	-0.1231003
-0.1	1.8080892	-0.1190501
0	1.802	-0.115
0.1	1.7959108	-0.1109499
0.2	1.7898216	-0.1068997
0.3	1.7837325	-0.1028496

Table 7.5: The changes in the parameters of x1 and x2

The last thing that needs to be mentioned, but equally important, is the baseline hazard function. Earlier in this chapter we referred to two possible ways of calculating this function, one due to Cox and the other one due to Kalbfleisch and Prentice. In this particular example we have used the later of the two, but as we can see in Figure 7.6 the differences are very small.

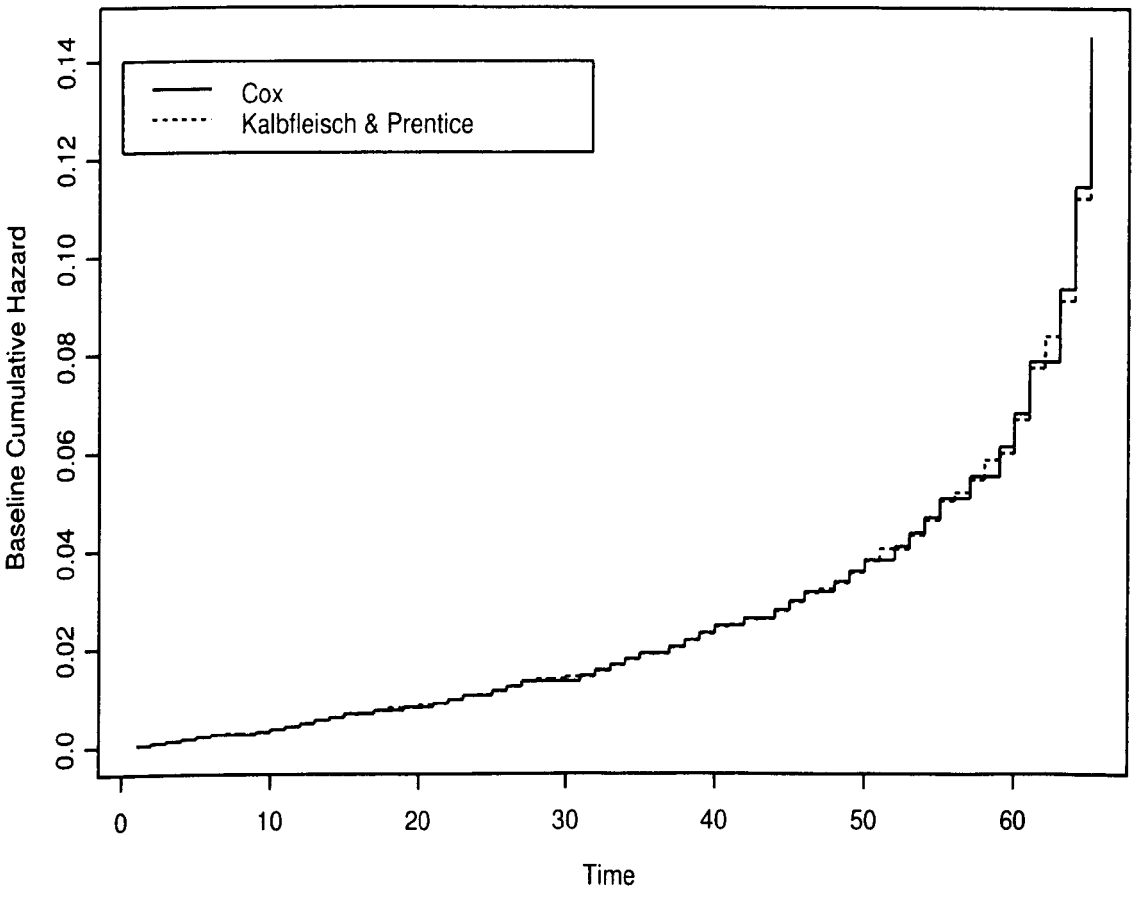

Figure 7.6: The baseline cumulative hazard function calculated with both ways.

Chapter 8

Simulation Studies

8.1 Introduction

In this chapter we focus on some examples based on simulated data, our aim being to assess the adequacy of our local approximation for small δ. Based on the Myeloma Data, we perform parametric bootstraps trying to demonstrate the use of model, using different parameterizations of the Weibull distribution.

8.2 Parametric Bootstrap

8.2.1 General Weibull

In the brief introduction we stated that we will explore different parameterizations of the Weibull distribution. It is a distribution with both proportional hazard and accelerated failure time properties. The PH property is of our main interest because as we have seen in the previous chapters, a substantial part of our research is related to PH models. A general form for the Weibull is

$$f(t, \lambda, \psi) = \lambda^{\psi} \psi t^{\psi-1} e^{-(\lambda t)^{\psi}} \tag{8.1}$$

where λ is the scale and ψ the shape parameters. Obviously, when $\psi = 1$ we have the exponential distribution.

In the example of chapter 3 we used the exponential distribution to analyze the data. If equivalently we take the log–cumulative hazard plot, Figure 8.1, we will see that the result is approximately a straight line, indicating that a Weibull model would be appropriate as well for the T–process. From the beginning of our research we have made clear that the C–process is just a nuisance process, and hence we never made an effort to explore which distribution would be most appropriate to describe the censoring mechanism. The censored times are approximately $\frac{1}{4}$ of the total number of observations in the myeloma data. Hence if we assume that the C–process is of main interest, we end up with a data set which is heavy "censored" (where the observations of the T–process are considered to be the "censored" times). If we take the log–cumulative hazard plot for the C–process as well, Figure 8.2, we see that it gives also an approximate straight line. Although we know that the plot for the C–process is based on fewer "failure" observations than the one of the T–process, we still get some useful information about the censoring process and how we should model it. Therefore, a Weibull distribution to describe the censoring mechanism seems appropriate.

For the purpose of our bootstrap examples we assume that both the processes follow a Weibull distribution of the form (8.1), $f_T(t, \theta, \alpha)$ and $f_C(c, \gamma, \beta)$. We aim to perform a sensitivity analysis on θ, the scale parameter which is the main parameter of interest, assuming that α is known. From the definition of our correlated model in chapter 3 and

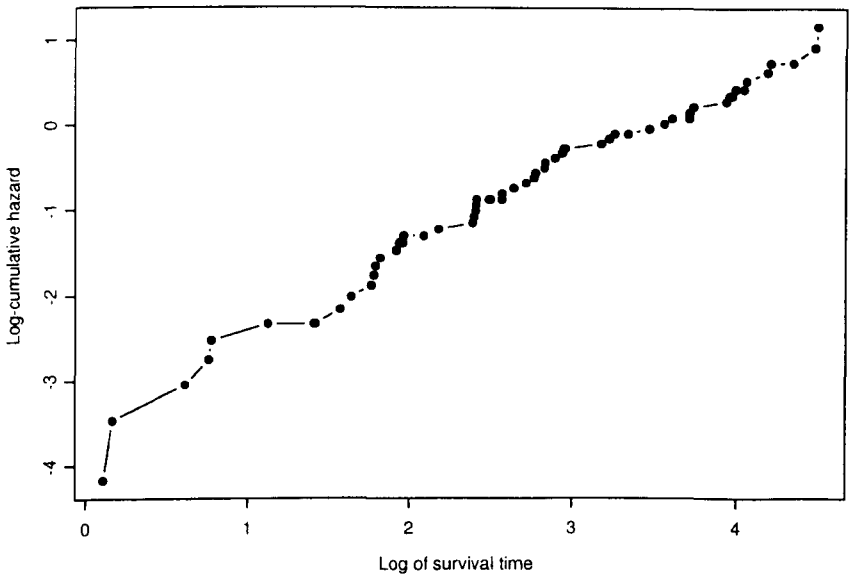

Figure 8.1: The log–cumulative hazard plot of the T–process

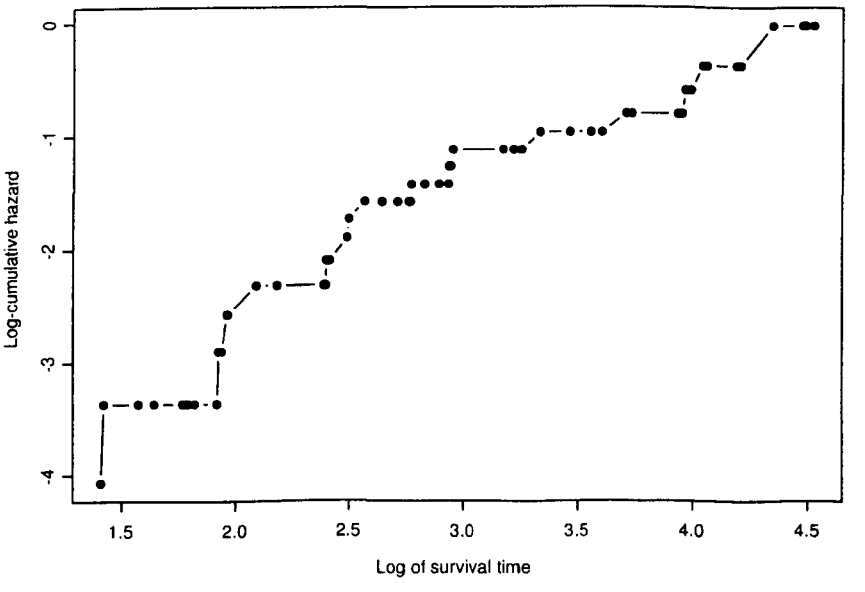

Figure 8.2: The log–cumulative hazard plot of the C–process

when $B(t,\theta) = 1 - H_T(t,\theta)$, we know that

$$\begin{aligned} Pr(C = c|T = t) &= f_C\big(c, \gamma + \delta\big[1 - H_T(t,\theta,\alpha)\big], \beta\big) \\ &= f_C(c,\gamma,\beta)\Big[1 + \delta\big[1 - H_T(t,\theta,\alpha)\big]\frac{\partial \log f_C(c,\gamma,\beta)}{\partial\gamma}\Big] \\ &\simeq f_C(c,\gamma,\beta)\exp\Big\{\delta\big[1 - H_T(t,\theta,\alpha)\big]\frac{\partial \log f_C(c,\gamma,\beta)}{\partial\gamma}\Big\}. \end{aligned} \tag{8.2}$$

The reason why we take the approximation in the above equation is because in this way we will make sure that no negative censored times will be generated. Furthermore, the conditional distribution function is

$$\begin{aligned} Pr(C < c|T = t) &= \int_0^c f_C(u,\gamma,\beta)\exp\Big\{\delta\big[1 - H_T(t,\theta,\alpha)\big]\frac{\partial \log f_C(u,\gamma,\beta)}{\partial\gamma}\Big\}du \\ &= 1 - \exp\Big\{-\Big[\gamma^\beta + \delta\beta\gamma^{\beta-1}\big[1 - h_T(t,\theta,\alpha)\big]\Big]c^\beta\Big\}, \end{aligned} \tag{8.3}$$

and the distribution function of the T–process is

$$F_T(t,\theta,\alpha) = 1 - \exp\big\{-(\theta t)^\alpha\big\}. \tag{8.4}$$

The procedure for generating a data set with survival times is the following. We generate a random number $u_{1i} \sim U[0,1]$. If we set (8.4) equal to u_{1i} we get

$$t_i = \frac{\big\{-\log(1 - u_{1i})\big\}^{\frac{1}{\alpha}}}{\theta}. \tag{8.5}$$

Similarly, if $u_{2i} \sim U[0,1]$ and with the value of t_i generated from (8.5), we set (8.3) equal

to u_{2i}, then we have

$$c_i = \left\{ \frac{-\log(1-u_{2i})}{\gamma^{\beta} \exp\left\{\delta \frac{\beta}{\gamma}\left[1 - H_T(t_i, \theta, \alpha)\right]\right\}} \right\}^{\frac{1}{\beta}}, \tag{8.6}$$

where $i = 1, \ldots, n$. In order to be able to generate censored times under the assumption of informative censoring, we need to specify a value for δ different from zero. Having the level of dependence fixed , we record $y_i = \min\{t_i, c_i\}$, the minimum of the two times, and we set the indicator variable to take the value

$$I_i = \begin{cases} 1, & \text{if} \quad t_i < c_i \\ 0, & \text{if} \quad c_i < t_i. \end{cases}$$

There are also some quantities that we are really interested to test. In chapter 4 we developed an expression for the statistical expectation of the bias under PH and when $B(t, \theta, \alpha) = 1 - H_T(t, \theta, \alpha)$, and a formula for the upper bound with PH as the only assumption, leaving $B(t, \theta, \alpha)$ arbitrary. At that stage we assumed PH to have the property

$$\frac{\partial H_T(t, \theta)}{\partial \theta} = H_T(t, \theta)$$

which is true in many cases. Weibull has the PH property, but it doesn't posses the above one. We have

$$\frac{\partial H_T(t, \theta, \alpha)}{\partial \theta} = \frac{\alpha}{\theta} H_T(t, \theta, \alpha)$$

and hence we need to make some adjustments to the formulas from chapter 4. More

specifically the expression of the bias now takes the form

$$\hat{\theta}_\delta - \hat{\theta}_0 \simeq \frac{\delta n}{\imath(\theta_0)} \frac{\alpha\beta}{\theta\gamma} \sum_{i=1}^{n} \Big[H_C(t_i,\gamma) H_T(t_i,\theta) - (1 - I_i) H_T(t_i,\theta) \Big]$$

and the expectation of the bias in (5.8) becomes

$$E\big[\hat{\theta}_\delta - \hat{\theta}_0\big] \simeq \frac{\delta n}{\imath(\theta_0)} \frac{\alpha\beta}{\theta\gamma} E_T\Big[H_C(t,\gamma) S_C(t,\gamma) \Big], \tag{8.7}$$

where

$$\imath(\theta_0) = \frac{\alpha}{\theta^2} \sum_{i=1}^{n} \Big[I_i + (\alpha - 1)(\theta t_i)^\alpha \Big].$$

Equivalently, when we do not make any assumptions about $B(t,\theta,\alpha)$ we have

$$\hat{\theta}_\delta - \hat{\theta}_0 \simeq \frac{\delta n}{\imath(\theta_0)} \frac{\beta}{\gamma} \sum_{i=1}^{n} \left[(1 - I_i) \Big[1 - H_C(t_i,\gamma) \Big] \frac{\partial \mu(t_i,\theta)}{\partial \theta} - I_i H_C(t_i,\gamma) \frac{\partial B(t_i,\theta)}{\partial \theta} \right],$$

and hence the expression (5.7) for the upper bound gives

$$\Big| E\big[\hat{\theta}_\delta - \hat{\theta}_0\big] \Big| \leq |\delta| \frac{n}{\imath(\theta_0)} \frac{\beta}{\gamma} \Big\{ E_T\Big[T_1^2(t,\theta,\gamma) \Big] \Big\}^{\frac{1}{2}}, \tag{8.8}$$

where

$$T_1(t,\theta,\gamma) = \int_0^t \Big[1 - H_C(x,\gamma) \Big] \Big[\frac{\alpha}{\theta} \big[1 - H_T(t,\theta) \big] + H_T(x,\theta) \Big] f_C(x,\gamma) dx.$$

In the particular case of the myeloma data, the estimates of the parameters under the independence assumption are $\hat{\theta}_0 = 0.0306$, $\hat{\alpha}_0 = 1.0358$, $\hat{\gamma}_0 = 0.0118$ and $\hat{\beta}_0 = 1.1028$. We observe that the shape parameters are not much different from one and hence the exponential case. Nevertheless, and despite the fact that θ is the main parameter of interest, we will still keep the Weibull assumption and the shape parameters different from

zero in order to observe any changes on them for different values of δ. Initially we will assume that it is known. However, we will obtain an estimate of α for each value of δ, trying in that way to explore if there is any impact on the sensitivity analysis of θ when we re–estimate α.

In order to illustrate our methodology we perform a parametric bootstrap on the myeloma data, based on the values of the parameters above. Additionally the correlation between the two processes is assumed to take the values $\rho = 0$, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3 and from (4.6) we can calculate the appropriate values of δ. The sample size is 65 and we generated 1000 such data sets for each value of δ. For every generated sample we calculate all the parameters under the independence assumption. In Table 8.1 we present the expected values of θ and the correction factor (CF) over the 1000 replication for each value of δ, along with their standard errors. The bias calculated for $\rho = 0$ is considered to be an estimate of the sampling bias, hence we simply subtract it from $\hat{\theta}_0$ in order to remove this source of bias. The expected values of the shape parameter α are also presented with it's standard error. In this parametric case we haven't standardized the score

ρ	δ	$E[\hat{\theta}_0 - \theta_T]$	$E[\hat{\theta}_0]$	$se[\hat{\theta}_0]$	E[CF]	se[CF]	$E[\hat{\alpha}_0]$	$se[\hat{\alpha}_0]$
0.30	0.0032	-0.0014	0.0295	0.0042	0.8756	0.2680	1.0284	0.1148
0.25	0.0027	-0.0011	0.0295	0.0043	0.8341	0.2091	1.0283	0.1172
0.20	0.0021	-0.0010	0.0296	0.0043	0.8051	0.2073	1.0294	0.1210
0.15	0.0016	-0.0007	0.0299	0.0044	0.7605	0.1931	1.0382	0.1194
0.10	0.0011	-0.0004	0.0302	0.0042	0.7269	0.1746	1.0481	0.1263
0.05	0.0005	0.0003	0.0309	0.0046	0.7002	0.1622	1.0511	0.1227
0.00	0.0000	0.0004	0.0310	0.0043	0.6577	0.1352	1.0636	0.1338

Table 8.1: Bootstrap results for the general Weibull

function of the C-process. This means that in Table 8.1 is the ordinary δ and not δ^*, the "standardized δ".

In this study our aim is to observe whether our model is capable to produce reasonable limits for the parameter of interest and for any other quantity that we might be interested in, like the median and the survival curve, when $\delta \neq 0$. The first thing someone would look at is the correlation bias. Figure 8.3 shows graphically the level of the mean bias and how good is our linear approximation. The solid line with (O) in the figure is the absolute value of the bootstrap expected correlation bias. It is calculated from the bias from Table 8.1 by subtracting the sampling bias (when $\rho = 0$). The second solid line with (Δ) is our linear approximation to the correlation bias ($\delta \times$ CF). We see that our approximation is good up to the level of $\rho = 0.2$, where the differences are due to random error, and when the correlation increases we tend to overestimate the correlation bias. It was of course expected that our methodology would work for values of δ close to zero. Nevertheless, the fact that we overestimate the correlation bias when $\rho > 0.2$ is not such a bad thing, because we know that our limiting values for θ_δ will always include the true value θ_{TR}. Maybe a choice of $\rho = 0.25$ would be more appropriate, but in our study we decided to choose $\rho = 0.3$ for illustration purposes and because it definitely provides limits that include θ_{TR}. The dotted line with (+) is a result of formula (8.7), and it shows what is our expectation of the correlation bias under the Weibull assumption. Finally the dashed line with ($\times$) comes from (8.8) and is the upper bound, calculated with no particular assumption about function $B(t, \theta, \alpha)$. Actually this is the worst possible bias we might have in our model under PH. All the above become clearer in Figure 8.4, where the actual intervals for

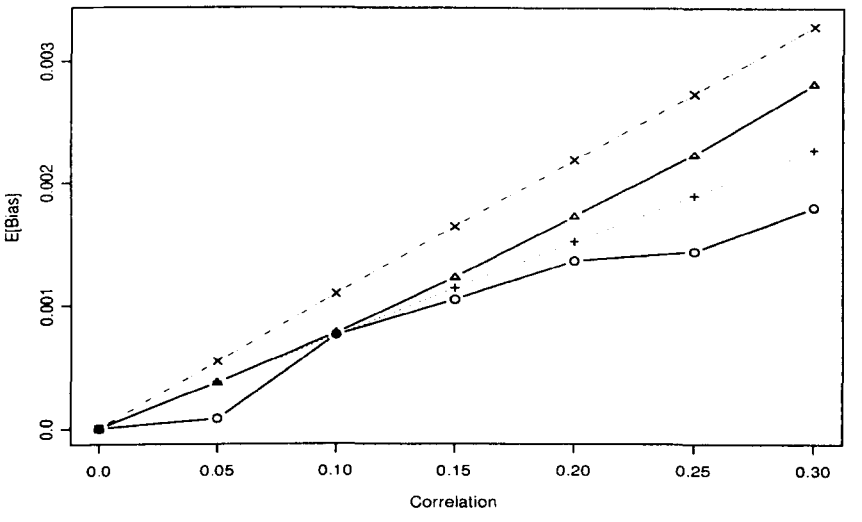

Figure 8.3: Sensitivity analysis on the expected correlation bias of θ. (O) bootstrap; (Δ) $\delta \times$ CF; (+) statistical expectation; ($\times$) upper bound.

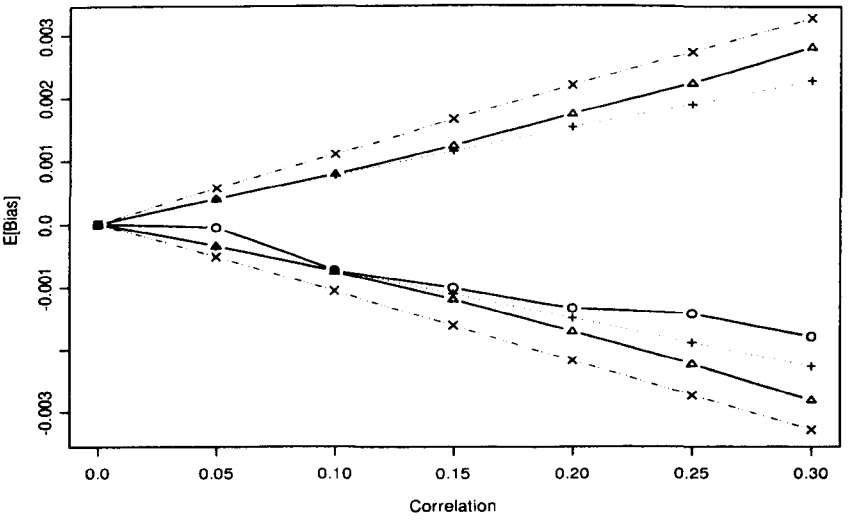

Figure 8.4: Intervals for the expected correlation bias. (O) bootstrap; (Δ) $\delta \times$ CF; (+) statistical expectation; ($\times$) upper/lower bound

positive and negative values of δ are presented. In this figure the correlation in the x–axis is the absolute value of the correlation, representing the actual level of dependence, without indicating any direction. Whether we have positive correlation or not is presented in the graph itself, where for positive δ we have a positive slope, and for negative δ we have negative slope. Our objective is that the true value of the correlation bias must lie within the intervals. In the above case we see that the bootstrap estimate of the bias is exactly where we want it to be, indicating that a sensitivity analysis over δ will produce "confidence" intervals for θ which will include θ_{TR}.

Another quantity of great importance is the median. In the Weibull case the median is given by

$$m = \frac{\left[\log(2)\right]^{1/\alpha}}{\theta}, \tag{8.9}$$

and in our particular case its true value is: $m_{TR} = 22.9406$. Keeping the same way of performing our sensitivity analysis as before we initially use the true value of α. Therefore in Figure 8.5 we perform a sensitivity analysis on the median, where again in the x–axis is the absolute value of the correlation. The straight dashed line is the exact value of the median (when $\delta = 0$) and the solid line with ($\Diamond$) is the median, calculated for positive values of δ. Our analysis will provide us with the dashed lines with (Δ), expecting the true value of the median to be included in these lines. Actually, we observe that for positive δ we get a very good approximation of the median up to a correlation of $\rho = 0.2$. After that we see that we under estimate the median, which still falls within the desired interval. This is expected because we have already stated that when $\rho > 0.2$ we over–estimate θ, a

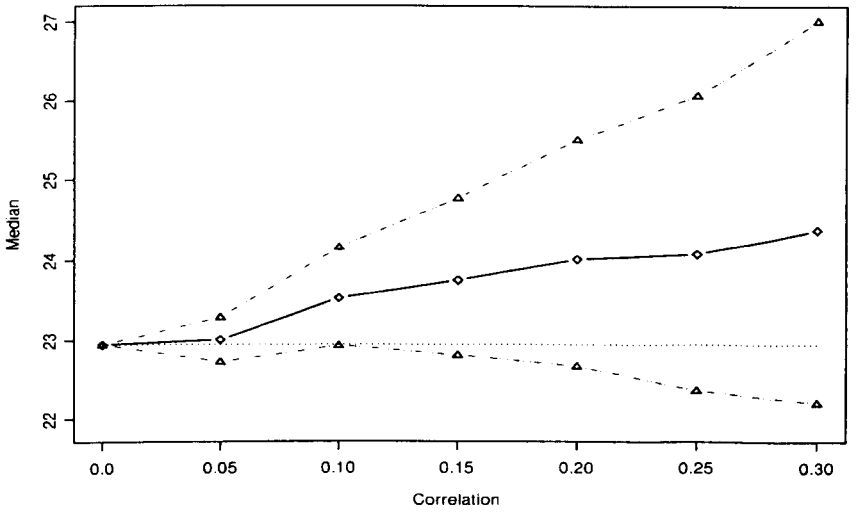

Figure 8.5: Sensitivity analysis on the median, with the shape parameter having the true value and δ taking both positive and negative values.

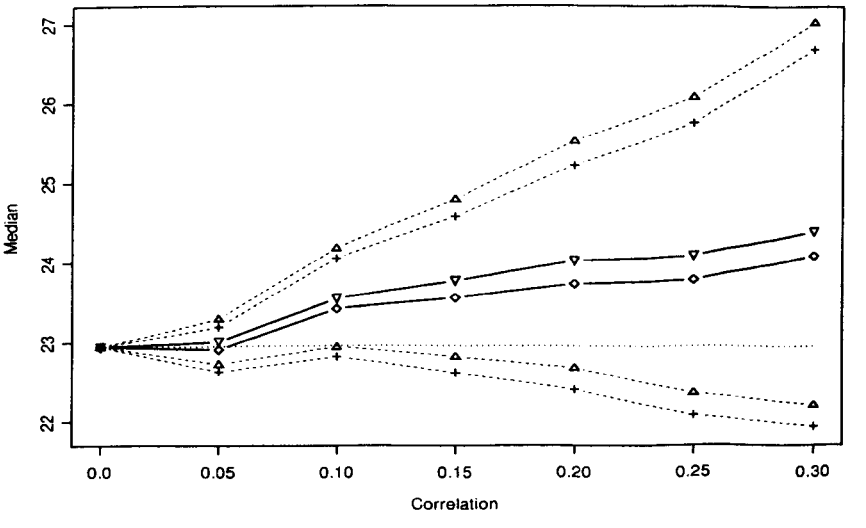

Figure 8.6: Sensitivity analysis on the median. ($\Diamond$) re–estimated α; (∇) true α; (+) limits when re–estimated α; (Δ) limits when true α.

result which immediately leads to the under -estimation of the median. Figure 8.6 is a more general graph than 8.5. It includes in addition the estimate of the median when we re–estimate α, solid line with ($\Diamond$), and the limiting lines resulted by this assumption, dashed lines with (+). This figure shows that we slightly under estimate the median, suggesting that a sensitivity analysis on both parameters might be worthwhile.

Finally, the last thing we would like to explore is the estimate of the survival curve. Figure 8.7 shows the survival curves for $\delta = 0.3$ when we re–estimate α and when we don't. In this particular case of estimating the survival curve we see that there is virtually no difference between the two curves. Hence in Figure 8.8 we perform the sensitivity analysis in the case when α is re–estimated. We see that we approximate the true survival curve, semi–dashed line, very well. The important thing though is that the true survival curve falls in the interval constructed by the curves for $\delta = \pm 0.3$.

After this bootstrap study we can have a good idea how our method works. When the parameters are more than one, we need to consider the case where we perform sensitivity analysis on all the parameters at the same time, although this would be definitely a difficult task in terms of computations. Despite that, our main objective was to explore the situation where the levels of dependence where know in advance, and then check the performance our methodology. As we expected our estimates are really good for small values of δ which correspond to $\rho \leq 0.2$. The fact that for values greater than that we tend to overestimate the parameter of interest might turn out to be in our favor if we can choose the right value for δ. This means that we can construct intervals which will definitely include θ_{TR} which is one of our major goals. Therefore, a value of δ that corresponds to $\rho = 0.25$ seems to

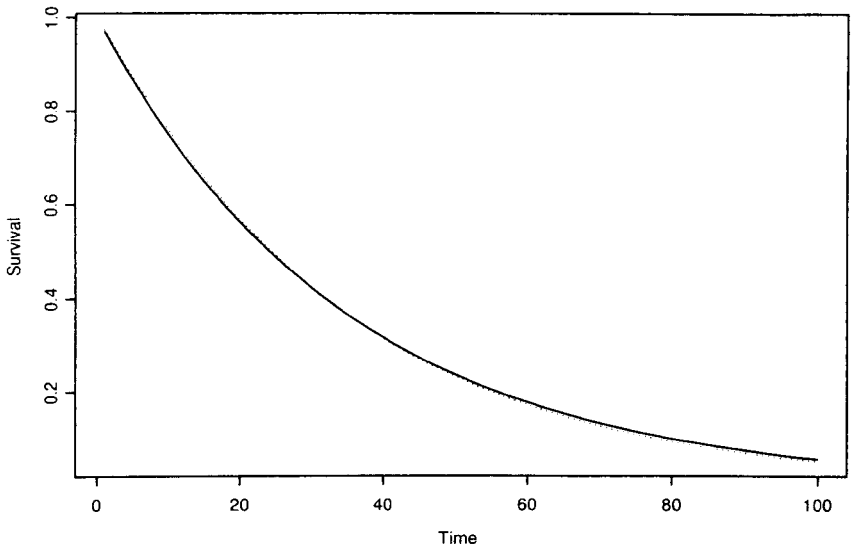

Figure 8.7: Survival curves with re–estimated (solid line) and fixed (dashed line) α, and θ_δ is taken for $\delta = 0.3$.

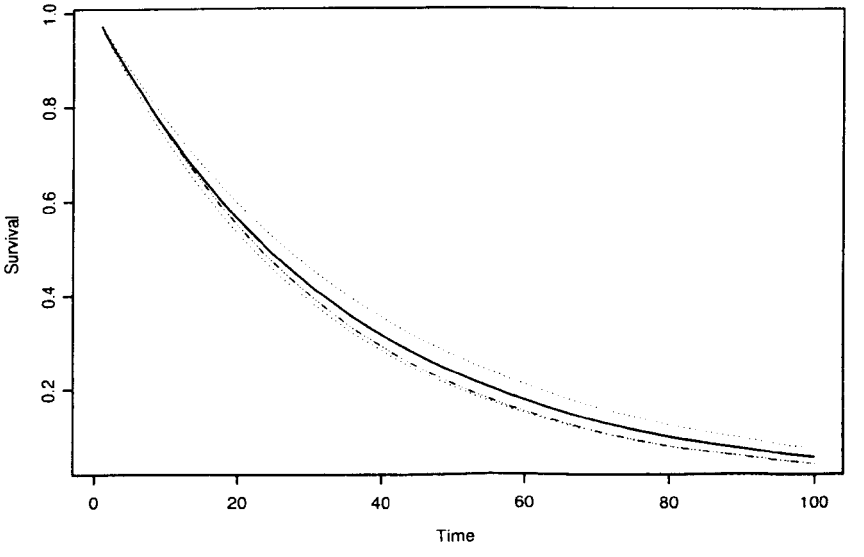

Figure 8.8: Sensitivity analysis on the survival curve with re–estimated α, for $\delta = 0.3$.

be a very good choice for our sensitivity analysis.

8.2.2 Re–parameterization of Weibull to include Median m

If we assume that m is the parameter of interest, we can use the following re-parameterization of the weibull. From (8.9) we have that

$$\theta = \frac{\left[\log(2)\right]^{\frac{1}{\alpha}}}{m}. \tag{8.10}$$

Substituting the above in (8.1) then the weibull distribution, including m, takes the form

$$f(t,m,\alpha) = \log(2)m^{-\alpha}\alpha t^{\alpha-1}\exp\Big\{-\log(2)\Big(\frac{t}{m}\Big)^{\alpha}\Big\}, \tag{8.11}$$

where the hazard function is $h_T(t,m,\alpha) = \log(2)\alpha m^{-\alpha}t^{\alpha-1}$, and the cumulative hazard is $H_T(t,m,\alpha) = \log(2)(\frac{t}{m})^{\alpha}$. The purpose of doing this is that in the case that we are interested in m, we would prefer to perform a direct sensitivity analysis on m, rather than estimate θ and then do the sensitivity analysis on the median with respect to θ.

The expression of the bias for m takes the form

$$\hat{m}_\delta - \hat{m}_0 = \frac{\delta}{\imath(m)}\sum_{i=0}^{n}\left[(1-I_i)\frac{\partial \log f_C(t_i,\gamma,\beta)}{\partial\gamma}\frac{\partial\mu(t_i,m,\alpha)}{\partial m} - I_i\frac{\partial H_C(t_i,\gamma,\beta)}{\partial\gamma}\frac{\partial B(t_i,m,\alpha)}{\partial m}\right], \tag{8.12}$$

where

$$\imath(m) = \frac{\alpha}{m^2}\sum_{i=1}^{n}\left[I_i - \log(2)(\alpha-1)\Big(\frac{t_i}{m}\Big)^{\alpha}\right].$$

It is obvious that the above re–parameterization is only for the T–process, while for the C–process we haven't changed anything, mainly due to it's secondary role. Therefore, if

we assume $B(t, m, \alpha) = 1 - H_T(t, m, \alpha)$, formula (8.12) takes the form

$$\hat{m}_\delta - \hat{m}_0 = \delta \frac{\alpha\beta}{m\gamma\iota(m)} \sum_{i=0}^{n} \left[(1 - I_i) H_T(t_i, m, \alpha) - H_T(t_i, m, \alpha) H_C(t_i, \gamma, \beta) \right]. \qquad (8.13)$$

ρ	δ	E$[\hat{m}_0 - m_{TR}]$	E$[m]$	se$[m]$	E[CF]	se[CF]	E$[\alpha]$	se$[\alpha]$
0.30	0.0032	1.1197	24.0603	3.7342	728.5192	258.5181	1.0297	0.1134
0.25	0.0027	1.3304	24.2710	3.8314	718.7331	253.9251	1.0296	0.1208
0.20	0.0021	1.1452	24.0858	3.6888	680.8189	242.5570	1.0291	0.1177
0.15	0.0016	0.9670	23.9077	3.6907	634.1373	220.6374	1.0415	0.1271
0.10	0.0011	0.6985	23.6391	3.6346	584.5753	204.0658	1.0440	0.1235
0.05	0.0005	0.6455	23.5861	3.6099	552.9551	178.8066	1.0470	0.1252
0.00	0.0000	0.03808	22.9787	3.3867	500.7274	153.4157	1.0556	0.1223

Table 8.2: Bootstrap results for the modified Weibull.

Now we perform a bootstrap study to the myeloma data, similar to the one of the previous section, using our modified Weibull distribution and the results are presented in Table 8.2. Again we assume only m is of interest and that no sensitivity analysis is performed on α. Figure 8.9 shows the limits that we construct for the median. We observe that when we know the exact value of δ we can approximate the correlation bias very well for $\rho \leq 0.2$. For $\rho > 0.2$ we overestimate the bias, something that is expected when δ becomes larger. Figure 8.10 shows the independent estimates of the median for different levels of dependence, with their differences being due to random variation.

The reason why we use this kind of modified Weibull is because we want to demonstrate a possible way of performing a direct sensitivity analysis on quantities, like the median, that are not directly included in the density function. The conclusion is that there are no major differences in our analysis whichever parameterization of the same distribution we

decide to use. The main problem that comes up is the calculation of the value of δ. This parameter is highly dependent on the way of modeling and the choice of the distributions of the two processes, even if there are different types of the same distribution. This is another reason why inferences about δ are not possible. In our case with the general and the modified Weibull, from Tables 8.1 and 8.2 we see that the values of δ are exactly the same. If we had the values of δ in more than 4 decimal places, we would have seen that there are differences. The main conclusion is that every time that we use our model, a careful calibration of the value of δ needs to be done, because similar parametric assumptions might imply completely different values of δ.

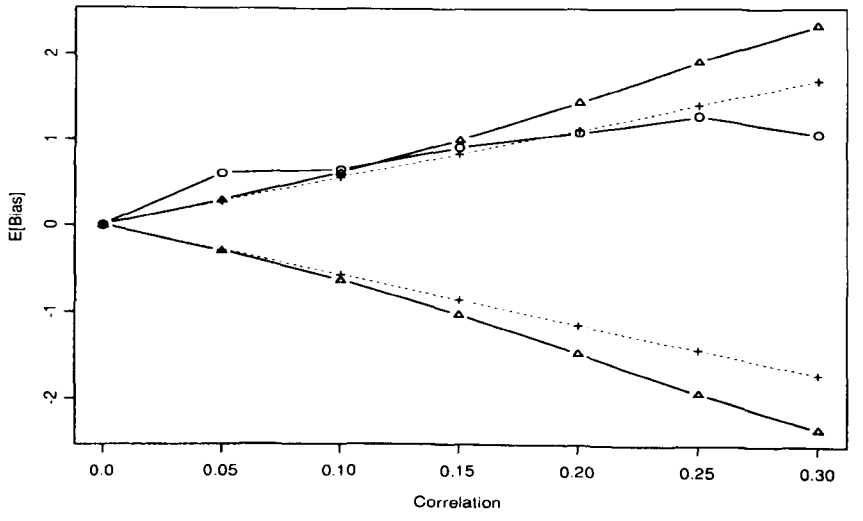

Figure 8.9: Sensitivity analysis on the bias of the median (with true α). (O) Bootstrap; ($\triangle$) $\delta \times$ CF; (+) Statistical expectation.

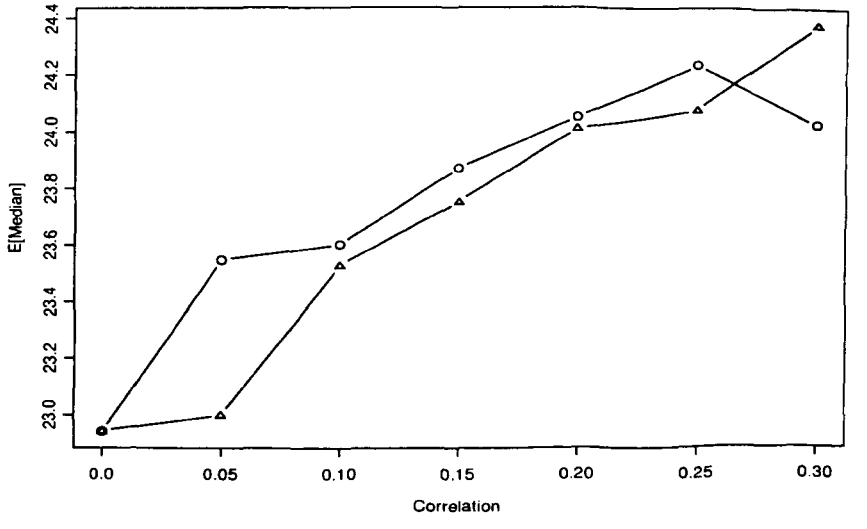

Figure 8.10: Independent estimates of the median for both types of weibull. (O) Modified Weibull; ($\triangle$) General Weibull.

Chapter 9

Conclusions

In this thesis we wanted to explore the problem of informative censoring. Knowing the problems related to this particular subject, we decided to focus on the case where the potential dependence between the failure and the censoring processes is small. We claim that almost all the cases of analysis of survival data fall into this category, in the sense that small dependencies may exist between the processes, even in the cases where we are confident that they don't. We have shown that in some cases even small dependencies of this kind can have a serious effect on the analysis.

We had to make assumptions in order to model in a reasonable way this situation. These assumptions led to models for the parametric and semi–parametric cases, where sensitivity analysis can be performed for parameters of interest. We managed to explore the relationship between the dependence parameter δ and the correlation between the two processes, while we believe that we proposed a reasonable choice for the bias function $B(t, \theta)$. The use of simulated data helped us discover firstly the validity of our model and secondly the borders where our approximation seems to collapse. An interesting part of this thesis is the analysis of the leukemia data in chapter 6, which demonstrated in a nice

way the power of having more information.

This work, of course, does not flog this subject to death. Given our way of modelling, further research on the possible choices for the bias function can be taken. In the semi-parametric case, the modified Cox's partial likelihood should be, somehow, related to the full likelihood, exactly like the partial likelihood, which will provide with an even better interpretation. In chapter 6 we discussed a model which can include both informative and non-informative censoring, provided that we have some additional information to make the distinction, although we didn't explore it to the end. This is the main area where additional research should be done, which will probably suggest that, for example, in clinical trials more information needs to be collected from each patient in order to improve our statistical analysis.

Appendix A

Expectation of the Correlation Bias–General Case

Proof: By substituting expression (5.5) in (5.4) we get

$$
\begin{aligned}
E[\hat{\theta}_\delta - \hat{\theta}_0] \;=\; & -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}} \Bigg\{ \int_0^\infty \Big[1 - H_C(c,\gamma)\Big] \frac{\int_c^\infty \frac{\partial B(u,\theta)}{\partial\theta} f_T(u,\theta)du}{S_T(c,\theta)} S_T(c,\theta) f_C(c,\gamma)dc \\
& + \int_0^\infty \Big[1 - H_C(c,\gamma)\Big] \frac{\int_c^\infty \frac{\partial f_T(u,\theta)}{\partial\theta} B(u,\theta)du}{S_T(c,\theta)} S_T(c,\theta) f_C(c,\gamma)dc \\
& + \int_0^\infty \Big[1 - H_C(c,\gamma)\Big] H_T(c,\theta) \frac{\int_c^\infty B(u,\theta) f_T(u,\theta)du}{S_T(c,\theta)} S_T(c,\theta) f_C(c,\gamma)dc \\
& - \int_0^\infty H_C(t,\gamma) \frac{\partial B(t,\theta)}{\partial\theta} f_T(t,\theta) S_C(t,\gamma)dt \Bigg\} \\
= \; & -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}} \Big\{ G_1 + G_2 + G_3 - G_4 \Big\} \qquad \text{(A.1)}
\end{aligned}
$$

where G_i is each one of the integrals above. By taking one at a time we have

$$\begin{aligned} G_1 &= \int_{c=0}^{\infty}\int_{u=c}^{\infty} \frac{\partial B(u,\theta)}{\partial\theta} f_T(u,\theta)\Big[1-H_C(c,\gamma)\Big] f_C(c,\gamma)dudc \\ &= \int_{u=0}^{\infty}\int_{c=0}^{u} \frac{\partial B(u,\theta)}{\partial\theta} f_T(u,\theta)\Big[1-H_C(c,\gamma)\Big] f_C(c,\gamma)dudc \\ &= \int_{0}^{\infty} \frac{\partial B(u,\theta)}{\partial\theta} H_C(u,\gamma)S_C(u,\gamma)f_T(u,\theta)du \\ &= G_4 \end{aligned} \tag{A.2}$$

and this is because

$$\int_0^u \Big[1-H_C(c,\gamma)\Big] f_C(c,\gamma)dc = H_C(u,\gamma)S_C(u,\gamma). \tag{A.3}$$

Now we take the second integral

$$\begin{aligned} G_2 &= \int_{c=0}^{\infty}\int_{u=c}^{\infty} \frac{\partial f_T(u,\theta)}{\partial\theta} B(u,\theta)\Big[1-H_C(c,\gamma)\Big] f_C(c,\gamma)dcdu \\ &= \int_{u=0}^{\infty}\int_{c=0}^{u} \frac{\partial f_T(u,\theta)}{\partial\theta} B(u,\theta)\Big[1-H_C(c,\gamma)\Big] f_C(c,\gamma)dcdu \\ &= \int_{u=0}^{\infty}\int_{c=0}^{u} f_T(u,\theta)\Big[1-H_T(u,\theta)\Big] B(u,\theta)\Big[1-H_C(c,\gamma)\Big] f_C(c,\gamma)dcdu \\ &= \int_{u=0}^{\infty} B(u,\theta)f_T(u,\theta)\Big[1-H_T(u,\theta)\Big] H_C(c,\gamma)S_C(c,\gamma)du \\ &= E_T\Big\{ B(t,\theta)\Big[1-H_T(t,\theta)\Big] H_C(t,\gamma)S_C(t,\gamma)\Big\} \end{aligned} \tag{A.4}$$

where under the PH assumption we have

$$\frac{\partial f_T(u,\theta)}{\partial \theta} = f_T(t,\theta)\Big[4 - H_T(t,\theta)\Big].$$

At last, if we take the third integral we have

$$\begin{aligned} G_3 &= \int_{c=0}^{\infty}\int_{u=c}^{\infty} f_T(u,\theta)B(u,\theta)H_T(c,\theta)\Big[1 - H_C(c,\gamma)\Big]f_C(c,\gamma)dcdu \\ &= \int_{u=0}^{\infty}\int_{c=0}^{u} f_T(u,\theta)B(u,\theta)H_T(c,\theta)\Big[1 - H_C(c,\gamma)\Big]f_C(c,\gamma)dcdu \\ &= \int_{u=0}^{\infty} f_T(u,\theta)B(u,\theta)\Big[\int_{c=0}^{u} H_T(c,\theta)\Big[1 - H_C(c,\gamma)\Big]f_C(c,\gamma)dc\Big]du \\ &= E_T\Big\{B(t,\theta)N(t,\theta,\gamma)\Big\} \end{aligned} \tag{A.5}$$

where

$$N(t,\theta,\gamma) = \int_{c=0}^{t} H_T(c,\theta)\Big[1 - H_C(c,\gamma)\Big]f_C(c,\gamma)dc.$$

Now, if we put (A.2),(A.4) and (A.5) into (A.1) we get

$$\begin{aligned} E[\hat{\theta}_\delta - \hat{\theta}_0] &= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}}\Bigg\{E_T\Big[B(t,\theta)\Big[1 - H_T(t,\theta)\Big]H_C(t,\gamma)S_C(t,\gamma)\Big] \\ &\qquad + E_T\Big[B(t,\theta)N(t,\theta,\gamma)\Big]\Bigg\} \\ &= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}}E_T\Big\{B(t,\theta)T(t,\theta,\gamma)\Big\} \end{aligned} \tag{A.6}$$

where

$$T(t,\theta,\gamma) = \int_0^t \Big[1 - H_C(x,\gamma)\Big]\Big[1 - H_T(t,\theta) + H_T(x,\theta)\Big] f_C(x,\gamma)dx.$$

This can be obtained by substituting (A.3) into (A.6) and combining the two integrals into one.

Appendix B

Expectation of the Correlation Bias–PH and $B(t,\theta) = 1 - H_T(t,\theta)$

Proof: If we substitute $B(t,\theta) = 1 - H_T(t,\theta)$ in (5.4) we get

$$
\begin{aligned}
E[\hat{\theta}_\delta - \hat{\theta}_0] &= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}} \Bigg\{ -\iint\limits_{C<T} \Big[1 - H_C(c,\gamma)\Big] H_T(c,\theta) f_T(t,\theta) f_C(c,\gamma) dt dc \\
&\quad + \iint\limits_{T<C} H_C(t,\gamma) H_T(t,\theta) f_T(t,\theta) f_C(c,\gamma) dt dc \Bigg\} \\
&= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}} \Bigg\{ \int\limits_{T=0}^{\infty} \int\limits_{C=T}^{\infty} H_C(t,\gamma) H_T(t,\theta) f_T(t,\theta) f_C(c,\gamma) dt dc \\
&\quad - \int\limits_{C=0}^{\infty} \int\limits_{T=C}^{\infty} \Big[1 - H_C(c,\gamma)\Big] H_T(c,\theta) f_T(t,\theta) f_C(c,\gamma) dt dc \Bigg\} \\
&= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}} \Bigg\{ \int\limits_{0}^{\infty} H_C(t,\gamma) H_T(t,\theta) f_T(t,\theta) S_C(t,\gamma) dt \\
&\quad - \int\limits_{0}^{\infty} \Big[1 - H_C(c,\gamma)\Big] H_T(c,\theta) S_T(c,\theta) f_C(c,\gamma) dc \Bigg\} \qquad \text{(B.1)}
\end{aligned}
$$

Using equation (A.3) from Appendix A in (B.1) we get

$$E[\hat{\theta}_\delta - \hat{\theta}_0] = -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^0}}\Bigg\{\int_0^\infty H_C(t,\gamma)H_T(t,\theta)f_T(t,\theta)S_C(t,\gamma)dt$$

$$-\int_{C=0}^{\infty}\Big[1-H_C(c,\gamma)\Big]f_C(c,\gamma)\int_{T=0}^{c}\Big[1-H_T(t,\theta)\Big]f_T(t,\theta)dtdc\Bigg\}$$

$$= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}}\Bigg\{\int_0^\infty H_C(t,\gamma)H_T(t,\theta)f_T(t,\theta)S_C(t,\gamma)dt$$

$$-\int_{T=0}^{\infty}\int_{C=T}^{\infty}\Big[1-H_C(c,\gamma)\Big]f_C(c,\gamma)\Big[1-H_T(t,\theta)\Big]f_T(t,\theta)dtdc\Bigg\}. \quad \text{(B.2)}$$

But,

$$\int_{C=T}^{\infty}\Big[1-H_C(c,\gamma)\Big]f_C(c,\gamma) = S_C(t,\gamma)+\int_T^\infty H_C(c,\gamma)\frac{\partial S_C(t,\gamma)}{\partial c}dc$$

$$= -H_C(t,\gamma)S_C(t,\gamma)$$

and hence, (B.2) becomes

$$
\begin{aligned}
E[\hat{\theta}_\delta - \hat{\theta}_0] &= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}} \Bigg\{ E_T\Big[H_C(t,\gamma) H_T(t,\theta) S_C(t,\gamma) \Big] \\
&\quad + E_T\Big[\Big[1 - H_T(t,\theta) \Big] H_C(t,\gamma) S_C(t,\gamma) \Big] \Bigg\} \\
&= -\frac{\delta}{\frac{\partial^2 LL(\theta,\gamma,\delta=0)}{\partial\theta^2}} E_T\Big[H_C(t,\gamma) S_C(t,\gamma) \Big]. \qquad \text{(B.3)}
\end{aligned}
$$

Appendix C

Equality of the log–likelihood functions

Proof: The cumulative hazard functions of the proxy model are

$$\Lambda_T(t) = H_T(t,\theta) + \delta \int_0^t h_T(s,\theta) \frac{\partial H_C(s,\gamma)}{\partial \gamma} \Big[\mu(s,\theta) - B(s,\theta)\Big] ds$$

and

$$\Lambda_C(c) = H_C(c,\gamma) + \delta \int_0^c h_C(s,\gamma)\mu(s,\theta) \frac{\partial \log h_C(s,\gamma)}{\partial \gamma} ds.$$

At the same time we have

$$\begin{aligned} \log \lambda_T(t) &= \log h_T(t,\theta) + \log \left[1 + \delta \frac{\partial H_C(t,\gamma)}{\partial \gamma} \Big[\mu(t,\theta) - B(t,\theta)\Big]\right] \\ &\simeq \log h_T(t,\theta) + \delta \frac{\partial H_C(t,\gamma)}{\partial \gamma} \Big[\mu(t,\theta) - B(t,\theta)\Big] \end{aligned}$$

and similarly

$$\log \lambda_C(c,\gamma) \simeq \log h_C(c,\gamma) + \delta\mu(c,\theta) \frac{\partial \log h_C(c,\gamma)}{\partial \gamma}$$

Therefore, substituting the above in the log- likelihood of (5.22) we have

$$\begin{aligned} LL_P =& \sum_{i=1}^{n} \Big[I_i \log \lambda_T(t_i) + (1 - I_i) \log \lambda_C(t_i) - \Lambda_T(t_i) - \Lambda_C(t_i) \Big] \\ =& LL(t; \theta, \gamma, \delta) + \delta \sum_{i=1}^{n} \Big[\frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \mu(t_i, \theta) - \int_0^{t_i} \mu(s, \theta) \frac{\partial h_C(s, \gamma)}{\partial \gamma} ds \\ & - \int_0^{t_i} h_T(s, \theta) \frac{\partial H_C(s, \gamma)}{\partial \gamma} \Big[\mu(s, \theta) - B(s, \theta) \Big] ds \Big] \end{aligned} \tag{C.1}$$

We need to prove that the part multiplied by δ in the above expression is zero. We will need

$$\frac{\partial \mu(u, \theta)}{\partial u} = \frac{\partial}{\partial u} \Big[\frac{\int_0^{\infty} B(s, \theta) f_T(s, \theta) ds}{S_T(u, \theta)} \Big] = h_T(u, \theta) \Big[\mu(u, \theta) - B(u, \theta) \Big]. \tag{C.2}$$

Therefore we have

$$\begin{aligned} & \sum_{i=1}^{n} \Big[\frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \mu(t_i, \theta) - \int_0^{t_i} h_T(s, \theta) \frac{\partial H_C(s, \gamma)}{\partial \gamma} \Big[\mu(s, \theta) - B(s, \theta) \Big] ds - \int_0^{t_i} \mu(s, \theta) \frac{\partial h_C(s, \gamma)}{\partial \gamma} ds \Big] \\ &= \sum_{i=1}^{n} \Big[\frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \mu(t_i, \theta) - \int_0^{t_i} \frac{\partial \mu(s, \theta)}{\partial s} \frac{\partial H_C(s, \gamma)}{\partial \gamma} - \int_0^{t_i} \mu(s, \theta) \frac{\partial h_C(s, \gamma)}{\partial \gamma} ds \Big] \\ &= \sum_{i=1}^{n} \Big[\frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \mu(t_i, \theta) - \Big[\mu(s, \theta) \frac{\partial H_C(t_i, \gamma)}{\partial \gamma} \Big]_0^{t_i} + \int_0^{t_i} \mu(s, \theta) \frac{\partial h_C(s, \gamma)}{\partial \gamma} - \int_0^{t_i} \mu(s, \theta) \frac{\partial h_C(s, \gamma)}{\partial \gamma} ds \Big] \\ &= 0 \end{aligned}$$

Appendix D

Sub–densities of both processes

In the case where we have no covariates, the joint survival function takes the form

$$\begin{aligned} S_{T,C}(x;\theta,\gamma) &= \int_x^\infty \int_x^\infty f_T(t,\theta) f_C(c,\gamma)\left[1+\delta B(t,\theta)\frac{\partial \log f_C(c,\gamma)}{\partial \gamma}\right] dtdc \\ &= S_T(x,\theta)S_C(x,\gamma)\int_x^\infty \int_x^\infty B(t,\theta)\frac{\partial \log f_C(c,\gamma)}{\partial \gamma} f_T(t,\theta) f_C(c,\gamma) dtdc \\ &= S_T(x,\theta)S_C(x,\gamma)\left[1-\delta\mu(x,\theta)\frac{\partial H_C(x,\gamma)}{\partial \gamma}\right]. \end{aligned} \tag{D.1}$$

Equivalently, the sub–density of the T–process becomes

$$\begin{aligned} f_T^\sharp(x,\theta) &= \left[-\frac{\partial}{\partial t}\left\{S_T(t,\theta)S_C(c,\gamma)\left[1-\delta\mu(t,\theta)\frac{\partial H_C(c,\gamma)}{\partial \gamma}\right]\right\}\right]_{t=c=x} \\ &= \left[f_T(t,\theta)S_C(c,\gamma)\left[1-\delta\mu(t,\theta)\frac{\partial H_C(c,\gamma)}{\partial \gamma}\right]\right. \\ &\qquad \left. +\delta S_T(t,\theta)S_C(c,\gamma)\frac{\partial \mu(t,\theta)}{\partial t}\frac{\partial H_C(c,\gamma)}{\partial \gamma}\right]_{t=c=x}, \end{aligned} \tag{D.2}$$

where

$$\begin{aligned}\frac{\partial \mu(t,\theta)}{\partial t} &= \frac{\partial}{\partial t}\left\{\frac{\int_t^\infty B(u,\theta) f_T(u,\theta)}{S_T(t,\theta)}\right\} \\ &= h_T(t,\theta)\Big[\mu(t,\theta) - B(t,\theta)\Big]. \end{aligned} \tag{D.3}$$

Therefore we have

$$f_T^\sharp(x,\theta) = f_T(x,\theta) S_C(x,\gamma)\left[1 - \delta B(x,\theta)\frac{\partial H_C(x,\gamma)}{\partial \gamma}\right]. \tag{D.4}$$

Following the same procedure for the C–process we finally get

$$f_C^\sharp(x,\gamma) = f_C(x,\gamma) S_T(x,\theta)\left[1 + \delta \mu(x,\theta)\frac{\partial \log f_C(x,\gamma)}{\partial \gamma}\right]. \tag{D.5}$$

Bibliography

[1] Collett,D.(1994).*Modelling Survival Data in Medical Research.* Chapman and Hall.

[2] Copas,J. and Eguchi,S.(2001). Local Sensitivity Appproximations for Selectivity Bias. *J.R. Statist. Soc. B* **63**,871–895.

[3] Copelan,E.A., Biggs,J.C., Thompson,J.M., Crilley,P., Szer,J., Klein,J.P., Kapoor,N.,Avalos,B.R., Cunningham,I., Atkinson,K., Downs,K., Harmon,G.S., Daly,M.B., Brodsky,I., Bulova,S.I., and Tutschka,P.J.(1991). Treatment for Acute Myelocytic Leukemia with Allogeneic Bone Marrow Transplantation Following Preparation with Bu/Cy. *Blood*, 838–843.

[4] Cox,D.R.(1959). The Analysis of Exponentially Distributed Life-Times with Two Types of Failure. *J.R. Statist. Soc. B*, **21**,411–421.

[5] Cox,D.R.(1972). Regression Models and Life-Tables(with discussion). *J.R. Statist. Soc. B*, **34**,187–220.

[6] Cox,D.R.(1975). Partial Likelihood. *Biometrika*, **62**,269–276.

[7] Cox,D.R. and Oakes,D.(1984). *Analysis of Survival Data.* Chapman and Hall.

[8] Crowder,M.(1991). On the Identifiability Crisis in Competing Risks Analysis. *Scand. J. Statist.*,**18**,223–233.

[9] Crowder,M.(1994). Identifiability Crises in Competing Risks. *Int. Statist. Rev.*,**62**, 379–391.

[10] David,H.A. and Moeschberger,M.L.(1978). *The theory of competing risks*. Griffin: High Wycombe.

[11] Farewell,V.T.(1977). A model for a binary variable with time–censored observations. *Biometrika*, **64**,43–46.

[12] Fisher,L. and Kanarek,P.(1974). Presenting Censored Survival Data When Censoring and Survival Times may not be Independent. *Reliability and Biometry: Statistical Analysis of Lifelength. SIAM:Philadelphia*, 303–326.

[13] Gail,M.(1975).A Review and Critique of some Models used in Competing Risks Analysis. *Biometrics*, **31**,209–222.

[14] Heckman,J.J. and Honore,B.E.(1989). The Identifiability of the Competing Risks Model. textslBiometrika,**76**,2,325-330.

[15] Heitjan,D.F. and Rubin,D.B.(1991).Ignorability and Coarse Data. *The Annals of Statistics*,**19**,2244–2253.

[16] Hoover,D.R. and Guess,F.M.(1990).Response linked censoring: Modelling and Estimation. *Biometrika*,**77**,893–896.

[17] Kalbfleisch,J.D. and Prentice,R.L.(1972). Contribution to discussion of paper by D.R.Cox. *J.R.Statist.Soc.B*,**34**,215–216.

[18] Kalbfleisch,J.D. and Prentice,R.L.(1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*,**60**,267–278.

[19] Kalbfleisch,J.D. and Prentice,R.L.(1980). *The Statistical Analysis of Failure Time Data*. New York:Wiley.

[20] Kaplan,E.L. and Meier,P.(1958). Nonparametric stimation from Incomplete Observations. *Journal of American Statistical Association*,**53**,457–481.

[21] Klein,J.P. and Moeschberger,M.L.(1988). Bounds on Net Survival Probabilities for Dependent Competing Risks. *Biometrics*,**44**,529–538.

[22] Klein,J.P. and Moeschberger,M.L.(1997). *Survival Analysis*. Springer.

[23] Krall,J.M. and Uthoff,V.A. and Harley, J.B.(1972). A Step–Up Procedure for Selecting Variables Associated with Survival. *Biometrics* **31**,49–57.

[24] Lagakos,S.W.(1979). General Right Censoring and Its Impact on the Analysis of Survival Data. *Biometrics*,**35**,139–156.

[25] Le,C.T.(1997). *Applied Survival Analysis*. New York:Wiley.

[26] Link,W.A.(1989). A Model for Informative Censoring. *Journal of the American Statistical Association*, **84**,749–752.

[27] Maller,R and Zhou,X.(1996).*Survival Analysis with Long Term Survivors.* John Wiley and Sons.

[28] Moeschberger,M.L. and David,H.A.(1971). Life Tests Under Competing Causes of Failure and the Theory of Competing Risks. *Biometrics*,**27**,909–933.

[29] Moeschberger,M.L. and Klein,J.P.(1984). Consequences of Departing From Independence in Exponential Series System. *Technometrics*,**26**,277–284.

[30] Moeschberger,M.L. and Klein,J.P.(1995). Statistical Methods for dependent Competing Risks. *Lifetime Data Analysis*,**1**,195–204.

[31] Peterson,A.V.(1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to Competing Risks. *Proc. Nat. Acad. Sci. USA*,**73**,11–13.

[32] Prentice,R.L., Kalbfleisch,J.D., Peterson,A.V., Flournoy,N., Farewell,V.T. and Breslow,N.E.(1976). The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics*,**34**,541–554.

[33] Slud,E.V. and Rubinstein,L.V.(1983). Dependent Competing Risks and Summary Survival Curves. *Biometrika*,**70**,643–649.

[34] Tsiatis,A.(1975). A Nonidentifiability Aspect of the Problem of Competing Risks. *Proc.Nat.Acad.Sci.USA*,**72**,20–22.

[35] Vaupel,J.W., Manton,K.G. and Stallard,E.(1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*,**16**,439–454.

[36] Williams,J.S. and Lagakos,S.W.(1977). Models for censored survival analysis: Constant–sum and variable–sum models. *Biometrika*,**64**,215–224.