

51024

0-315-03128-X



National Library of Canada

Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR LUC JANICOT

TITLE OF THESIS/TITRE DE LA THÈSE LARGE SAMPLE PROPERTIES OF THE COX
TECHNIQUE IN SURVIVAL TIME ANALYSIS

UNIVERSITY/UNIVERSITÉ SIMON FRASER UNIVERSITY

DEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE M. Sc.

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ 1981

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE DR. R. LOCKHART

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE December 17th 1980 SIGNED/SIGNÉ _____

PERMANENT ADDRESS/RÉSIDENCE FIXÉ



National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents, qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

LARGE SAMPLE PROPERTIES OF THE COX
TECHNIQUE IN SURVIVAL TIME ANALYSIS

by

Luc Janicot

Ingenieur de L'ecole Centrale

Des Arts et Manufactures

(Ecole Centrale Paris), Paris, France, 1979

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Mathematics



Luc Janicot 1980

SIMON FRASER UNIVERSITY

December 1980

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.

APPROVAL

Name: Luc Janicot
Degree: Master of Science
Title of Thesis: Large sample properties of the Cox
technique in survival time analysis.

Examining Committee:

Chairperson: G.A.C. Graham

R. Lockhart
Senior Supervisor

K.L. Weldon

B.S. Thomson

R. Routledge
External Examiner
Department of Mathematics
Simon Fraser University

Date Approved: December 16, 1980

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis or dissertation (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this thesis for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis/Dissertation:

LARGE SAMPLE PROPERTIES OF THE COX
TECHNIQUE IN SURVIVAL TIME ANALYSIS

Author:

(signature)

LUC JANICOT

(name)

December 17th 1980

(date)

ABSTRACT

The Cox partial likelihood technique is explored in a large sample setting. For a variety of parametrizations of the proportional hazard rate model, consistency and asymptotic normality of the Maximum Partial Likelihood estimators are proved. The covariates are taken to be random and observations are censored on the right. Previous assumptions on the distribution of the covariates are relaxed. An extension to the case of non-random covariates is considered.

RESUME

Les propriétés asymptotiques de l'estimateur obtenu en maximisant la fonction de vraisemblance partielle de Cox sont étudiées. Sa convergence et sa normalité asymptotique sont prouvées pour diverses paramétrisations du modèle. Les covariates sont assumées aléatoires. Les hypothèses sur leur distribution faites dans les articles parus à ce jour sont supprimées. Enfin le cas où les covariates sont des constantes non-aléatoires est exploré.

ACKNOWLEDGEMENTS

I would like to thank Dr. R. Lockhart for his assistance during this thesis and more generally during my graduate work.

I would also like to thank the staff of the mathematics department especially Ms. Sylvia Holmes and Ms. Betty Dwyer and my fellow graduate students, especially John Spinelli, Chris Mah, and Wenni Kuo, for their kindness and help during my academic year at Simon Fraser University.

Last but not least I would like to thank my friend Catherine Baillie for her help during the stressful typing stage of this thesis.

TABLE OF CONTENTS

Approval	(ii)
Abstract	(iii)
Acknowledgements	(iv)
Introduction	1
Literature Survey on Hazard Rate Models with Covariates	5
Chapter 1: Model, Assumptions, Notation	12
Chapter 2: Consistency	16
Chapter 3: Asymptotic Normality	58
Chapter 4: Extension and Problems	78
Part A) Consistency of Cox's estimate when the covariates are given constants.	78
Part B) Infinite Cox's estimate	81
Appendix	83
Bibliography	93

Introduction

In a carcinogenic experiment, the cancer-inducing properties of a substance (for example a new food additive) have to be evaluated. This substance is injected into mice in different doses. The times of appearance of a tumor are recorded. Some mice might die before developing cancer. In cancer research a new drug is found, but its effectiveness seems to depend on several covariates: for example the age of the patient, its sex, its blood count at the beginning of the experiment, etc... Then an experiment is designed and the times of death, for example, are recorded. Some patients might decide to leave the experiment before the end. Some might die of causes other than cancer. These two examples are censored survival data: some subjects are withdrawn from the experiment for causes other than those of interest.

A useful tool to analyse survival data is the hazard rate. It gives the "intensity of risk" of failure at any time t , given that the individual has not failed prior to t . More formally the hazard rate at time t among subjects with covariate Z is defined as:

$$\lambda(t, z) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T \leq t + \Delta t | T \geq t, z]}{\Delta t}$$

where T is the time of failure of the subject (time at which the mouse develops a tumor, time at which the individual dies), Z in our examples would be the dose of substance, the age of the patient, etc... A model commonly accepted is the

proportional hazard rate model in which the hazard rate can be factored into one term depending on t and one term depending on the covariate.

Cox (1972) assumes a partly parametrized proportional hazard rate model according to:

$$\lambda(t, z) = \lambda_0(t) h(\beta, z)$$

where $\lambda_0(t)$ is an unspecified function of t and where

$h(\beta, z)$ is usually $e^{\beta'z}$. However in carcinogenic experiments, where only small doses are considered, $h(\beta, z)$ is commonly a polynomial in the dose Z . Then Cox introduces the partial likelihood function. At a time t , the risk set $R(t)$ at that time is the set of all subjects still in the experiment.

Let us consider a failure time t_i . Given the fact that one subject failed at that time and given $R(t_i)$, the conditional probability that subject j failed is

$$\frac{h(\beta, z_j)}{\sum_{k \in R(t_i)} h(\beta, z_k)}$$

Simply multiplying these terms gives Cox's partial likelihood. This function is not really a likelihood function since the factor related to the intervals between two failure times is ignored. Nor is it a conditional or marginal likelihood except in special cases.

Therefore the classical results on the asymptotics of maximum likelihood estimates cannot be used. Cox (1972) gave a rather informal justification of the consistency and asymptotic normality. Since then no fully satisfying

paper has been published. Tsiatis (1978) proves consistency and asymptotic normality for bounded real-valued covariates drawn from a continuous population. In his argument, the experiment is stopped at some prespecified time T_f . Liu and Crowley (to appear) assume their covariates to be drawn from a finite set of possible values. Their argument is conditional on the observed values of the covariates.

Our first chapter will be a literature survey on the hazard rate models with covariates. It will give us an idea on how to analyze our two examples of survival data. The following chapter will give some more precise formulation of the model, assumptions and notations used thereafter. We shall then prove consistency and asymptotic normality for vector-valued covariates drawn from a continuous population. The proof of the consistency in the exponential case $h(\beta, z) = e^{\beta z}$ will need no assumption on the boundedness of the covariates, only on their moments. Furthermore we will not assume that the experiment is stopped after the prespecified time.

In the case of other parametrizations of $h(\beta, z)$ it will be proved that the maximum partial likelihood equation has a consistent root when the experiment is stopped at some prespecified time.

The asymptotic normality will be proved for vector-valued covariates not bounded but with some further assumptions on the function h and its first, second and third

order derivatives.

Then an extension of our arguments to the case where the covariates are give constants is considered.

Literature survey on hazard rate models with covariates

The analysis of time to occurrence or "failure data" is of considerable interest for medical experiments. The growing importance of medical research has triggered numerous papers on specialized methods to analyze this kind of data. The classical notion of hazard function is a tool commonly used. The hazard function $\lambda(t, z)$ gives the "risk" of failure at any time of a subject having covariates Z , given that the subject has not failed prior to t . In his pioneering paper in 1972, Cox assumed (as we do in the following chapters) that the hazard function could be factorized in one term depending on t and one term depending on Z and a parameter β according to:

$$\lambda(t, z) = \lambda_0(t) h(\beta, z) \quad (1)$$

More precisely, Cox specified $h(\beta, z) = e^{\beta'z}$ as do the majority of papers. Another fruitful partially parametric model was introduced by Prentice and Kalbfleisch (1979):

$$\lambda(t, z) = \lambda_0(t e^{\beta'z}) e^{\beta'z} \quad (2)$$

The covariates then act multiplicatively on the time of failure itself. A prospective type of sampling is usually considered. A population of subjects is given at some specified origin of time measurement and followed forward to observe their respective times of failure. Kalbfleisch and Prentice (1979) investigated the case of retrospective studies in which subjects are selected on the basis of

their time of failure after which one "looks back" to ascertain the corresponding covariates values or covariates functions.

I) Analysis of the models:

There are several approaches to the analysis of the above models. The most natural one is to parametrize them completely, by using for example a two-parameter Weibull regression model for model (1) with $\lambda_0(t) = \lambda_0 (t)^{q-1}$. Other parametric special cases of (2) include the log-normal, the log-logistic and the generalized gamma regression models (e.g. Farewell and Prentice 1977). Then standard methods such as maximum likelihood can be used.

The other approach to deal with model (1) is the technique considered in the following chapters: 'Cox's partial likelihood method.

II) Cox's partial likelihood technique:

Cox's technique is a partial parametrization of model (1) where $\lambda_0(t)$ is allowed to be arbitrary. The main interest is in the regression parameters: for example we want to know if the substance studied induces tumors. Cox's likelihood function is then introduced and the vector of estimates of the regression parameters is the maximum of this likelihood function. Cox (1972) gave a rather informal justification to this likelihood function. In 1975 he considered it in more detail under the term partial likelihood. Kalbfleisch and Prentice (1973) showed that in the absence

of censoring, Cox's likelihood is precisely the likelihood based on the marginal distribution of the ranks of the failure times. A group invariance argument showed the rank vector to be "partially" sufficient, in the sense of Barnard (1963), for β in the "absence of knowledge" of λ_0 . Their group invariance argument breaks down with censored data.

In this case they showed only that it is the likelihood corresponding to the marginal probability of the set of possible underlying rank vectors consistent with the observed data. There are no classical results on this "likelihood".

No tied failure times are assumed in the above models. Unfortunately the data will often be recorded with ties. If the number of ties is small, an ad hoc modification of the above procedures is satisfactory (Cox 1972, Kalbfleisch and Prentice 1979). Otherwise it is preferable to use a discrete failure time model. Kalbfleisch and Prentice (1973) showed that grouping the continuous model (1) gives a discrete model with simple properties. Cox (1972) introduced a generalized partial likelihood function.

III) Asymptotics:

Two main papers, still unpublished, deal with consistency and asymptotic normality of the estimator obtained by using Cox's technique. These are Tsiatis's (1978) and Liu-Crowley's (to appear) but neither of them can be considered as fully satisfactory. Liu and Crowley consider a discrete approach to the problem: their covariates take values in a finite set of possible values.

Their argument is conditional on the observed values of the covariates. This is an interesting step towards proving consistency and asymptotic normality when the covariates are not assumed to be drawn from a population but are given constants. The main drawback is that they consider a finite set of possible values and their argument seems to be difficult to extend to an infinite set of possible values. Tsiatis considers as we do a continuous approach. He assumes that the real-valued covariates are bounded and that the experiment is stopped at a prespecified time. We remove both these assumptions and consider vector-valued covariates.

Cox's technique is asymptotically fully efficient only in special circumstances. However, the amount of information lost in any specific data, with the λ_0 function restricted, is usually small. Many have tried to give some more formal basis to that argument. Kalbfleisch (1974) assumes that there is no censoring, that the covariate vector has dimension 1 and does not depend on time. He then shows that the Cox likelihood estimator has full efficiency with respect to the M.L.E. relative to any parametric model in (1) of the form $\lambda_0(t) = \lambda h_0(t)$ with $h_0(t)$ known and λ scale parameter to be estimated if the true value β is equal to 0. He then derives the approximate expression $\exp\{-\beta^2 \text{Var}(z)\}$ for the asymptotic relative efficiency valid near $\beta=0$. The dependence of

efficiency on the regression parameter is a situation unfamiliar to ordinary linear regression. In an important paper, Efron (1977) attacks the problem from an interesting viewpoint. Taking the covariate functions and censoring times to be fixed, he derives expressions for the finite sample information matrices and gives conditions for Cox's method to be asymptotically fully efficient. In his parametrization he introduces a notional "average hazard if all items are on test". In his formulation, the underlying hazard function may depend on the regression coefficients as well as the nuisance parameters. He then discusses the relative efficiency in the two-sample problem and gives some simulation results in this case.

This thesis provides a rigorous foundation for some of the efficiency calculations.

An attractive feature of inference based on Cox's likelihood is the robustness implied by the arbitrariness in the λ_0 function. Steve Samuels, in a 1977 unpublished doctoral dissertation at the University of Washington has examined the robustness of $\hat{\beta}$ more formally.

IV) Other parametrizations of the hazard rate—Parametrizations of $h(\beta, z)$

In the preceding sections we factorized the hazard rate $\lambda(t, z)$ according to $\lambda(t, z) = \lambda_0(t) h(\beta, z)$. The most common parametrization for $h(\beta, z)$ is the exponential

model. Feigl and Zelen (1965) treated the case $\lambda_0(t) = \lambda$ constant, and proposed several forms for h , notably the exponential model and $h(\beta, z) = \frac{1}{1 + \beta z}$. In carcinogenic testing, scientists are interested in the relationship between the dose of a potential cancer inducing substance injected into an animal and the time at which that animal develops cancer. In this field, many workers stipulate the Weibull form $\lambda_0(t) = k(t-w)^{k-1}$ in which w is usually taken as known and hence without loss of generality taken to be 0. Hartley and Sielken (1977) introduced a polynomial form for λ_0 according to $\lambda_0(t) = \sum \beta_r t^r$ with $\beta_r \geq 0$ and some standardization rules. The model form of the dose dependent $h(\beta, z)$ has been of particular concern, as the interest is usually in small doses (see Mantel and Bryan (1961)) and it is in this dose range that monitoring data is difficult, particularly if the spontaneous rate is zero. The multiphase models of Armitage and Doll (1961) and Armitage (1974) specify $h(\beta, z)$ as a polynomial of the dose level z of the form $h(\beta, z) = c \prod_{s=1}^q (1 + \gamma_s z)$ in which all γ_s are strictly positive. Peto (1974) adopts a more general model of the form $h(\beta, z) = \sum_{s=0}^q \alpha_s z^s$, $\alpha_s > 0$.

In this thesis Cox's technique is studied with more general parametrization of h such as those mentioned here.

Hartley and Sielken (1975b) dropped the polynomial form and stipulated only that $h(\beta, z)$ is a smooth positive

convex function (unpublished paper). As in this case

$\lambda_0(t)$ is parametrized, their technique is a partly parametric one similar to Cox's.

V) Survivor function estimation:

Surprisingly little attention has been given to the estimation of the survival function and its standard error. The underlying survivor function $F(t, z) = P[\tau \geq t, z]$

can be written $F(t, z) = F_0(t)^{h(\beta, z)}$ where

$$F_0(t) = \exp\left(-\int_0^t \lambda_0(u) du\right) = \exp(-\Lambda_0(t))$$

At any specified β a non-parametric maximum likelihood estimation can be carried out (Kalbfleisch and Prentice 1973) to give an estimator of $F_0(t)$. Breslow and Crowley (1974) consider a continuous case, where there are no ties among observations. Using likelihood arguments, they introduce an empirical integrated hazard function according to

$$\hat{\Lambda}_0(t, \beta) = \sum_{i \in D(t)} \frac{1}{\left[\sum_{j \in R(t_i)} \exp \beta z_j \right]}$$

They then establish that if the true value of β is 0 the random function $\sqrt{n} [\hat{\Lambda}_0(t) - \Lambda_0(t)]$ converges to a mean zero Gaussian process, $\Lambda_0(t)$ being the true

integrated hazard function. Tsiatis has extended that result to $\hat{\Lambda}_0(t, \hat{\beta})$ where $\hat{\beta}$ is Cox's partial likelihood estimate again under rather severe restrictions such as bounded covariates.

We have not studied the problem of extending this result to the more general setting of this thesis.

Chapter 1: Model, Assumptions, Notation

In this chapter we shall give a precise mathematical formulation of the model and introduce the notation and assumptions needed later on. In the examples given in the introduction, the available data were the covariates for each subject, the time at which each subject left the experiment, the reason why it left and the size of the population. Let us introduce the following notation:

- let Z_i be the vector of observed covariates for subject i .
- let Y_i be the time of disappearance of subject i from the experiment.
- we define an indicator variable δ_i to indicate for which reason the subject i left the experiment according to $\delta_i = 0$ if subject i was censored, $\delta_i = 1$ otherwise.
- let N be the size of the population at the start of the experiment.
- let β_0 be the true value of the regression parameter.

Distribution of the random variables:

Let us assume that for each subject i we have two underlying random variables: T_i , the true underlying survival time, and S_i time of censoring. The observed variable Y_i is just the minimum of T_i and S_i . T and S will

be assumed independent given the vector of covariate Z . This is not a mild assumption. For example, in a medical experiment, a person cured (very late time of death) may leave the experiment early. It seems very hard to remove this assumption as a censoring dependent on the time of death could grossly mislead the statistician by concealing all the information relative to some values of the covariates.

To solve our problem, we shall take a continuous approach. The covariate vectors will be assumed to be of dimension p and to be drawn from a continuous population. It is our conviction that the method we use in the following chapter can be extended to the case where the covariates are given constants drawn from an infinite population of possible values.

Let us introduce the notation: $G(t|z) = P[S \geq t | z]$
 $G(t|z)$ is an unknown function of t . S does not necessarily have a conditional hazard function. We have the freedom to allow censoring in groups, even to assume that, at a certain time T_f , everyone still in the experiment is censored. For technical reasons, we shall assume that there is only a finite number of mass-points in the density of S given Z .

The model for the hazard rate:

We shall assume that the random variable T has a conditional hazard rate $\lambda(t, z)$. We shall assume a commonly accepted form of the hazard rate $\lambda(t, z) = \lambda_0(t) h(\beta, z)$

where $\lambda_0(t)$ depends only on t and is otherwise unspecified, and $h(\beta, z)$ depends on a parameter β and the covariate vector z . β is the parameter of interest and has dimension p . This model is called the proportional hazard rate model since the ratio of the hazard rates of two subjects depends on the covariates and not on time. Following Cox (1972) we shall use a partial parametrization of $\lambda(t, z)$.

$\lambda_0(t)$ is assumed unknown and without any constraint.

On the other hand $h(\beta, z)$ is assumed to be a known function of β and z . Usually $h(\beta, z)$ is assumed to be $e^{\beta'z}$.

In the following we shall assume only that $h(\beta, z)$ is such that $h(\beta, z), \frac{\partial h}{\partial \beta_k}(\beta, z) (k=1, \dots, p)$ are monotone functions of β_k for each z and each k . This assumption is satisfied for all the parametrizations mentioned in the literature survey.

Assumptions on the moments of $h(\beta, z)$:

Tsiatis (1978) assumed that the covariates are bounded. Instead we shall make assumptions only on the moments of $h(\beta, z)$ and

$E[h(\beta, z)^r]$ exists for $r = -3, -2, -1, 1, 2, 3$

and for every β . In what follows there is actually some freedom of choice in which moments of h are assumed to

exist. For instance we could assume $E[h(\beta, z)^r] < \infty$

for each β and $-1 \leq r \leq +5$. (See page 38.)

$E\left[\left\|\frac{\partial h}{\partial \beta}\right\|\right]$ and $E\left[\left\|\frac{\partial h}{\partial \beta}\right\|^2\right]$ exist for

every β .

Although it would be desirable to do so it seems difficult to eliminate the assumptions about inverse moments of h .

Obviously in the exponential case it is sufficient that $E[e^{\beta z}]$ exists for every β .

Classical results:

Let $\varphi_z(t)$ be the density of T given Z . We know that:

$$\lambda(t, z) = \frac{\varphi_z(t)}{1 - P[T \leq t | z]}$$

Straightforward computations yield that:

$$\varphi_z(t) = h(\beta_0, z) \lambda_0(t) \exp\left(-\int_0^t \lambda_0(u) h(\beta_0, z) du\right)$$

The conditional probability of surviving until time t without being censored given that $Z=z$ is:

$$H(t|z) = G(t|z) \exp\left(-\int_0^t \lambda_0(x) h(\beta_0, z) dx\right)$$

Chapter 2: Consistency

In this chapter we shall prove the consistency of the estimate obtained by maximizing Cox's partial likelihood function. More precisely we shall prove the following theorem:

- In the exponential case, Cox's partial likelihood has only one local maximum and this maximum converges in probability to the true value of the parameter.
- In the general case where $h(\beta, Z)$ is such that $h(\beta, Z)$ and $\frac{\partial h(\beta, Z)}{\partial \beta_k}$ ($k = 1, \dots, p$) are monotone functions of β_l for each Z and each l , if the time of censoring is bounded by a finite time T_f , and if this time T_f is such that $P[Y \geq T_f] > 0$, then there is a root of the maximum Cox's partial likelihood equation (a local maximum of Cox's likelihood function) that is consistent in probability.

The log of Cox's partial likelihood is:

$$\text{Log } L_c = \sum_{i=1}^N \delta_i \text{Log } h(\beta, z_i) - \sum_{i=1}^N \delta_i \text{Log} \left[\sum_{j=1}^N 1(y_j \geq T_i) h(\beta, z_j) \right]$$

Let us consider

$$\Phi_N(\beta) = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } h(\beta, z_i) - \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} \left[\frac{1}{N} \sum_{j=1}^N 1(y_j \geq T_i) h(\beta, z_j) \right]$$

$\Phi_N(\beta)$ and $\text{log } L_c$ have the same shape. Studying the maxima of $\Phi_N(\beta)$ is equivalent to studying the maxima of $\text{log } L_c$. From now on we will study $\Phi_N(\beta)$ instead of $\text{log } L_c$.

$\Phi_N(\beta)$ is a sum of dependant random variables.

$$\text{Let } f_N(t) = \frac{1}{N} \sum_{j=1}^N 1(y_j \geq t) h(\beta, z_j)$$

By the strong law of large numbers, for fixed t , $f_N(t)$ converges to $f(t) = E[h(\beta, z) 1(y \geq t)]$. We will prove that

we can replace $f_N(T_i)$ by $f(T_i)$ in $\Phi_N(\beta)$. $\Phi_N(\beta)$ is then a sum of i.i.d random variables. We then shall apply the strong law of large numbers to prove that $\Phi_N(\beta)$ converges point-wise to a function $\Phi(\beta)$.

We shall then study the function $\Phi(\beta)$. We shall prove two important results:

- 1) β_0 is an extremum of $\Phi(\beta)$
- 2) $\Phi(\beta)$ is concave on a small neighbourhood of the true value β_0 .

The exponential case and the general case will then be distinguished.

In the exponential case, we shall prove ϕ is concave on the whole domain of β . Then $\phi_N(\beta)$ has only one local maximum $\hat{\beta}_N$. In a neighbourhood V of the true value β_0 , $\phi_N(\beta)$ and $\phi(\beta)$ are two continuous concave functions. We shall prove (lemma 7) this implies that the location of the maximum of ϕ_N in V converges to the location of the maximum of ϕ , that is β_0 . $\phi_N(\beta)$ has only one local maximum. This implies the consistency of $\hat{\beta}_N$.

In the general case, we will have to reduce our study to the neighbourhood V of β_0 . As we don't know anything about the shape of ϕ_N in V we shall need to prove the uniform convergence of ϕ_N to ϕ in V . We will then prove that ϕ_N has a local maximum in V and that this local maximum converges to β_0 . To prove the uniform convergence, we shall prove that the sequence is equicontinuous. In order to do so, we shall have to assume that the censoring time, S , for an individual is bounded by a prespecified time T_f . This assumption is needed more for mathematical reasons than real statistical ones. Anyway it is a mild one, since most survival studies are ended after some prespecified time T_f .

A) Lemmas:

The four following lemmas are mainly computations needed later on. First let us introduce the following notations: Let $k(z, t)$ be a function of z and t :

$E_{z, t}(k(z, t))$ means expectation of $k(z, t)$ where the summation is done over z and t .

$E_t(k(z, t))$ means conditional expectation of $k(z, t)$ given z . The proofs of these lemmas are given in the appendix.

Lemma 1:
$$E[S|z] = \int_{t=0}^{t=\infty} \lambda_0(t) h(\beta_0, z) \exp\left\{-\int_0^t \lambda_0(u) h(\beta_0, z) du\right\} G(t|z) dt$$

$$E[S|z] = E_t[G(t|z)]$$

Lemma 2: Let $f(w) = E[h(\beta, z) | (Y \geq w)]$

$$f(w) = E_z \left[h(\beta, z) \left(\exp - \int_0^w \lambda_0(u) h(\beta_0, z) du \right) G(w|z) \right]$$

Lemma 3:

$$E[S \text{ Log } f(t)] = E_{z, t} [\text{Log } f(t) G(t|z)]$$

Lemma 4:

$$E[S | (Y \geq w) | z] = \int_{t=w}^{t=\infty} \lambda_0(t) h(\beta_0, z) \left(\exp - \int_0^t \lambda_0(u) h(\beta_0, z) du \right) G(t|z) dt$$

Lemma 5: Let $\{a_n\}$ a sequence of real numbers converging to 0, let $h(\beta, z)$ be a known function of β and z .

Then

$$\text{Sup}_t \left\{ a_n \sqrt{n} \left| \frac{1}{n} \sum_{j=1}^n 1(y_j \geq t) h(\beta, z_j) - E[1(y \geq t) h(\beta, z)] \right| \right\}$$

converges to 0 in probability, for β fixed.

Proof: First let us give some definitions. Let D be the space of functions on $[0, +\infty($ that are right continuous and have left-hand limits. Let us define the Skorokhod topology on D . Let \mathcal{D} be the σ -field generated by the open sets of D . Let x and y be 2 functions of D . We want to define a distance $d(x, y)$. The idea is that we cannot measure time with perfect accuracy anymore than we can position.

Definition: Let Λ denote the class of strictly increasing, continuous mappings of $[0, +\infty($ onto itself. If $\lambda \in \Lambda$ then $\lambda(0) = 0$ and $\lambda(+\infty) = +\infty$. Then set $d(x, y) = \inf_{\lambda \in \Lambda} \varepsilon$

where

$$S = \left\{ \varepsilon; (\exists \lambda \in \Lambda) \left\{ \begin{array}{l} \sup_t |\lambda(t) - t| \leq \varepsilon \\ \sup_t |x(t) - y(\lambda(t))| \leq \varepsilon \end{array} \right. \right\}$$

Let us consider the random process $Z_n(t) =$

$$\sqrt{n} \left[\frac{1}{n} \sum_{j=1}^n 1(y_j \geq t) h(\beta, z_j) - E[1(y \geq t) h(\beta, z)] \right]$$

The random process Z_n is an application from (Ω, \mathcal{B}, P)

where Ω is the sample space, to D . Then for ω belonging to Ω , $Z_n(\omega)$ is an element of D . The random process

Z_n induces a probability measure P_n on (D, \mathcal{D}) according

to: $P_n(A) = P_n(Z_n(\omega) \in A)$ where A is a

measurable subset of D .

We shall prove that the sequence $\{Z_n\}$ of random processes converges in distribution to a Gaussian process Z which has mean 0. By definition this means that the distribution P_n of the Z_n converge weakly to the distribution of P of Z : $P_n \implies P$. From Billingsley (1968), we have to prove that the finite dimensional distributions of $Z_n(t)$ are asymptotically multivariate normal with mean zero and that the sequence of distributions of $Z_n(t)$ is tight

Step 1: A simple application of the multivariate central theorem yields that the finite dimensional distributions of $Z_n(t)$ are asymptotically distributed as a multivariate normal with mean zero.

Step 2: The sequence $\{P_n\}$ is tight. From Billingsley (1968) p. 128, it suffices to show that:

$$E \left[(Z_n(t) - Z_n(t_1))^2 (Z_n(t_2) - Z_n(t))^2 \right] \leq C_4 [F(t_2) - F(t_1)]^{4/3}$$

where $F(t) = P[Y < t]$, $0 \leq t_1 \leq t \leq t_2$

The proof is given in the appendix. Therefore we have:

$$P_n \implies P$$

Now let us prove it implies that

$$\sup_t \left\{ a_n \sqrt{n} \left[\frac{1}{n} \sum_{j=1}^n 1(Y_j \geq t) h(\beta, z_j) - E[1(Y \geq t) h(\beta, z)] \right] \right\}$$

converges to 0 in probability.

Let us give first some further results and definitions:

- let us define the supremum norm according to, for x belonging to D :

$$\rho_T(x) = \sup_t |x(t)|$$

- Result: (Billingsley) convergence to a continuous limit (such as $Z(w)$)

in ρ_T is equivalent to convergence in the Skorokhod metric. Therefore, if A_ϵ is the subset of D such that

$$A_\epsilon = \{x \in D; \rho_T(x) > \epsilon\},$$

$$P_n(Z_n(\omega) \in A_\epsilon) \longrightarrow P(Z(\omega) \in A_\epsilon)$$

Similarly

$$P_n(a_n Z_n(\omega) \in A_\epsilon) \longrightarrow P(a_n Z(\omega) \in A_\epsilon)$$

The sequence of real numbers

$$y_n = P(a_n Z(\omega) \in A_\epsilon)$$

converges to 0. We then have proved the result:

$$\sup_t a_n \sqrt{n} \left[\frac{1}{n} \sum_{j=1}^n 1(y_j \geq t) h(p, z_j) - E[1(y \geq t) h(p, z)] \right]$$

converges to 0 in probability.

Lemma 6: Let us consider the i.i.d random variables

$$|\log h(\beta, z_j)| \quad . \quad \text{We have: } \frac{\max_{j=1, \dots, N} |\log h(\beta, z_j)|}{n^{1/3 - 1/p}} \text{ converges}$$

almost surely to 0 when $N \rightarrow +\infty$ with p large. Later on we shall need to know "how fast" the maximum over the whole population of $|\log h(\beta, z_j)|$ goes to $+\infty$. This lemma gives us an upper bound.

Proof:

Note: Borel-Cantelli lemma: given events A_n ; $n=1, 2, \dots$, if $\sum P A_n < \infty$ then $P(A_n \text{ i.o.}) = 0$ (to be read "A_n's occur infinitely often").

First by using the Borel-Cantelli lemma we shall prove that for any ε

$$P \left[\frac{|\log h(\beta, z_n)|}{n^{1/3 - 1/p}} > \varepsilon \text{ i.o.} \right] = 0,$$

$|\log h(\beta, z_n)|$ being the value of $\log h(\beta, z)$ for the N th subject. We shall then prove it yields that

$$P \left[\frac{\max_{j=1, \dots, N} |\log h(\beta, z_j)|}{N^{1/3 - 1/p}} > \varepsilon \text{ i.o.} \right] = 0$$

Let us introduce the following notation: $x_j = |\log h(\beta, z_j)|$ and $M_n = \max_{j=1, \dots, N} X_j$ Let A_n be the event $\frac{X_n}{n^{1/3 - 1/p}} > \varepsilon$

Let us prove $\sum P A_n < \infty$

The Tchebitcheff inequality yields

$$P[Y > \varepsilon] \leq \frac{E[Y^k]}{\varepsilon^k}$$

Then $P \left[\frac{X_n}{n^{1/3 - 1/p}} > \varepsilon \right] \leq \frac{E[X_n^k]}{\varepsilon^k n^{k/3 - k/p}}$ provided $E(X^k)$ exists.

Then $\sum P A_n \leq \frac{E[X^k]}{\varepsilon^k} \sum_{n=1}^{+\infty} \frac{1}{n^{k/3 - k/p}}$

To have convergence of $\sum P_n$ it is sufficient that

$$\frac{K}{3} - \frac{K}{p} > 1, \text{ for example } K = 3 + \delta, \text{ where } \delta \text{ is}$$

an arbitrary small number. As

$$|\log h(\beta, z)| \leq h(\beta, z) + \frac{1}{h(\beta, z)}$$

the existence of $E[|\log h(\beta, z)|^{3+\epsilon}]$ is implied by

$$\text{the existence of } E[h(\beta, z)^{3+\epsilon}] \text{ and } E[h(\beta, z)^{-(3+\epsilon)}]$$

Therefore there is convergence of $\sum P_n$. (Recall our

assumption in chapter 1) The Borel-Cantelli lemma

implies then that

$$P\left[\frac{X_n}{n^{1/3-1/p}} > \epsilon \text{ i.o.}\right] = 0$$

Let us prove by contradiction this implies:

$$P\left[\frac{M_n}{n^{1/3-1/p}} > \epsilon \text{ i.o.}\right] = 0$$

Let us assume there exists a sequence $\{M_{n_i}\}$ such

$$\text{that } \frac{M_{n_i}}{n_i^{1/3-1/p}} > \epsilon$$

Consider first $n_{(1)} \in \mathbb{N}$ such that

$$\frac{M_{n_{(1)}}}{n_{(1)}^{1/3-1/p}} > \epsilon$$

$$\text{If } X_{n_{(1)}} \leq \epsilon n_{(1)}^{1/3-1/p} \text{ then } M_{n_{(1)}} \geq \epsilon n_{(1)}^{1/3-1/p} > \epsilon (n_{(1)}-1)^{1/3-1/p}$$

n is not the first n such that $M_n \geq \epsilon n^{1/3-1/p}$.

it is a contradiction. Then

$$X_{n_{(1)}} > \epsilon n_{(1)}^{1/3-1/p}$$

$$\exists n^* \text{ such that } M_{n_{(1)}} \leq \epsilon (n^*)^{1/3-1/p}$$

$$(\exists n' \in \mathbb{N}) \begin{cases} n' > n^* \\ M_{n'} > \epsilon (n')^{1/3-1/p} \end{cases}$$

As $M_{n_{(1)}} \leq \epsilon (n^*)^{1/3-1/p}$ it implies $\{X_i\}_{i=1, \dots, n_{(1)}}$

are such that $X_i \leq \epsilon (n^*)^{1/3-1/p}$

But $M_{n'} > \varepsilon (n')^{1/3 - 1/p}$ implies there is $n_{(2)}$ such that

$$\begin{cases} n_{(1)} < n_{(2)} < n' \\ X_{n_{(2)}} > \varepsilon n_{(2)}^{1/3 - 1/p} \end{cases}$$

And so on...

We can build in this way a sequence $\{X_{n_{(i)}}\}$ such that

$$(\forall n_{(i)}) \frac{X_{n_{(i)}}}{n_{(i)}^{1/3 - 1/p}} > \varepsilon$$

This yields a contradiction with the earlier result:

$$P\left[\frac{X_n}{n^{1/3 - 1/p}} > \varepsilon i.o.\right] = 0$$

Then

$$(\exists N_2) (\forall N > N_2) \left(\frac{M_N}{N^{1/3 - 1/p}} < \varepsilon \right)$$

Since ε was arbitrarily chosen this gives:

$$\frac{\max_{j=1, N} |\text{Log } h(\beta, z_j)|}{n^{1/3 - 1/p}} \xrightarrow{\text{a.s.}} 0$$

Lemma 7:

R.T. Rockafellar (1970) proved that if the continuous functions F_n are convex on an open set S , if they converge pointwise to a function F on a subset D dense in S , then the functions F_n converge uniformly to F on any compact set included in S .

As this result is not widely known, we have included in the appendix a proof of the following simpler result: if the continuous functions F_n are convex on an open set S , if they converge pointwise to a function F on S , then the location of the maximum of F_n converges to the location of the maximum of F .

B. $\Phi_N(\beta)$ converges point-wise to $\Phi(\beta)$ when $N \rightarrow \infty$

We know that:

$$\Phi_N(\beta) = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} h(\beta, z_i) - \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} \left[\frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_i) h(\beta, z_j) \right]$$

Let us first consider the first term

$$\frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} h(\beta, z_i)$$

The terms $\delta_i \text{Log} h(\beta, z_i)$ are i.i.d random variables.

Therefore we can apply the strong law of large numbers

to prove that $\frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} h(\beta, z_i)$ converges to

$$E[\delta \text{Log} h(\beta, z)]$$

which has been computed in lemma 1.

Now let us consider

$$\frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} \left[\frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_i) h(\beta, z_j) \right]$$

That term is a sum of dependant random variables, we want

to replace it by a sum of i.i.d random variables. More

precisely, let $f_N(w) = \frac{1}{N} \sum_{j=1}^N 1(y_j \geq w) h(\beta, z_j)$

and let $f(w) = E[h(\beta, z) 1(y \geq w)]$,

$$\hat{\alpha}_N = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} f(\tau_i) \quad , \quad \alpha_N = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} f(\tau_i)$$

Let us prove that $\hat{\alpha}_N$ converges point-wise to α_N .

(i.e. $\hat{\alpha}_N - \alpha_N$ converges to 0 pointwise as function

of β) The proof of this convergence is an essential

part of this chapter on Consistency. This proof follows.

Then we can apply the strong law of large numbers to

α_N to prove that α_N converges point-wise to $E[\delta \text{Log} f(\tau)]$

which has been computed in Lemma 3.

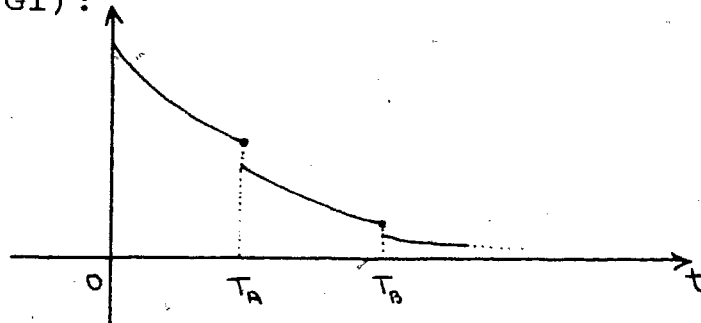
Proof of the convergence of $\hat{\alpha}_N$ to α_N :

Remember that we used the notation $f(t) = E[1(y \geq t) h(p, z)]$

As we allowed mass-points in the density of the censoring time, $f(t)$ is only assumed to be left-continuous. But we shall assume there is only a finite number of discontinuities: after a certain time $f(t)$ is continuous. Let us give

an example of the function $f(t)$. Consider a medical experiment, where sex and age are among the covariates.

At a specified time T_A all the women, for example, are withdrawn from the experiment, at T_B all the men over 40 years are withdrawn. In this example $f(t)$ would have the graph (G1):



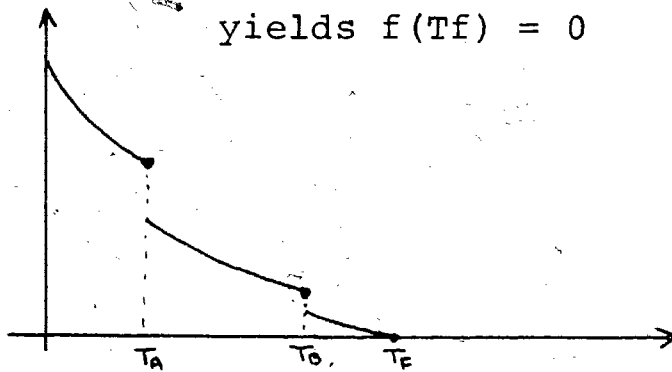
In order to prove the convergence of $\hat{\alpha}_N$ to α_N , we shall have to consider different cases depending on the shape of $f(t)$.

Case A: There is a time T_f such that

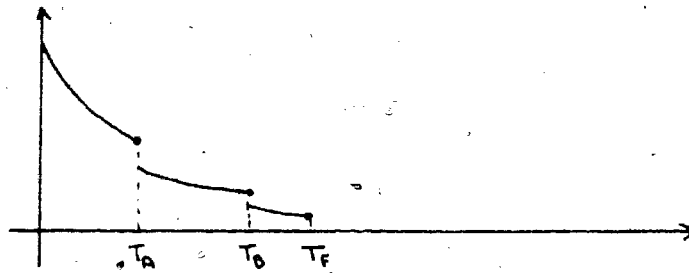
$$(\forall t > T_f) f(t) = 0$$

The experiment is stopped at a certain specified time T_f . We shall have to consider two different subcases:

Subcase A1: $f(t)$ is continuous at T_f which yields $f(T_f) = 0$

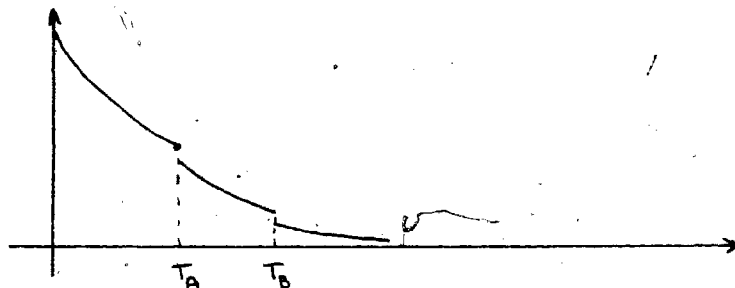


Subcase A2: $f(t)$ is not continuous at T_f , $f(T_f) > 0$



At T_f we stop the experiment, every subject not withdrawn is censored. This Subcase A2 is the only case studied by Tsiatis (1978). It is the most commonly encountered situation.

Case B: $(\forall t) f(t) > 0$



It is in this case that the proof of the convergence of $\hat{\alpha}_N$ to α_N is the most difficult. To prove the consistency in this case is interesting first for theoretical reasons. If we stop an experiment at T_f , we obtain by Cox's technique an estimate $\hat{\beta}_N$. But we didn't consider what is happening after T_f . The theoretical estimate β_N^* obtained by

observing the experiment till $t \rightarrow \infty$ is the estimate using all the information useful for Cox's technique. It would be quite worrisome if $\hat{\beta}_N^*$ was not consistent. Furthermore if the asymptotic approximations are to be applied in situations where the experiment ends with the death or failure of virtually all the subjects, then we will not have confidence in the approximations unless we can establish consistency in this case B.

Case A there is a time T_f such that $(\forall t > T_f) f(t) = 0$

Subcase A1: $f(T_f) = 0$

Let us prove that this subcase A1 can be treated as a case B by changing the time axis. For example let us change t in t' according to $t' = \frac{1}{T_f - t} - \frac{1}{T_f}$, the range of

t' is then $[0, +\infty[$.

The order of times of death is unchanged. As Cox's likelihood uses only the order of time of death, Cox's likelihood is unchanged.

Let us prove that the structure of the model is not changed: we still have a proportional hazard rate model.

Let us compute the hazard rate $\lambda^*(t', z)$ of the subject with covariate z at time t' .

Remember the density of t given z is:

$$\varphi_z(t) = \lambda_0(t) h(\beta_0, z) \exp\left(-\int_0^t \lambda_0(u) h(\beta_0, z) du\right)$$

We have $t = T_f - \frac{1}{t' + \frac{1}{T_f}}$ this gives $dt = \frac{1}{(t' + \frac{1}{T_f})^2} dt'$

Then density of t' given z is:

$$\varphi_z^*(t') = \frac{\lambda_0\left(T_f - \frac{1}{t' + \frac{1}{T_f}}\right) h(\beta_0, z) \exp\left(-\int_0^{T_f - \frac{1}{t' + \frac{1}{T_f}}} \lambda_0(u) h(\beta_0, z) du\right)}{\left(t' + \frac{1}{T_f}\right)^2}$$

$$\text{Let } F(y^*) = P[t' \leq y^*] = \int_0^{y^*} \varphi_z^*(t') dt' = \int_0^{T_f - \frac{1}{y^* + \frac{1}{T_f}}} \varphi_z(\alpha) d\alpha = 1 - \exp\left(-\int_0^{T_f - \frac{1}{y^* + \frac{1}{T_f}}} \lambda_0(u) h(\beta_0, z) du\right)$$

Therefore:

$$\lambda^*(t, z) = \frac{\varphi_z^*(t)}{1 - F(t)}, \quad \text{we have}$$

$$\lambda^*(t, z) = \frac{\lambda_0 \left(T_F - \frac{1}{t + 1/T_F} \right)}{\left(t + \frac{1}{T_F} \right)^2} h(\beta, z)$$

We still have a proportional hazard rate model with the same parametrization h .

Subcase A2: $f(T_F) > 0$

We want to prove that $\hat{\alpha}_N = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f_N(T_i)$ converges in probability to $\alpha_N = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f(T_i)$

$$\text{We have: } \left| \hat{\alpha}_N - \alpha_N \right| \leq \frac{1}{N} \sum_{i=1}^N \left| \text{Log } \frac{f_N(T_i)}{f(T_i)} \right|$$

$$\text{We know that } \left| \text{Log } \frac{f_N(T_i)}{f(T_i)} \right| = \left| \frac{f_N(T_i) - f(T_i)}{f(T_i)} \right| + R \left(\frac{f_N(T_i) - f(T_i)}{f(T_i)} \right)$$

$R(x)$ means a function of x of smaller order than x .

$(\forall T_i) f(T_i) \geq f(T_F)$ as $f(t)$ is a decreasing function of t , and $T_i \leq T_F$

$$\text{Therefore } \left| \frac{f_N(T_i) - f(T_i)}{f(T_i)} \right| \leq \left| \frac{f_N(T_i) - f(T_i)}{f(T_F)} \right|$$

But lemma 5 and the Skorokhod construction yields that

$$\sup_t |f_n(t) - f(t)| \text{ converges to 0 almost surely.}$$

Therefore

$$(\exists N_0) (\forall N > N_0) \left| \frac{f_n(T_i) - f(T_i)}{f(T_F)} \right| < \varepsilon$$

This proves the convergence in probability of $\hat{\alpha}_N$ to α_N in this subcase A2.

Case B

Remember $\hat{\alpha}_N = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f_N(T_i)$, $\alpha_N = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f(T_i)$

The difficulty here comes from the fact that when $T_i \rightarrow +\infty$, $f(T_i) \rightarrow 0$, $\text{Log } f(T_i) \rightarrow -\infty$

To overcome that difficulty, we are going to divide $\hat{\alpha}_N$ and α_N in two terms: $\hat{\alpha}_N = \hat{a}_n + \hat{b}_n$, $\alpha_N = a_n + b_n$

with

$$\hat{a}_n = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f_n(T_i) \mathbb{1}(T_i \leq t_n); \quad \hat{b}_n = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f_n(T_i) \mathbb{1}(T_i > t_n)$$

$$a_n = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f(T_i) \mathbb{1}(T_i \leq t_n); \quad b_n = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f(T_i) \mathbb{1}(T_i > t_n)$$

We are going to choose the sequence $\{t_n\}$ going fast enough to $+\infty$ so that \hat{b}_n and b_n converges to 0, and slow enough so that \hat{a}_n converges to a_n . In this case B, the theorem will be proved in 3 steps.

Step 1: $\hat{b}_n \longrightarrow 0$ for a suitable sequence $\{t_n\}$

Step 2: $b_n \longrightarrow 0$ for any $\{t_n\}$ such that $t_n \rightarrow \infty$

Step 3: $\hat{a}_n \longrightarrow a_n$ point wise in β for the sequence $\{t_n\}$ (step 1)

We have to point out that the convergence of \hat{a}_n to a_n is quite straightforward when t_n is fixed: it is nothing else than the case A2.

Step 1: Let $\{T_i\}$ set of times of deaths such that $\delta_i = 1$. Let us assume first that $t_n < \text{Sup } T_i$. Then there are $\{T_k\}$ such that $T_k > t_n$.

We are going to find a lower-bound and an upper-bound for $f_N(\tau_k)$.

$$f_N(\tau_k) = \frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_k) h(\beta, z_j)$$

Let $m(\tau_k)$ be the number of individuals in the experiment such that $y_j \geq \tau_k$. Obviously $1 \leq m(\tau_k) \leq m(t_n)$

Then $\text{Min}_{j=1, N} \frac{h(\beta, z_j) \cdot m(\tau_k)}{N} \leq f_N(\tau_k)$ which gives $f_N(\tau_k) \geq \frac{\text{Min}_{j=1, N} h(\beta, z_j)}{N}$

We have found a common lower-bound. Now let us find a common upper-bound.

Let ϵ given. There exists a compact $K_{\beta, \epsilon}$ depending on β such that $E[h(\beta, z) 1(z \notin K)] < \frac{\epsilon}{2}$ as $E[h(\beta, z)]$ exists.

$$f_N(\tau_k) = \frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_k) h(\beta, z_j) 1(z_j \in K_{\beta, \epsilon}) + \frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_k) h(\beta, z_j) 1(z_j \notin K_{\beta, \epsilon})$$

We can prove that the 2nd term is bounded by ϵ as:

$$\frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_k) h(\beta, z_j) 1(z_j \notin K_{\beta, \epsilon}) \leq \frac{1}{N} \sum_{j=1}^N h(\beta, z_j) 1(z_j \notin K_{\beta, \epsilon})$$

$$(\exists N_0(\beta)) (\forall N > N_0(\beta)) \left| \frac{1}{N} \sum_{j=1}^N h(\beta, z_j) 1(z_j \notin K_{\beta, \epsilon}) - E[h(\beta, z) 1(z \notin K_{\beta, \epsilon})] \right| < \frac{\epsilon}{2}$$

$$(\forall N > N_0(\beta)) \frac{1}{N} \sum_{j=1}^N h(\beta, z_j) 1(z_j \notin K_{\beta, \epsilon}) < \epsilon$$

$$\frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_k) h(\beta, z_j) 1(z_j \in K_{\beta, \epsilon}) \leq \text{Max}_{z \in K_{\beta, \epsilon}} h(\beta, z)$$

Now

$$(\forall N > N_0(\beta)) \frac{1}{N} \text{Min} h(\beta, z_j) \leq f_N(\tau_k) \leq \epsilon + \text{Max}_{z \in K_{\beta, \epsilon}} h(\beta, z)$$

From the bounds of $f_N(\tau_k)$, we shall deduce

bounds on \hat{b}_N

$$\text{Log} \left[\text{Min}_{j=1, N} h(\beta, z_j) \right] - \text{Log } N \leq \text{Log } f_N(t_n) \leq \text{Log} \left[\epsilon + \text{Max}_{z \in K_{\epsilon, \beta}} h(\beta, z) \right]$$

Remember

$$\hat{b}_N = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log } f_N(T_i) \mathbb{1}(T_i > t_n)$$

Consider $M^*(t_n)$, the number of individuals such that $\delta = 1$ and

$T > t_n$. Then:

$$\left(\text{Log} \left[\text{Min}_{j=1, N} h(\beta, z_j) \right] - \text{Log } N \right) \frac{M^*(t_n)}{N} \leq \hat{b}_N(t_n) \leq \frac{M^*(t_n)}{N} \left(\text{Log} \left[\epsilon + \text{Max}_{z \in K_{\epsilon, \beta}} h(\beta, z) \right] \right)$$

Remember we want to prove that $\hat{b}_N(t_n) \rightarrow 0$

for a sequence $\{t_n\}$. It is sufficient to choose $\{t_n\}$

such that $\frac{\log N}{N} M^*(t_n) \rightarrow 0$ and $\text{Log} \left[\text{Min}_{j=1, N} h(\beta, z_j) \right] \frac{M^*(t_n)}{N} \rightarrow 0$

In the remainder of the theorem it will be shown we need

to choose the sequence $\{t_n\}$ such that $\sqrt{n} f(t_n) \rightarrow +\infty$

Let us choose $\{t_n\}$ such that $\sqrt{n} f(t_n) = \text{Log } n$. This is

possible as there is a finite number of discontinuities

and we are in the case $(\forall t) f(t) > 0$

$$f(t_n) = E[h(\beta, z) \mathbb{1}(y \geq t_n)] = \frac{\text{Log } n}{\sqrt{n}}$$

We need two more steps to prove that $\hat{b}_N(t_n)$ converges

to 0 almost surely.

Step 1(a): $\frac{\text{Log } n}{n} M^*(t_n) \xrightarrow{\text{a.s.}} 0$ in probability

Step 1(b): $\text{Log} \left[\text{Min}_{j=1, N} h(\beta, z_j) \right] \frac{M^*(t_n)}{n} \xrightarrow{\text{a.s.}} 0$
in probability

Step 1(a):

Let $M(t_n)$ be the number of individuals such

that $y \geq t_n$. $M^*(t_n) \leq M(t_n)$

When N goes to $+\infty$, $\frac{M(t_n)}{N}$ converges to $E[1(y \geq t_n)]$

But $E[1(y \geq t_n)] = E[h(\beta, z)^{-1/2} h(\beta, z)^{1/2} 1(y \geq t_n)]$

Remember Hölder's inequality: $E|xy| \leq E^{1/r} |x|^r \cdot E^{1/s} |y|^s$

where $\frac{1}{r} + \frac{1}{s} = 1$

Let $r=6$, $s=6/5$

Then we obtain

$$E[1(y \geq t_n)] \leq E[h(\beta, z)^{-3}]^{1/6} E[h(\beta, z)^{3/5} 1(y \geq t_n)]^{5/6}$$

But:

$$E[h(\beta, z)^{3/5} 1(y \geq t_n)] = E[h(\beta, z)^{4/5} 1(y \geq t_n) h(\beta, z)^{-1/5}]$$

Let us apply again Hölder's inequality: $r=5/4$, $s=5$

Then:

$$E[h(\beta, z)^{3/5} 1(y \geq t_n)] \leq E[h(\beta, z) 1(y \geq t_n)]^{4/5} E[h(\beta, z)^{-1}]^{1/5}$$

Therefore:

$$E[1(y \geq t_n)] \leq E[h(\beta, z)^{-3}]^{1/6} E[h(\beta, z)^{-1}]^{1/6} E[h(\beta, z) 1(y \geq t_n)]^{2/3}$$

We assumed that $E[h(\beta, z)^{-3}]$ and $E[h(\beta, z)^{-1}]$ existed.

Therefore if $c = E[h(\beta, z)^{-3}]^{1/6} E[h(\beta, z)^{-1}]^{1/6}$, as

$$E[h(\beta, z) 1(y \geq t_n)] = \frac{\text{Log } n}{n}$$

we have:

$$E[1(y \geq t_n)] \leq c \frac{(\text{Log } n)^{2/3}}{n^{1/3}}$$

Therefore $\text{log } n E[1(y \geq t_n)]$ converges to 0.

Lemma 5 yields that

$$\text{Sup}_w \left| \text{Log } n \left\{ \frac{M(w)}{n} - E[1(y \geq w)] \right\} \right| \rightarrow 0 \text{ in probability}$$

Therefore $\frac{\text{log } n M(t_n)}{n} \rightarrow 0$ But $M^*(t_n) \leq M(t_n)$

Therefore we proved step (1a): $\text{Log } n \frac{M^*(t_n)}{n} \rightarrow 0$ in probability

The assumptions we made on the moments of $h(\beta, z)$

$(E[h(\beta, z)^{-3}], E[h(\beta, z)^{-1}], E[h(\beta, z)^3])$ are only a compromise between the assumptions needed in this step 1(a) and

those needed in lemma (6) where $\sum_{n=1}^{+\infty} \left(\frac{1}{n^{1/3-1/p}} \right)^k$ has to converge. By using different values of r and s in

Holder's inequalities, we could obtain different necessary values of the order of moments assumed converging. It

must be pointed out that in the exponential case, this problem is of no importance as all moments exist if we

simply assume $E[e^{\beta'z}]$ for all β .

Step 1(b): We want to prove that

$$\text{Log} \left[\text{Min}_{h(\beta, z)} \frac{M^*(t_n)}{n} \right] \text{ converges to } 0. \text{ We have}$$

$$\left| \text{Log} \left[\text{Min}_{j=1, n} h(\beta, z_j) \right] \right| \leq \text{Max}_{j=1, n} \left| \text{Log} h(\beta, z_j) \right|$$

We know $M^*(t_n) \leq M(t_n)$

Therefore

$$0 \leq \left| \text{Log} \left(\text{Min}_{j=1, n} h(\beta, z_j) \right) \right| \frac{M^*(t_n)}{N} \leq \text{Max}_{j=1, n} \left| \text{Log} h(\beta, z_j) \right| \frac{M(t_n)}{n}$$

But

$$\left(\text{Max}_{j=1, N} \left| \text{Log} h(\beta, z_j) \right| \right) \frac{M(t_n)}{N} \leq \text{Max}_{j=1, N} \left| \text{Log} h(\beta, z_j) \right| \left\{ E[1(y \geq t_n)] + \text{Sup}_w \left| \frac{M(w)}{N} - E[1(y \geq w)] \right| \right\}$$

Remember $E[1(y \geq t_n)] \leq c \frac{\text{Log} n}{n^{2/3}}$

Therefore

$$\text{Max}_{j=1, N} \left| \text{Log} h(\beta, z_j) \right| E[1(y \geq t_n)] \leq \frac{\text{Max}_{j=1, N} \left| \text{Log} h(\beta, z_j) \right|}{n^{1/3-1/p}} \cdot \frac{n^{1/3-1/p}}{n^{2/3}} \cdot c \cdot (\text{Log} n)^{2/3}$$

lemma 6 proved $\text{Max}_{j=1, N} \left| \text{Log} h(\beta, z_j) \right| \cdot n^{-1/3+1/p}$ converges to 0.

Moreover from lemma 6:

$$\text{Sup}_w \left\{ \text{Max}_{j=1, N} \left| \text{Log} h(\beta, z_j) \right| \cdot \frac{\sqrt{n}}{\sqrt{n}} \left| \frac{M(w)}{n} - E[1(y \geq w)] \right| \right\} \rightarrow 0$$

Then step 1(b) is proved. Step 1(a) and 1(b) yield that:

$$\hat{b}_n(t_n) \rightarrow 0$$

Step 2: $b_n \rightarrow 0$

Remember $b_n = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} f(\tau_i) \mid (\tau_i > t_n)$

We can apply

the S.L.L.N. to prove b_n converges to $E[\delta \text{Log} f(t) \mid (t > t_n)]$

for t_n fixed. But $E[\delta \text{Log} f(t)]$ exists. Then when

$t_n \rightarrow +\infty$, $E[\delta \text{Log} f(t) \mid (t > t_n)]$ converges to 0.

Thus $b_n \rightarrow 0$

Step 3: $\hat{a}_n \rightarrow a_n$, point wise in p :

Remember

$$\hat{a}_n = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} f_N(\tau_i) \mid (\tau_i \leq t_n) ; a_n = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} f(\tau_i) \mid (\tau_i \leq t_n)$$

Therefore

$$|\hat{a}_n - a_n| \leq \frac{1}{N} \sum_{i=1}^N \left| \text{Log} \frac{f_N(\tau_i)}{f(\tau_i)} \right| \mid (\tau_i \leq t_n)$$

We know that:
$$\left| \text{Log} \frac{f_N(\tau_i)}{f(\tau_i)} \right| = \left| \frac{f_N(\tau_i) - f(\tau_i)}{f(\tau_i)} \right| + R\left(\frac{f_N(\tau_i) - f(\tau_i)}{f(\tau_i)}\right)$$

$R(x)$ means a function of x of smaller order than x .

$$(\forall \tau_i \leq t_n) f(\tau_i) \geq f(t_n)$$

therefore
$$\left| \frac{f_N(\tau_i) - f(\tau_i)}{f(\tau_i)} \right| \leq \left| \frac{f_N(\tau_i) - f(\tau_i)}{f(t_n)} \right|$$

But $\frac{1}{f(t_n)} = \frac{\sqrt{n}}{\text{Log} n}$

then

$$\left| \frac{f_N(\tau_i) - f(\tau_i)}{f(\tau_i)} \right| = \frac{\sqrt{n}}{\text{Log} n} |f_N(\tau_i) - f(\tau_i)|$$

But lemma 5 yields:

$$\sup_t a_n \sqrt{n} |f_N(t) - f(t)| \rightarrow 0 \text{ in}$$

probability provided that $a_n \rightarrow 0$. Let ε be fixed.

Then:

$$(\exists N_0)(\forall N > N_0) \left| \frac{f_N(\tau_i) - f(\tau_i)}{f(t_n)} \right| < \varepsilon . \text{ It yields } \left| \frac{f_N(\tau_i) - f(\tau_i)}{f(\tau_i)} \right| < \varepsilon$$

$$\text{Then } \left| \text{Log} \frac{f_N(\tau_i)}{f(\tau_i)} \right| \leq 2\varepsilon$$

$$\text{Then } |\hat{a}_N - a_N| \leq \frac{2}{N} \varepsilon N$$

We have proved step 3: \hat{a}_N converges to a_N . Therefore, steps 1, 2, 3, imply that $\hat{\alpha}_N$ converges to α_N in case B.

We then have proved:

$$\Phi_N(\beta) = \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} h(\beta, z_i) - \frac{1}{N} \sum_{i=1}^N \delta_i \text{Log} \left[\frac{1}{N} \sum_{j=1}^N 1(y_j \geq \tau_i) h(\beta, z_j) \right]$$

converges point wise in probability to

$$\Phi(\beta) = E_z \left[\text{Log} h(\beta, z) E[\delta|z] \right] - E_{t,z} \left[\text{Log} f(t) G(t|z) \right]$$

with

$$f(w) = E \left[h(\beta, z) 1(y \geq w) \right] = E_z \left[h(\beta, z) G(w|z) \exp - \int_0^w \lambda_0(u) h(\beta_0, z) du \right]$$

In the next two paragraphs we shall study the function

$\Phi(\beta)$. As we don't know the shape of $h(\beta, z)$, we will only be able to prove some results in the neighbourhood of β_0 , namely:

- (i) β_0 is an extremum of $\Phi(\beta)$, and
- (ii) $\Phi(\beta)$ is concave in a neighbourhood V of β_0 .

C β_0 is an extremum of $\phi(\beta)$

We shall prove $\left(\frac{\partial \phi}{\partial \beta}\right)_{\beta_0} = 0$

Let β_i the i th coordinate of β . Let us prove $\left(\frac{\partial \phi}{\partial \beta_i}\right)_{\beta_0} = 0$

$$\left(\frac{\partial \phi}{\partial \beta_i}\right)_{\beta_0} = E_z \left[E[\delta|z] \frac{\left(\frac{\partial h(\beta, z)}{\partial \beta_i}\right)_{\beta_0}}{h(\beta_0, z)} \right] - E_{t,z} \left[G(t|z) \frac{\left(\frac{\partial f}{\partial \beta_i}\right)_{\beta_0}}{f(t)} \right] = A - B$$

Remember: - density of t given z is:

$$\lambda_0(t) h(\beta_0, z) \left(\exp - \int_0^t \lambda_0(u) h(\beta_0, z) du \right) dt$$

- density of z is $\chi(z) dz$

B is a double integral in t and z . By using

Fubini's theorem, we can express it as:

$$\int_{t=0}^{t=+\infty} \lambda_0(t) \frac{\left(\frac{\partial f}{\partial \beta_i}\right)_{\beta_0}}{f(t)} \left[\int_{z=-\infty}^{z=+\infty} h(\beta_0, z) \left(\exp - \int_0^t \lambda_0(u) h(\beta_0, z) du \right) G(t|z) \chi(z) dz \right] dt$$

The term inside the brackets is nothing else than $f(t)$.

Then
$$B = \int_{t=0}^{t=+\infty} \lambda_0(t) \left(\frac{\partial f(t)}{\partial \beta_i}\right)_{\beta_0} dt$$

Now let us consider
$$A = E_z \left[E[\delta|z] \frac{\left[\frac{\partial h(\beta, z)}{\partial \beta_i}\right]_{\beta_0}}{h(\beta_0, z)} \right]$$

Remember we proved in lemma 1 that:

$$E[S|z] = \int_{t=0}^{t=\infty} \lambda_0(t) h(\beta_0, z) \left(\exp - \int_0^t \lambda_0(u) h(\beta_0, z) du \right) G(t|z) dt$$

We can express as before A in a double integral in t and z:

$$A = \int_{t=0}^{t=\infty} \lambda_0(t) \left[\int_{z=-\infty}^{z=\infty} \left(\frac{\partial h}{\partial \beta_i} \right)_{\beta_0} \left(\exp - \int_0^t \lambda_0(u) h(\beta_0, z) du \right) G(t|z) X(z) dz \right] dt$$

But the term inside brackets is $\left(\frac{\partial f(t)}{\partial \beta_i} \right)_{\beta_0}$. Thus

A = B and so

$$\left(\frac{\partial \Phi}{\partial \beta_i} \right)_{\beta_0} = 0$$

β_0 is an extremum of $\Phi(\beta)$

As yet we do not know if it is a maximum or minimum nor do we know if it is the only extremum of $\Phi(\beta)$.

D Concavity of $\Phi(\beta)$ in a neighbourhood $V(\beta_0)$ of β_0 .

We assumed that $\frac{\partial^2 \Phi}{\partial \beta^2}$ was continuous. Therefore to prove that there is a neighbourhood $V(\beta_0)$ of β_0 , on which

$\Phi(\beta)$ is concave, it is sufficient to prove that $\left(\frac{\partial^2 \Phi}{\partial \beta^2}\right)$ is a positive definite matrix at $\beta = \beta_0$.

Let us consider $\frac{\partial^2 \Phi}{\partial \beta_i \partial \beta_k}$

$$\begin{aligned} \frac{\partial^2 \Phi}{\partial \beta_i \partial \beta_k} &= E_z \left[E[\delta|z] \frac{\frac{\partial^2 h}{\partial \beta_i \partial \beta_k}}{h(\beta, z)} \right] - E_z \left[\frac{E[\delta|z]}{h^2} \left(\frac{\partial h}{\partial \beta_i} \right) \left(\frac{\partial h}{\partial \beta_k} \right) \right] \\ &\quad - E_{z,t} \left[\frac{G(t|z)}{f(t)} E_z \left[\frac{\partial^2 h}{\partial \beta_i \partial \beta_k} \left(\exp - \int_0^t \lambda_0(u) h(\beta, z) du \right) G(t|z) \right] \right] \\ &\quad + E_{z,t} \left[\frac{G(t|z)}{f(t)^2} E_z \left[\frac{\partial h}{\partial \beta_i} \left(\exp - \int_0^t \lambda_0(u) h(\beta, z) du \right) G(t|z) \right] E_z \left[\frac{\partial h}{\partial \beta_k} G(t|z) \exp - \int_0^t \lambda_0(u) h du \right] \right] \end{aligned}$$

Let

$$\frac{\partial^2 \Phi}{\partial \beta_i \partial \beta_k} = A_{ik} - B_{ik} - C_{ik} + D_{ik}$$

We shall prove $\left(\frac{\partial^2 \Phi}{\partial \beta_i \partial \beta_k}\right)_{\beta_0}$ is a positive definite matrix in two steps. In the first step we shall prove $A_{ik} = B_{ik}$.

In the second step we shall prove $[D_{ik}] - [B_{ik}]$ is a positive definite matrix under a mild assumption on the set of z 's.

First step: $A_{ik} = C_{ik}$

We shall use arguments similar to C. C_{ik} is a double integral in t and z that can be expressed as:

$$C_{ik} = \int_{t=0}^{t=\infty} \frac{\lambda_0(t)}{f(t)} E_z \left[\frac{\partial^2 h}{\partial p_i \partial p_k} \exp - \int_0^t \lambda_0(u) h(p_0, z) du \cdot G(t|z) \right] E_z \left[h(p_0, z) \exp - \int_0^t \lambda_0(u) h(p_0, z) du \cdot G(t|z) \right]$$

But if we remember that $f(t) = E_z \left[h(p_0, z) G(t|z) \exp - \int_0^t \lambda_0(u) h(p_0, z) du \right]$

$$C_{ik} = \int_{t=0}^{t=\infty} \lambda_0(t) E_z \left[\frac{\partial^2 h}{\partial p_i \partial p_k} \left(\exp - \int_0^t \lambda_0(u) h(p_0, z) du \right) G(t|z) \right] dt$$

Remember that

$$E[S|z] = \int_{t=0}^{t=\infty} \lambda_0(t) h(p_0, z) \left(\exp - \int_0^t \lambda_0(u) h(p_0, z) du \right) G(t|z) dt$$

Therefore

$$A_{ik} = E_z \left[\frac{\partial^2 h}{\partial p_i \partial p_k} \int_{t=0}^{t=\infty} \lambda_0(t) G(t|z) \left(\exp - \int_0^t \lambda_0(u) h(p_0, z) du \right) dt \right]$$

By inverting the order of the integrals A_{ik} can be transformed into C_{ik} .

$$\underline{A_{ik} = C_{ik}}$$

Second step: $(D_{ik}) - (B_{ik})$ is a definite positive matrix.

By similar arguments as before, we can express B-D as:

$$\int_{t=0}^{t=\infty} \frac{\lambda_0(t)}{f(t)} \left[E_z \left[h(p_0, z) \exp - \int_0^t \lambda_0(u) h(p_0, z) du \cdot G(t|z) \right] E_z \left[\frac{G(t|z)}{h(p_0, z)} \left(\frac{\partial h}{\partial p} \right)^2 \exp - \int_0^t \lambda_0(u) h(p_0, z) du \right] - E_z \left[\frac{\partial h}{\partial p} \left(\exp - \int_0^t \lambda_0(u) h(p_0, z) du \right) G(t|z) \right]^2 \right] dt$$

Let us call $M(t)$ the term in large brackets. In the appendix A5 it will be shown that $M(t)$ can be expressed as a double integral:

$$M(t) = \int_{x=-\infty}^{x=+\infty} \int_{z=x}^{z=+\infty} \left[\frac{\frac{\partial h}{\partial \beta}(\beta, z)}{h(\beta, z)} - \frac{\frac{\partial h}{\partial \beta}(\beta, x)}{h(\beta, x)} \right]^2 dp(t, z, x)$$

Therefore $B - D =$

$$B - D = \int_{t=0}^{t=+\infty} \frac{\lambda_0(t)}{f(t)} M(t) dt$$

- obviously $B - D$ is positive.

- let us prove $B - D$ is definite:

$$V(z, x) = \frac{\frac{\partial h}{\partial \beta}(\beta, z)}{h(\beta, z)} - \frac{\frac{\partial h}{\partial \beta}(\beta, x)}{h(\beta, x)}$$

Let

Let X be a vector.

$$\begin{aligned} X'(B-D)X = 0 &\Leftrightarrow (\forall x)(\forall z) X'(xA) \sqrt{X} \sqrt{zA} \sqrt{X} = 0 \Leftrightarrow (\forall x)(\forall z) \sqrt{xA} \sqrt{zA} \sqrt{X} = 0 \\ &\Leftrightarrow (\forall z)(\forall x) \frac{\frac{\partial h}{\partial \beta}(\beta, z)}{h(\beta, z)} X = \frac{\frac{\partial h}{\partial \beta}(\beta, x)}{h(\beta, x)} X \end{aligned}$$

Let us take x such that

$$\frac{\partial h}{\partial \beta}(\beta, x) = 0$$

This implies

$$(\forall z) \frac{\partial h}{\partial \beta}(\beta, z) X = 0$$

If there was such an X , it would mean we didn't need so many dimensions for the parameter β .

For example if $h(\beta, z) = e^{\beta z}$, $(\forall z) \frac{\partial h}{\partial \beta}(\beta, z) X = 0$ means

$$(\forall z) \sum_{k=1}^p z_k X_k = 0$$

, i.e. the z 's are in a hyperplane.

A covariate can be expressed as a linear combination of the others. Then we have no reason to consider this covariate; we don't need as many dimensions for the parameter β . We shall assume there is no such X. Then B-D is positive definite. Therefore -

$\left(\frac{\partial^2 \Phi}{\partial \beta^2} \right)_{\beta_0}$ is a positive definite matrix

Hence there is a compact neighbourhood $v(\beta_0)$ of β_0 , which $\Phi(\beta)$ is strictly concave.

E Exponential case: $h(\beta, z) = e^{\beta'z}$

We are now considering the special case $h(\beta, z) = e^{\beta'z}$.

This case is the most used model, the one used by Cox in his original paper. Let us prove that in this case,

$\Phi_n(\beta)$ is concave everywhere. For obvious reasons

$\Phi_n(\beta)$ has the same concavity as the following function

$$\Phi_n^*(\beta) = \sum_{i=1}^N \delta_i \beta'z_i - \sum_{i=1}^N \delta_i \text{Log} \left[\sum_{j \in R_i} e^{\beta'z_j} \right]$$

Let us study $\frac{\partial^2 \Phi_n^*}{\partial \beta^2}$:

$$-\frac{\partial^2 \Phi_n^*}{\partial \beta^2} = \sum_{i=1}^N \delta_i \left\{ \left[\sum_{j \in R_i} \left(\frac{e^{\beta'z_j}}{\sum_{j \in R_i} e^{\beta'z_j}} \right) z_j z_j' \right] - \left[\sum_{j \in R_i} z_j \left(\frac{e^{\beta'z_j}}{\sum_{j \in R_i} e^{\beta'z_j}} \right) \right] \left[\sum_{j \in R_i} z_j \left(\frac{e^{\beta'z_j}}{\sum_{j \in R_i} e^{\beta'z_j}} \right) \right]' \right\}$$

Let $A_i = \sum_{j \in R_i} z_j z_j' \left(\frac{e^{\beta'z_j}}{\sum_{j \in R_i} e^{\beta'z_j}} \right)$ and $B_i = \sum_{j \in R_i} z_j \left(\frac{e^{\beta'z_j}}{\sum_{j \in R_i} e^{\beta'z_j}} \right)$

Then $-\frac{\partial^2 \Phi_n^*}{\partial \beta^2} = \sum_{i=1}^N \delta_i [A_i - B_i B_i']$

Let T_i be a time of death. Given the risk set R_i , the conditional probability that subject k died is $\frac{e^{\beta'z_k}}{\sum_{j \in R_i} e^{\beta'z_j}}$

We can define the conditional

expectation of z given R_i as: $E[z | R_i] = \sum_{j \in R_i} z_j \left(\frac{e^{\beta'z_j}}{\sum_{j \in R_i} e^{\beta'z_j}} \right)$

$$E[z | R_i] = B_i$$

Similarly $E[zz' | R_i] = A_i$

Therefore $A_i - B_i B_i' = E[zz' | R_i] - E[z | R_i] E[z | R_i]'$

is the variance-covariance matrix of z given R_i .

$$\text{But } - \frac{\partial^2 \Phi_N^*}{\partial \beta^2} = \sum_{i=1}^N \delta_i [A_i - B_i B_i']$$

Therefore

$$\left(- \frac{\partial^2 \Phi_N^*}{\partial \beta^2} \right)$$

is a positive matrix.

Let us prove that under a mild assumption $-\frac{\partial^2 \Phi_N^*}{\partial \beta^2}$ is

definite. Let us consider:

$$M_k = E \left[(z - E[z | R_k]) (z - E[z | R_k])' | R_k \right]$$

Consider γ a vector

$$\gamma' M_k \gamma = 0 \Leftrightarrow E \left[\gamma' (z - E[z | R_k]) (z - E[z | R_k])' \gamma | R_k \right] = 0$$

$$\Leftrightarrow E \left[(\gamma' (z - E[z | R_k]))^2 | R_k \right] = 0 \Leftrightarrow (\forall z \in R_k) \gamma' (z - E[z | R_k]) = 0$$

It means all the z belonging to R_k are in a hyperplane.

Then as $-\frac{\partial^2 \phi_N^*}{\partial \beta^2}$ is a sum of matrices M_k , $-\frac{\partial^2 \phi_N^*}{\partial \beta^2}$ is

definite unless each of these M_k is not definite. That can happen only if all the z 's at $t=0$ are in a hyperplane. As we are looking for the asymptotic properties of ϕ_N , provided that the domain of definition of z is not a hyperplane (as in the discussion on the concavity of ϕ , it would mean that β has too many dimensions) for a sample size large enough, the matrix M_0 (M_k at $t=0$) is definite.

Therefore $-\frac{\partial^2 \phi_N^*}{\partial \beta^2}$ is positive definite.

This proves that:

ϕ_N is strictly concave on the whole domain of β

Then ϕ_N has only one maximum $\hat{\beta}_N$. As we can't infer anything about the shape of ϕ_N in the general case, it might be interesting to study some other particular models. For example, for a particular model, if ϕ_N has only one local maximum and if ϕ_N is concave in the neighbourhood of the true value, the proof of the consistency is similar to the proof of the consistency in the exponential case.

Let us go back to this exponential case: we

have proved that $\Phi(\beta)$ is concave on a compact neighbourhood V of its maximum β_0 (indeed in this case Φ is convex on all \mathbb{R}^r), that $\{\Phi_n(\beta)\}$ is a sequence of functions converging point-wise in probability to $\Phi(\beta)$ and concave on V .

To apply Lemma 7 we need to know $\Phi_n(\beta, \omega) \rightarrow \Phi(\beta, \omega)$ for all β in some countable dense set and for all ω except for a set $N \subset \Omega$ with $\Pr[N] = 0$. The point is that N must not depend on β .

We now argue that $\hat{\beta}_n \rightarrow \beta_0$ in probability by contradiction. If $\hat{\beta}_n$ does not converge to β_0 in probability, there is a subsequence n' such that

$$\Pr[|\hat{\beta}_{n'} - \beta_0| > \varepsilon] \geq \delta > 0 \quad \text{for some fixed } \varepsilon, \delta > 0.$$

Since $\Phi_n(\beta) \rightarrow \Phi(\beta)$ in probability for each β , it is possible to choose a subsequence n'' of n' such that

$$\Phi_{n''}(\beta) \rightarrow \Phi(\beta) \quad \text{for all } \beta \text{ with rational coordinates}$$

(the set of such β is a countable dense set) and for all ω except for a set $N \subset \Omega$, independent of β , with

$\Pr(N) = 0$. For this sub-subsequence Rockafellar's theorem

proves that $\Phi_{n''} \rightarrow \Phi$ uniformly on compacts. Hence

$$\hat{\beta}_{n''} \rightarrow \beta \quad \text{almost surely. Then } \Pr[|\hat{\beta}_{n''} - \beta_0| > \varepsilon] \rightarrow 0$$

which contradicts the selection of n'' as a subsequence of n' .

Therefore:

$\hat{\beta}_n$ is consistent in probability.

F General Case:

As we can prove very little about the shape of $\phi_N(\beta)$ in the general case, we shall use another tack. We shall prove that the sequence $\{\phi_N(\beta)\}$ is equicontinuous in a neighbourhood of β_0 . This will yield the uniform convergence of ϕ_N to ϕ . That will give us an "idea" about the shape of ϕ_N when N gets large. For reasons mathematical more than statistical, we shall assume that the time of censoring is bounded by a certain finite time T_f . This assumption is mild since most survival studies are ended after some prespecified time. We shall assume moreover that that time T_f is such that $P[Y \geq T_f] > 0$.

F1 Equicontinuity of $\{\phi_N(\beta)\}$ on a neighbourhood $V(\beta_0)$ of the true value β_0 .

We will need the following classical result:

Theorem: If E is a compact metric space and f_n an equicontinuous sequence of real valued functions of E converging pointwise for $x \in E$ in a countable dense subset of E to a continuous function f then $f_n \rightarrow f$ uniformly on E . We claim that $\{\phi_N(\beta)\}$ is an equicontinuous sequence on a compact neighbourhood of β_0 .

$$\phi_N(\beta) = \frac{1}{N} \sum_{i=1}^N \delta_i \log h(\beta, z_i) - \frac{1}{N} \sum_{i=1}^N \delta_i \log \left[\frac{1}{N} \sum_{j \in R_i} h(\beta, z_j) \right]$$

Let β_1 and β_2 be in a neighbourhood of β_0 .

$$\Phi_N(\beta_1) = \Phi_N(\beta_2) + (\beta_1 - \beta_2) \left(\frac{\partial \Phi_N}{\partial \beta} \right)_{\beta^*}$$

with β^* lying between

β_1 and β_2 . To prove the equicontinuity of $\{\Phi_N\}$ it is sufficient to prove $\left(\frac{\partial \Phi}{\partial \beta} \right)$ is bounded in $V(\beta_0)$.

Let us prove that $\left\| \frac{\partial \Phi_N}{\partial \beta} \right\|$ is bounded for β close enough to β_0 :

$$\frac{\partial \Phi_N}{\partial \beta} = \frac{1}{N} \sum_{i=1}^N \frac{\delta_i}{h(\beta, z_i)} \left[\frac{\partial h}{\partial \beta}(\beta, z_i) \right] = \frac{1}{N} \sum_{i=1}^N \frac{\left(\frac{1}{N} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j) \right)}{\left(\frac{1}{N} \sum_{j \in R_i} h(\beta, z_j) \right)}$$

First consider the 2nd term $B =$

$$B = \frac{1}{N} \sum_{i=1}^N \frac{\frac{1}{N} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j)}{\frac{1}{N} \sum_{j \in R_i} h(\beta, z_j)}$$

We prove B is bounded in 2 steps:

1st step: $\left\| \frac{1}{N} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j) \right\|$ bounded for N large

enough in a certain $V(\beta_0)$.

2nd step: $\frac{1}{N} \sum_{j \in R_i} h(\beta, z_j)$ bounded for N large enough

in a certain $V(\beta_0)$.

First step: We shall assume β is in hypercube H of center β_0 such that $(\forall k) |\beta_k - \beta_k^0| < \epsilon$. We shall use the hypotheses we made on $\frac{\partial h}{\partial \beta_k}$: $\frac{\partial h}{\partial \beta_k}$ is a monotone

function of β_l for any l . (β_l l th coordinate of β).

$$\left\| \frac{1}{N} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j) \right\| \leq \frac{1}{N} \sum_{j \in R_i} \left\| \frac{\partial h}{\partial \beta}(\beta, z_j) \right\|$$

The parameter β has p dimensions.

$$\left\| \frac{\partial h}{\partial \beta}(\beta, z_j) \right\| = \sqrt{\left[\frac{\partial h}{\partial \beta_1}(\beta, z_j) \right]^2 + \dots + \left[\frac{\partial h}{\partial \beta_p}(\beta, z_j) \right]^2}$$

$\frac{\partial h}{\partial \beta_i}(\beta, z_j)$ is a function of β : let $\beta_1, \beta_2, \dots, \beta_{q-1}, \beta_{q+1}, \dots, \beta_p$

be fixed. We then have a monotone function of β_q . Its maximum in $[\beta_0^q - \epsilon, \beta_0^q + \epsilon]$ is either at $\beta_0^q - \epsilon$, or at $\beta_0^q + \epsilon$. Therefore the maximum of $\frac{\partial h}{\partial \beta_i}$ in H is on

one of the vertices: $\beta_0 + \delta_r \epsilon$, δ_r being a vector such that $\delta_r^i = \pm 1$. There are 2^p different vectors δ_r , corresponding to the 2^p vertices of the hypercube. Let

\mathcal{L}_r be that set of vectors δ_r .

Then obviously:

$$\left[\frac{\partial h}{\partial \beta_i} (\beta, z_j) \right]^2 \leq \sum_{\tilde{\delta}_r \in \Lambda_r} \left[\frac{\partial h}{\partial \beta_i} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z_j) \right]^2$$

Then

$$\left\| \frac{\partial h}{\partial \beta} (\beta, z_j) \right\|^2 \leq \sum_{\tilde{\delta}_r \in \Lambda_r} \left\| \frac{\partial h}{\partial \beta} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z_j) \right\|^2$$

Then

$$\left\| \frac{\partial h}{\partial \beta} (\beta, z_j) \right\| \leq \sum_{\tilde{\delta}_r \in \Lambda_r} \left\| \frac{\partial h}{\partial \beta} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z_j) \right\|$$

Therefore:

$$\frac{1}{N} \sum_{j \in R} \left\| \frac{\partial h}{\partial \beta} (\beta, z_j) \right\| \leq \sum_{\tilde{\delta}_r \in \Lambda_r} \left\{ \frac{1}{N} \sum_{j=1}^N \left\| \frac{\partial h}{\partial \beta} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z_j) \right\| \right\}$$

For each $\tilde{\delta}_r$, $\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon$ is a fixed vector, so we can apply the Strong Law of Large Numbers. Therefore there exists

N_r such that:

$$\frac{1}{N} \sum_{j=1}^N \left\| \frac{\partial h}{\partial \beta} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z_j) \right\| \leq \varepsilon_r + E \left[\left\| \frac{\partial h}{\partial \beta} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z) \right\| \right]$$

But there exists only a finite number of N_r , then by

taking $N > \text{Max}_r N_r$, we have:

$$\sum_{\tilde{\delta}_r \in \Lambda_r} \left\{ \frac{1}{N} \sum_{j=1}^N \left\| \frac{\partial h}{\partial \beta} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z_j) \right\| \right\} \leq 2^p \varepsilon_r + \sum_{\tilde{\delta}_r \in \Lambda_r} E \left[\left\| \frac{\partial h}{\partial \beta} (\tilde{\beta}_0 + \tilde{\delta}_r \varepsilon, z) \right\| \right]$$

We assumed that $E \left[\left\| \frac{\partial h}{\partial \beta} (\beta, z) \right\| \right]$ exists for all β .

We therefore have proved that for $N > \text{Max}_r N_r$, and for

β in the hypercube H

$$\left\| \frac{1}{N} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j) \right\|$$

is bounded by $M = 2\epsilon_1 + \epsilon_1 E \left[\left\| \frac{\partial h}{\partial \beta}(\beta + \delta_{\epsilon_1}, z) \right\| \right]$

2nd step:

$$\frac{1}{N} \sum_{j \in R_i} h(\beta, z_j)$$

is bounded.

Let

$$C = \frac{1}{N} \sum_{j \in R_i} h(\beta, z_j) = \frac{1}{N} \sum_{j=1}^N I(y_j \geq T_i) h(\beta, z_j)$$

We assumed $T_i \leq T_F$ so $C \geq \frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) h(\beta, z_j)$

Let us prove that for N large enough $\frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) h(\beta, z_j)$

has a lower bound. Let us consider a compact K_n such that

$$P[z \in K_n] > 0$$

$$C \geq \frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) I(z_j \in K_n) h(\beta, z_j)$$

We could now apply the Strong Law of Large Numbers, but

$$\frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) I(z_j \in K_n) h(\beta, z_j)$$

$$E[I(z \in K_n) I(y \geq T_F) h(\beta, z)]$$

would be close to

for N depending on β .

We have to overcome this difficulty:

$$\frac{1}{N} \sum_{j \in I} I(y_j \geq T_F) I(z_j \in K_n) h(\beta, z_j) =$$

$$\frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) I(z_j \in K_n) h(\beta, z_j) + \frac{1}{N} \sum_{\substack{j=1 \\ z_j \in K_n}}^N I(y_j \geq T_F) [h(\beta, z_j) - h(\beta_0, z_j)]$$

Then there is $N(\beta_0)$ depending on β_0 such that:

$$(\forall N > N(\beta_0))$$

$$\left| \frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) I(z_j \in K_1) h(\beta_0, z_j) - E[I(z \in K_1) I(y \geq T_F) h(\beta_0, z)] \right| < \varepsilon_1$$

We can make the 2nd term $\frac{1}{N} \sum_{\substack{j=1 \\ z_j \in K_1}}^N I(y_j \geq T_F) [h(\beta_0, z_j) - h(\beta, z_j)]$ small compared to ε_1 , since:

$$\frac{1}{N} \sum_{\substack{j=1 \\ z_j \in K_1}}^N I(y_j \geq T_F) [h(\beta_0, z_j) - h(\beta, z_j)] \leq \sup_{z \in K_1} [h(\beta_0, z) - h(\beta, z)] \left[\frac{1}{N} \sum_{j=1}^N I(z_j \in K_1) I(y_j \geq T_F) \right]$$

$$\frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) I(z_j \in K_1) \text{ is close to } E[I(z \in K_1) I(y \geq T_F)]$$

for $N > N_1$.

$$\text{And } \sup_{z \in K_1} [h(\beta_0, z) - h(\beta, z)] < \varepsilon_1 \quad \|\beta - \beta_0\| < \alpha_1$$

Then for β such that $\|\beta - \beta_0\| < \alpha_2$, $N > N_2$

$$\frac{1}{N} \sum_{j=1}^N I(y_j \geq T_F) [h(\beta_0, z_j) - h(\beta, z_j)] < \varepsilon_1$$

$$\text{Let } \varepsilon_1 \leq E \left[\frac{I(z \in K_1) I(y \geq T_F) h(\beta_0, z)}{4} \right]$$

Then for $N > \sup\{N_2, N(\beta_0)\}$, for β such that

$$\|\beta - \beta_0\| < \alpha_2: \quad \frac{1}{N} \sum_{j \in R_1} h(\beta, z_j) \geq \frac{E[I(z \in K_1) I(y \geq T_F) h(\beta_0, z)]}{2}$$

Hence:

$$\|B\| = \left\| \frac{1}{N} \sum_{i=1}^N \delta_i \frac{\frac{1}{N} \sum_{j \in R_1} \frac{\partial h}{\partial \beta}(\beta, z_j)}{\frac{1}{N} \sum_{j \in R_1} h(\beta, z_j)} \right\| \leq M \left\| \frac{1}{N} \sum_{i=1}^N \delta_i \right\| \text{ for suitable } M$$

Therefore: B is bounded for a neighbourhood $v(\beta_0)$ of β_0 .

Now let us consider the term $A = \frac{1}{N} \sum_{i=1}^N \delta_i \frac{\frac{\partial h}{\partial \beta}(\beta, z_i)}{h(\beta, z_i)}$

As before

$$\|A\| \leq \frac{1}{N} \sum_{i=1}^N \frac{\delta_i}{h(\beta, z_i)} \left\| \frac{\partial h}{\partial \beta}(\beta, z_i) \right\| ; \left\| \frac{\partial h}{\partial \beta}(\beta, z_i) \right\| \leq \sum_{\delta_r \in \mathcal{L}_r} \left\| \frac{\partial h}{\partial \beta}(\beta_0 + \delta_r \varepsilon, z_i) \right\|$$

We have assumed too that $h(\beta, z)$ is monotone in β_1 , for z fixed, for any l .

$$\text{Then } \left(\exists \delta_r \in \mathcal{L}_r \right) \frac{1}{h(\beta, z_i)} \leq \frac{1}{h(\beta_0 + \delta_r \varepsilon, z_i)}$$

Therefore

$$\frac{1}{N} \sum_{i=1}^N \delta_i \frac{\left\| \frac{\partial h}{\partial \beta}(\beta_0 + \delta_r \varepsilon, z_i) \right\|}{h(\beta, z_i)} \leq \sum_{\delta_r^* \in \mathcal{L}_r} \left[\frac{1}{N} \sum_{i=1}^N \delta_i \frac{\left\| \frac{\partial h}{\partial \beta}(\beta_0 + \delta_r^* \varepsilon, z_i) \right\|}{h(\beta_0 + \delta_r^* \varepsilon, z_i)} \right]$$

There exists N_{r_1, r_2} such that for $(\delta_{r_1}, \delta_{r_2})$

$$\frac{1}{N} \sum_{i=1}^N \delta_i \frac{\left\| \frac{\partial h}{\partial \beta}(\beta_0 + \delta_{r_1} \varepsilon, z_i) \right\|}{h(\beta_0 + \delta_{r_2} \varepsilon, z_i)}$$

is bounded by

$$\varepsilon \left[\frac{\left\| \frac{\partial h}{\partial \beta}(\beta_0 + \delta_{r_1} \varepsilon, z) \right\|}{h(\beta_0 + \delta_{r_2} \varepsilon, z)} \right] + \varepsilon$$

$$\text{But } E \left[\frac{\int \left\| \frac{\partial h(\beta_0 + \delta_{r_1} \varepsilon, z)}{\partial \beta} \right\|}{h(\beta_0 + \delta_{r_2} \varepsilon, z)} \right] \leq \sqrt{E \left[\int \left\| \frac{\partial h(\beta_0 + \delta_{r_1} \varepsilon, z)}{\partial \beta} \right\|^2 \right]} \sqrt{E \left[\frac{1}{h(\beta_0 + \delta_{r_2} \varepsilon, z)^2} \right]}$$

$$\text{Then } E \left[\frac{\int \left\| \frac{\partial h(\beta_0 + \delta_{r_1} \varepsilon, z)}{\partial \beta} \right\|}{h(\beta_0 + \delta_{r_2} \varepsilon, z)} \right] \text{ exists. } \quad \downarrow$$

Therefore, as there is a finite number of $(\delta_{r_1}, \delta_{r_2})$ it yields that $\|(A)\|$ is bounded.

Therefore: $\left\| \frac{\partial \Phi_N}{\partial \beta} \right\|$ is bounded for β close enough to β_0 .

Let us recall what we proved in the general case:

- $\Phi_N(\beta)$ converges point-wise in probability to a given function $\Phi(\beta)$.
- The sequence $\{\Phi_N(\beta)\}$ is equicontinuous on a compact neighbourhood K_1 of the true value β_0 .
- β_0 is a maximum of $\Phi(\beta)$.
- $\Phi(\beta)$ is concave on a compact K_2 .

Let K be the intersection of K_1 and K_2 . An argument similar to the one used on page 49 (but using equicontinuity in place of Rockafellar's theorem) proves that Φ_N has a local maximum on K which converges to the true value in probability.

We therefore have proved:

The Maximum Partial Likelihood equation has a root that is consistent in probability.

In the case of several roots, further work might extend some results of the classical M.L.E. theory to the Cox partial likelihood technique. For example, we could examine the change in the sign of $\frac{\partial \phi_N}{\partial \beta}$ from positive to negative and searching the intervals in which these changes occur, to locate, evaluate and compare the maxima. V.D. Barnett (1966) discusses a systematic method of doing this, using the "Method of false positions". Another method (Le Cam) could be to find a consistent estimator $\hat{\beta}_N$ preferably easy to compute, then find either a nearby local maximum of Cox's likelihood or to do a one-step Newton Raphson from $\hat{\beta}_N$ towards the maximum partial likelihood estimate. Some further study might show that as in the case of the maximum likelihood estimate, this new estimator is asymptotically minimum variance unbiased (i.e. fully efficient) when β has one dimension.

Chapter 3: Asymptotic Normality

Let β be the vector of parameters of interest and β_0 the true value of this parameter. Cox's likelihood is proportional to :

$$\phi_n(\beta) = \sum_{i=1}^N \delta_i \text{Log } h(\beta, z_i) - \sum_{i=1}^N \delta_i \text{Log} \left[\frac{1}{N} \sum_{j \in R_i} h(\beta, z_j) \right]$$

Let $\hat{\beta}_n$ be a consistent root of the partial likelihood equation. In this chapter we shall assume that there exists T_f such that the times of censoring are bounded by a certain time T_f , that $P[Y \geq T_f] > 0$ and

that $\frac{\partial h}{\partial \beta}(\beta, z)$, $\frac{\partial^2 h}{\partial \beta^2}$, $\frac{\partial^3 h}{\partial \beta_k \partial \beta_l \partial \beta_q}$ are monotone functions in β for z fixed and for any r . This last assumption is verified for the exponential model and more generally for any function h such that $h(\beta, z) = g(\beta'z)$ where $g(x)$ is a function such that $g(x)$, $g'(x)$, $g''(x)$ are monotone. For example, $h(\beta, z) = (1 + \beta'z)^{-1}$, $h(\beta, z) = 1 + \beta'z$ verify this assumption.

We shall prove that the random vector

$\sqrt{n}(\hat{\beta}_n - \beta_0)$ is asymptotically distributed as a multivariate normal random variable with mean zero and variance-covariance matrix V . An approach quite similar to the one used in classical M.L.E. theory will be used. Taylor's expansion gives us:

$$\left(\frac{\partial \phi_n}{\partial \beta}\right)_{\hat{\beta}_n} = \left(\frac{\partial \phi_n}{\partial \beta}\right)_{\beta_0} + \left(\frac{\partial^2 \phi_n}{\partial \beta^2}\right)_{\tilde{\beta}} (\hat{\beta}_n - \beta_0)$$

where $\tilde{\beta}$ lies between $\hat{\beta}_n$ and β_0 . As $\hat{\beta}_n$ is a root of the partial likelihood equation, we have $\left(\frac{\partial \phi_n}{\partial \beta}\right)_{\hat{\beta}_n} = 0$.

$$\text{Hence we have: } \left(-\frac{1}{n} \frac{\partial^2 \phi_n}{\partial \beta^2}\right)_{\tilde{\beta}} \sqrt{n} (\hat{\beta}_n - \beta_0) = \sqrt{n} \left(\frac{1}{n} \frac{\partial \phi_n}{\partial \beta}\right)_{\beta_0}$$

As in the M.L.E. theory, we shall prove in a first step

that $\sqrt{n} \left(\frac{1}{n} \frac{\partial \phi_n}{\partial \beta}\right)_{\beta_0}$ is asymptotically distributed as

a multivariate normal random variable. In a second

step, we shall prove that $\left(-\frac{1}{n} \frac{\partial^2 \phi_n}{\partial \beta^2}\right)_{\tilde{\beta}}$ has the same limit

as $\left(-\frac{1}{n} \frac{\partial^2 \phi_n}{\partial \beta^2}\right)_{\beta_0}$ and that this term converges to a finite limit. Unfortunately these terms are not averages of i.i.d random variables.

A First Step: $C_n = \sqrt{n} \left[\frac{1}{n} \frac{\partial \phi_n}{\partial \beta}\right]_{\beta_0}$ is asymptotically

distributed as a multivariate normal random variable:

Let us introduce some notation: (following Tsiatis)

$$\hat{\epsilon}_z(t) = \frac{1}{n} \sum_{j=1}^n 1(y_j \geq t) \frac{\partial h}{\partial \beta}(\beta_0, z_j) ; \epsilon_z(t) = E\left[1(y \geq t) \frac{\partial h}{\partial \beta}\right]$$

where Y_j is the time of disappearance of subject having

covariate z_j .

$$\hat{\epsilon}(t) = \frac{1}{n} \sum_{j=1}^n 1(y_j \geq t) h(\beta_0, z_j) \quad , \quad \epsilon(t) = E[1(y \geq t) h(\beta_0, z)]$$

$$\hat{Q}(t) = \frac{1}{n} \sum_{i=1}^n \delta_i 1(y_i \geq t) \quad , \quad Q(t) = E[\delta 1(y \geq t)]$$

$Q(t)$ is the probability of surviving until time t without being censored and eventually dying before being censored so $Q(0) < 1$. In the chapter on consistency, it was proved that

$$E_z \left[\frac{\delta}{h(\beta_0, z)} \frac{\partial h(\beta_0, z)}{\partial \beta} \right] = E_{t,z} \left[\frac{\delta \epsilon_2(t)}{\epsilon(t)} \right]$$

Therefore, C_n can be written as:

$$\sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n \delta_i \frac{\partial h(\beta_0, z_i)}{\partial \beta} - E_z \left[\frac{\delta}{h(\beta_0, z)} \frac{\partial h(\beta_0, z)}{\partial \beta} \right] + E_{t,z} \left[\frac{\delta \epsilon_2(t)}{\epsilon(t)} \right] - \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\hat{\epsilon}_2(t_i)}{\hat{\epsilon}(t_i)} \right]$$

Let us first consider:

$$F_n = \sqrt{n} \left[E_{t,z} \left[\frac{\delta \epsilon_2(t)}{\epsilon(t)} \right] - \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\hat{\epsilon}_2(t_i)}{\hat{\epsilon}(t_i)} \right]$$

We have:

$$E_{t,z} \left[\frac{\delta \epsilon_2(t)}{\epsilon(t)} \right] = \int_{[0, T_F]} (-dQ) \frac{\epsilon_2(t)}{\epsilon(t)}$$

Similarly

$$\frac{1}{n} \sum_{i=1}^n \delta_i \frac{\hat{\epsilon}_2(t_i)}{\hat{\epsilon}(t_i)} = \int_{[0, T_F]} (-d\hat{Q}) \frac{\hat{\epsilon}_2(t)}{\hat{\epsilon}(t)}$$

Hence we are thinking of Q as the cumulative distribution function of a signed finite measure.

$$\text{Therefore } F_n = \sqrt{n} \left[\int_{[0, T_F]} (-dQ) \frac{\epsilon_2(t)}{\epsilon(t)} - \int_{[0, T_F]} (-d\hat{Q}) \frac{\hat{\epsilon}_2(t)}{\hat{\epsilon}(t)} \right]$$

Then C_n can be expressed as:

$$C_n = \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \left[\frac{\delta_i}{h(\beta_0, z_i)} \frac{\partial h(\beta_0, z_i)}{\partial \beta} - E \left[\frac{\delta}{h(\beta_0, z)} \frac{\partial h(\beta_0, z)}{\partial \beta} \right] \right] \right. \\ \left. + \int_{[0, T_F]} (-dQ) \frac{\varepsilon_2(t)}{\varepsilon(t)} - \int_{[0, T_F]} (-d\hat{Q}) \frac{\hat{\varepsilon}_2(t)}{\varepsilon(t)} \right\}$$

The term $\int_{[0, T_F]} (-dQ) \frac{\varepsilon_2(t)}{\varepsilon(t)} - \int_{[0, T_F]} (-d\hat{Q}) \frac{\hat{\varepsilon}_2(t)}{\varepsilon(t)}$ is similar

to an expression of the form $ab - \hat{a}\hat{b}$ where a, b are two theoretical functions and \hat{a}, \hat{b} the empirical corresponding functions. A classical approach is to write $ab - \hat{a}\hat{b} = a(b - \hat{b}) + b(a - \hat{a}) + (\hat{a} - a)(b - \hat{b})$. In our case $(a - \hat{a})(b - \hat{b})$ will be a quantity of order $\frac{1}{\sqrt{n}}$, therefore "negligible", and the terms $(a - \hat{a}), (b - \hat{b})$ will be asymptotically normal variables. More formally it will be shown in the appendix that C_n can be expressed as:

$C_n = C_{1n} + C_{2n} + C_{3n} + C_{4n} + C_{5n} + C_{6n} + R_{1n} + R_{2n} + R_{3n} + R_{4n}$ where

$$C_{1n} = \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{h(\beta_0, z_i)} \frac{\partial h(\beta_0, z_i)}{\partial \beta} - E \left[\frac{\delta}{h(\beta_0, z)} \frac{\partial h(\beta_0, z)}{\partial \beta} \right] \right]$$

$$C_{2n} = -\sqrt{n} \left[\hat{Q}(0) - Q(0) \right] \frac{\varepsilon_2(0)}{\varepsilon(0)}$$

$$C_{3n} = -\sqrt{n} \int_{[0, T_F]} \frac{(\hat{Q} - Q)}{\varepsilon} d\varepsilon_2$$

$$C_{4n} = \int_{[0, T_F]} \sqrt{n} \frac{(\hat{Q} - Q) \varepsilon_2}{\varepsilon^2} d\varepsilon \quad ; \quad C_{5n} = \int_{[0, T_F]} \sqrt{n} \frac{(\hat{\varepsilon}_2 - \varepsilon_2)}{\varepsilon} dQ$$

$$C_{6n} = - \int_{[0, T_F]} \sqrt{n} \frac{(\hat{\varepsilon} - \varepsilon)}{\varepsilon^2} \varepsilon_2 dQ$$

$$R_{1n} = \int_{[0, T_F]} \sqrt{n} \frac{(\hat{\varepsilon}_2 - \varepsilon_2)}{\hat{\varepsilon}} d(\hat{Q} - Q) \quad , \quad R_{2n} = - \int_{[0, T_F]} \sqrt{n} \frac{(\hat{\varepsilon}_2 - \varepsilon_2)(\hat{\varepsilon} - \varepsilon)}{\hat{\varepsilon} \varepsilon} dQ$$

$$R_{3n} = - \int_{[0, T_F]} \sqrt{n} \frac{(\hat{\varepsilon} - \varepsilon) \varepsilon_2}{\hat{\varepsilon} \varepsilon} d(\hat{Q} - Q) \quad ; \quad R_{4n} = \int_{[0, T_F]} \sqrt{n} \frac{(\hat{\varepsilon} - \varepsilon)^2 \varepsilon_2}{\hat{\varepsilon} \varepsilon^2} dQ$$

We shall prove that C_n is asymptotically distributed as a multivariate normal random variable in two steps:

First step: $D_n = C_{1n} + \dots + C_{6n}$ is asymptotically distributed as a multivariate normal random variable with covariance matrix Σ and mean 0.

2nd step: $R_{1n}, R_{2n}, R_{3n}, R_{4n}$ go to 0 in probability.

First step: Let us consider the sum S of the random terms in D_n .

$$S = \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{h(\beta_0, z_i)} \frac{\partial h(\beta_0, z_i)}{\partial \beta} - \hat{Q}(0) \frac{\varepsilon_2(0)}{\varepsilon(0)} + \int_{[0, T_F]} \frac{\hat{Q}}{\varepsilon^2} \varepsilon_2 d\varepsilon \right. \\ \left. - \int_{[0, T_F]} \frac{\hat{Q}}{\varepsilon} d\varepsilon_2 + \int_{[0, T_F]} \frac{\hat{\varepsilon}_2}{\varepsilon} dQ - \int_{[0, T_F]} \frac{\hat{\varepsilon}}{\varepsilon^2} \varepsilon_2 dQ \right\}$$

Rewriting this we have:

$$S = \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{h(\beta_0, z_i)} \frac{\partial h(\beta_0, z_i)}{\partial \beta_0} - \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\varepsilon_2(0)}{\varepsilon(0)} - \frac{1}{n} \sum_{i=1}^n \delta_i \int_{[0, T_F]} \frac{1(y_i \geq t)}{\varepsilon} d\varepsilon_2 \right. \\ \left. + \frac{1}{n} \sum_{i=1}^n \delta_i \int_{[0, T_F]} \frac{\varepsilon_2}{\varepsilon^2} 1(y_i \geq t) d\varepsilon + \frac{1}{n} \sum_{i=1}^n \frac{\partial h}{\partial \beta}(\beta_0, z_i) \int_{[0, T_F]} \frac{1(y_i \geq t)}{\varepsilon(t)} dQ \right. \\ \left. - \frac{1}{n} \sum_{i=1}^n h(\beta_0, z_i) \int_{[0, T_F]} \frac{\varepsilon_2}{\varepsilon^2} 1(y_i \geq t) d\varepsilon \right\}$$

Therefore S is a sum of independent, identically distributed random vectors.

Then S is asymptotically distributed as a M.V.N $(0, \frac{1}{n})$

2nd step: $R_{1n}, R_{2n}, R_{3n}, R_{4n}$, go to 0 in probability:

In the following we shall consider only the k th coordinate of R_{1n} , ($k = 1, \dots, p$).

1) Let us consider R_{2n} :

$$R_{2n} = - \int_{[0, T_F]} \frac{\sqrt{n} (\hat{\varepsilon}_2 - \varepsilon_2) (\hat{\varepsilon} - \varepsilon)}{\hat{\varepsilon} \varepsilon} dQ$$

When N goes to $+\infty$, $\hat{\varepsilon}(T_F)$ converges to $\varepsilon(T_F)$. We assumed

that $\varepsilon(T_F) > 0$. Let ε_1 be given such that $0 < \varepsilon_1 < \varepsilon(T_F)$

Then there exists N_1 such that for any $N > N_1$, $\hat{\varepsilon}(T_F) \geq \varepsilon(T_F) - \varepsilon_1$

$$|R_{2n}| \leq \frac{\sup_t n^{1/4} (\hat{\varepsilon}_2(t) - \varepsilon_2(t)) \cdot \sup_t n^{1/4} (\hat{\varepsilon}(t) - \varepsilon(t))}{\varepsilon(T_F) (\varepsilon(T_F) - \varepsilon_i)}$$

Since for any t such that $t \leq T_F$ we have

$$\varepsilon(t) \geq \varepsilon(T_F) \quad ; \quad \hat{\varepsilon}(t) \geq \hat{\varepsilon}(T_F).$$

Lemma 5 proved that

$$\sup_t n^{1/4} (\hat{\varepsilon}_2(t) - \varepsilon_2(t)) \quad , \quad \sup_t n^{1/4} (\hat{\varepsilon}(t) - \varepsilon(t))$$

go to 0 in probability.

Therefore

R_{2n} goes to 0 in probability

2) A similar argument shows that R_{4n} goes to 0 in probability.

3) Consider $R_{1n} = \int_{[0, T_F]} \frac{\sqrt{n} [\hat{\varepsilon}_2 - \varepsilon_2]}{\hat{\varepsilon}} d(\hat{Q} - Q)$

As in lemma 5, let $Z_n(t) = \sqrt{n} [\hat{\varepsilon}_2(t) - \varepsilon_2(t)]$ and let $Z(t)$ be the Gaussian process limit of $Z_n(t)$.

Therefore

$$R_{1n} = \int_{[0, T_F]} \frac{Z_n(t)}{\hat{\varepsilon}(t)} d(\hat{Q} - Q)$$

$$R_{1n} = \int_{[0, T_F]} \frac{Z_n(t) - Z(t)}{\hat{\varepsilon}} d(\hat{Q} - Q) + \int_{[0, T_F]} Z \left(\frac{1}{\hat{\varepsilon}} - \frac{1}{\varepsilon} \right) d(\hat{Q} - Q) + \int_{[0, T_F]} \frac{Z}{\varepsilon} d(\hat{Q} - Q)$$

Since we have proved the weak convergence of $Z_n(t)$ to $Z(t)$ for β fixed, it is possible (Skorokhod (1956)) to construct processes $X_n(t, \omega, \beta_0)$ and $X(t, \omega, \beta_0)$, possibly on a new probability space, such that the finite dimensional distributions of $X_n(t, \omega, \beta_0)$ and $Z_n(t, \omega, \beta_0)$, of $Z(t, \omega, \beta_0)$ and $X(t, \omega, \beta_0)$ are the same, such that $X_n(t, \omega, \beta_0)$ and $X(t, \omega, \beta_0)$ belong to \mathcal{D} for every ω in a subspace $\Omega_0(\beta_0)$ depending on β_0 such that $P[\Omega_0(\beta_0)] = 1$ and such that $X_n(t, \omega, \beta_0)$ converges to $X(t, \omega, \beta_0)$ for every ω in $\Omega_0(\beta_0)$.

Let

$$A = \int_{[0, T_F]} \frac{Z_n(t) - Z(t)}{\hat{\varepsilon}} d(\hat{Q} - Q)$$

$$B = \int_{[0, T_F]} Z(t) \frac{\varepsilon - \hat{\varepsilon}}{\hat{\varepsilon} \varepsilon} d(\hat{Q} - Q)$$

$$C = \int_{[0, T_F]} \frac{Z}{\varepsilon} d(\hat{Q} - Q)$$

Therefore $Z_n(t, \omega)$ converges almost surely to $Z(t, \omega)$ with the Skorokhod topology in D , (after the Skorokhod construction). Billingsley (1968) shows that the convergence to a continuous limit, as $Z(t)$ is, in the Skorokhod topology is equivalent to uniform convergence. Therefore if $\rho_T(Z_n, Z) = \sup_t (Z_n, Z)$, $\rho_T(Z_n, Z)$ converges to 0. As before there exists N_1 such that, for any $N > N_1$, $\hat{\varepsilon}(T_F) \geq \varepsilon(T_F) - \varepsilon_1$.

First term:

$$A = \int_{[0, T_F]} \frac{Z_n(t) - Z(t)}{\hat{\varepsilon}} d(\hat{Q} - Q)$$

$$|A| \leq \frac{\rho_T(Z_n, Z)}{(\varepsilon(T_F) - \varepsilon_1)} \left[\left| \int_{[0, T_F]} d\hat{Q} \right| + \left| \int_{[0, T_F]} dQ \right| \right]$$

$$\left| \int_{[0, T_F]} d\hat{Q} \right| + \left| \int_{[0, T_F]} dQ \right| \leq 2$$

Therefore $|A|$ converges to 0

$$\text{Second term: } B = \int_{[0, T_F]} Z(t) \frac{\varepsilon - \hat{\varepsilon}}{\hat{\varepsilon} \varepsilon} d(\hat{Q} - Q)$$

By the same arguments as before:

$$|B| \leq \frac{\rho_T(Z, 0) \rho_T(\varepsilon, \hat{\varepsilon})}{\varepsilon(T_F) (\varepsilon(T_F) - \varepsilon_1)} \left| \int_{[0, T_F]} d(\hat{Q} - Q) \right|$$

But as before

$$\left| \int_{[0, T_f]} d(\hat{Q} - Q) \right| \leq 2$$

Lemma 5 yields that $\rho_T(\varepsilon, \hat{\varepsilon})$ converges to 0 in probability.

Therefore B converges to 0 in probability.

$$\text{Third term: } C = \int_{[0, T_f]} \frac{Z}{\varepsilon} d(\hat{Q} - Q)$$

Consider a subset Ω_0 of the underlying probability space such that:

$$i) P[\Omega_0] = 1$$

ii) for $\omega \in \Omega_0$, Z is uniformly continuous in $(0, T_f)$

Choose a partition (depending on ω) of $[0, T_f]$ into K_0 intervals $I_k =]\xi_{k-1}, \xi_k[$ such that

$$\sup_{t \in I_k} \left| \frac{Z(t)}{\varepsilon(t)} - \frac{Z(\xi_k)}{\varepsilon(\xi_k)} \right| < \varepsilon \quad (k=1, \dots, K_0)$$

K_0 is fixed given ε .

Then:

$$C = \int_{[0, T_f]} \frac{Z}{\varepsilon} d(\hat{Q} - Q) = \sum_{k=0}^{K_0-1} \int_{\xi_k}^{\xi_{k+1}} \frac{Z(t)}{\varepsilon(t)} d(\hat{Q} - Q)$$

But we have the inequalities:

$$\sum_{k=0}^{k_0-1} \frac{Z(\xi_{k+1})}{\varepsilon(\xi_{k+1})} \int_{\xi_k}^{\xi_{k+1}} d(\hat{Q}-Q) - \sum_{k=0}^{k_0-1} \int_{\xi_k}^{\xi_{k+1}} \sup_{t \in I_{k+1}} \left(\frac{Z}{\varepsilon} - \frac{Z(\xi_{k+1})}{\varepsilon(\xi_{k+1})} \right) d(\hat{Q}-Q) \leq A$$

and

$$C \leq \sum_{k=0}^{k_0-1} \frac{Z(\xi_{k+1})}{\varepsilon(\xi_{k+1})} \left[\hat{Q}(\xi_{k+1}) - Q(\xi_{k+1}) - (\hat{Q}(\xi_k) - Q(\xi_k)) \right] \\ + \sum_{k=0}^{k_0-1} \int_{\xi_k}^{\xi_{k+1}} \sup_{t \in I_{k+1}} \left(\frac{Z}{\varepsilon} - \frac{Z(\xi_{k+1})}{\varepsilon(\xi_{k+1})} \right) d(\hat{Q}-Q)$$

The second term is bounded by 2ε .

The first term can be rewritten as:

$$\frac{Z(\xi_k)}{\varepsilon(\xi_k)} \left[\hat{Q}(\xi_k) - Q(\xi_k) \right] + \sum_{k=1}^k \left[\frac{Z(\xi_k)}{\varepsilon(\xi_k)} - \frac{Z(\xi_{k-1})}{\varepsilon(\xi_{k-1})} \right] \left[\hat{Q}(\xi_{k-1}) - Q(\xi_{k-1}) \right] \\ + \frac{Z(\xi_1)}{\varepsilon(\xi_1)} \left[\hat{Q}(0) - Q(0) \right]$$

Therefore the first term is bounded by

$$(k_0-1)\varepsilon \rho_T(\hat{Q}, Q) + 2 \rho_T\left(\frac{Z}{\varepsilon}, 0\right) \rho_T(\hat{Q}, Q)$$

Hence $|C|$ is bounded by:

$$2\varepsilon + \left[(k_0-1)\varepsilon + 2 \rho_T\left(\frac{Z}{\varepsilon}, 0\right) \right] \rho_T(\hat{Q}, Q)$$

Lemma 5 implies that $f_T(\hat{Q}, Q)$ converges to 0 when N goes to $+\infty$. As ϵ was arbitrary, this implies C converges to 0 in probability. Therefore:

Rln goes to 0 in probability

- 4) R3n goes to 0 in probability by the same arguments as before.

We therefore have proved that:

$C_n = \sqrt{n} \left[\frac{1}{n} \frac{\partial \phi_n}{\partial \beta} \right]_{\beta_0}$ is asymptotically distributed as a multivariate normal random variable.

B Second Step: Consider $\hat{M} = -\frac{1}{n} \frac{\partial^2 \phi_n}{\partial \beta^2}$.

$\hat{\beta}$ lies between $\hat{\beta}_n$ and β_0 . We shall prove that M converges to a finite limit in two steps:

First step: $\hat{M}_0 = -\frac{1}{n} \frac{\partial^2 \phi_n}{\partial \beta_0^2}$ converges in probability to a finite limit M_0 .

Second step:

The sequence $\left\{ -\frac{1}{N} \frac{\partial^2 \phi_N}{\partial \beta^2} \right\}$

is equicontinuous in a neighbourhood of β_0 .

It will therefore imply that \hat{M} converges to M_0 .

First Step: convergence of \hat{M}_0

\hat{M}_0 is equal to:

$$\hat{M}_0 = \frac{1}{N} \sum_{i=1}^N \delta_i \left\{ \frac{\frac{\partial h}{\partial \beta}(\beta_0, z_i)^2}{h(\beta_0, z_i)^2} - \frac{\frac{\partial^2 h}{\partial \beta^2}(\beta_0, z_i)}{h(\beta_0, z_i)} \right\}$$

$$+ \frac{1}{N} \sum_{i=1}^N \delta_i \left\{ \frac{\frac{1}{n} \sum_{j \in R_i} \frac{\partial^2 h}{\partial \beta^2}(\beta_0, z_j)}{\frac{1}{n} \sum_{j \in R_i} h(\beta_0, z_j)} - \frac{\left[\frac{1}{n} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta_0, z_j) \right]^2}{\left[\frac{1}{n} \sum_{j \in R_i} h(\beta_0, z_j) \right]^2} \right\}$$

An application of strong law of large numbers gives that the first term converges almost surely to

$$E \left[\delta \left\{ \frac{\frac{\partial h}{\partial \beta}(\beta_0, z)^2}{h(\beta_0, z)^2} - \frac{\frac{\partial^2 h}{\partial \beta^2}(\beta_0, z)}{h(\beta_0, z)} \right\} \right]$$

Now let us consider the second term.

We claim that

$$\hat{A} = \frac{1}{N} \sum_{i=1}^N \delta_i \frac{\frac{1}{n} \sum_{j=1}^n I(y_j \geq T_i) \frac{\partial h}{\partial \beta}(\beta_0, z_j)}{\frac{1}{n} \sum_{j=1}^n I(y_j \geq T_i) h(\beta_0, z_j)} \quad \text{converges to:}$$

$$M_n = E \left[\delta \frac{E \left[I(y \geq t) \frac{\partial h}{\partial \beta^2}(\beta_0, z) \right]}{E \left[I(y \geq t) h(\beta_0, z) \right]} \right] \quad \text{in probability.}$$

Let us introduce some further notation:

$$\hat{\varepsilon}_2(t) = \frac{1}{n} \sum_{j=1}^n I(y_j \geq t) \frac{\partial^2 h}{\partial \beta^2}(\beta_0, z_j); \quad \varepsilon_2(t) = E \left[I(y \geq t) \frac{\partial^2 h}{\partial \beta^2}(\beta_0, z) \right]$$

$$A = \frac{1}{N} \sum_{i=1}^N \int_{\tau_i} \frac{\varepsilon_2(\tau_i)}{\varepsilon(\tau_i)}$$

Therefore

$$\hat{A} - A = \frac{1}{N} \sum_{i=1}^N \int_{\tau_i} \left[\frac{\hat{\varepsilon}_2(\tau_i)}{\hat{\varepsilon}(\tau_i)} - \frac{\varepsilon_2(\tau_i)}{\varepsilon(\tau_i)} \right]$$

Let us first prove the following lemma:

Lemma: $\alpha = \sup \left\{ \left| \frac{\hat{\varepsilon}_2(t)}{\hat{\varepsilon}(t)} - \frac{\varepsilon_2(t)}{\varepsilon(t)} \right| \right\}$ converges in probability to 0.

We have
$$\frac{\hat{\varepsilon}_2(t)}{\hat{\varepsilon}(t)} - \frac{\varepsilon_2(t)}{\varepsilon(t)} = \frac{\hat{\varepsilon}_2(t) - \varepsilon_2(t)}{\hat{\varepsilon}(t)} + \varepsilon_2(t) \frac{\varepsilon(t) - \hat{\varepsilon}(t)}{\hat{\varepsilon}(t) \varepsilon(t)}$$

As we have seen in the preceding pages, for n large enough, $\hat{\varepsilon}(t) \geq \varepsilon(\tau_F) - \varepsilon_1$

Therefore

$$\alpha \leq \frac{1}{(\varepsilon(\tau_F) - \varepsilon_1)} \sup_t |\hat{\varepsilon}_2(t) - \varepsilon_2(t)| + \frac{\sup_t \varepsilon_2(t)}{\varepsilon(\tau_F)(\varepsilon(\tau_F) - \varepsilon_1)} \sup_t |\hat{\varepsilon}(t) - \varepsilon(t)|$$

Lemma 5 yields that $\sup_t |\hat{\varepsilon}_2(t) - \varepsilon_2(t)|$ and

$\sup_t |\hat{\varepsilon}(t) - \varepsilon(t)|$ converges in probability to 0.

Therefore a converges in probability to 0.

We have then $\|\hat{A} - A\| < \alpha \left(\frac{1}{n} \sum_{i=1}^n \delta_i \right)$

The strong law of large numbers yields that $\left(\frac{1}{n} \sum_{i=1}^n \delta_i \right)$ converges almost surely to $E(\delta)$.

Therefore \hat{A} converges to A in probability.

Another application of the strong law of large numbers gives us that A converges to M .

Now let us consider $\hat{B} = \frac{1}{N} \sum_{i=1}^N \delta_i \frac{\hat{\varepsilon}_2(\tau_i)^2}{\hat{\varepsilon}(\tau_i)^2}$

Let us prove that has the same limit as $B = \frac{1}{N} \sum_{i=1}^N \delta_i \frac{\varepsilon_2(\tau_i)^2}{\varepsilon(\tau_i)^2}$

$$\hat{B} - B = \frac{1}{N} \sum_{i=1}^N \delta_i \left[\frac{\hat{\varepsilon}_2(\tau_i)}{\hat{\varepsilon}(\tau_i)} - \frac{\varepsilon_2(\tau_i)}{\varepsilon(\tau_i)} \right] \left[\frac{\hat{\varepsilon}_2(\tau_i)}{\hat{\varepsilon}(\tau_i)} + \frac{\varepsilon_2(\tau_i)}{\varepsilon(\tau_i)} \right]$$

Therefore

$$\|\hat{B} - B\| < \sup_t \left| \frac{\hat{\varepsilon}_2(t)}{\hat{\varepsilon}(t)} - \frac{\varepsilon_2(t)}{\varepsilon(t)} \right| \cdot \frac{1}{N} \sum_{i=1}^N \delta_i \left\| \frac{\hat{\varepsilon}_2(\tau_i)}{\hat{\varepsilon}(\tau_i)} + \frac{\varepsilon_2(\tau_i)}{\varepsilon(\tau_i)} \right\|$$

By arguments similar to those of the preceding paragraph,

it can be proved that $\sup_t \left| \frac{\hat{\epsilon}_2(t)}{\hat{\epsilon}(t)} - \frac{\epsilon_2(t)}{\epsilon(t)} \right|$ converges to

0 and $\frac{1}{N} \sum_{i=1}^N \delta_i \left\| \frac{\hat{\epsilon}_2(T_i)}{\hat{\epsilon}(T_i)} + \frac{\epsilon_2(T_i)}{\epsilon(T_i)} \right\|$ converges to a finite

limit. Therefore \hat{B} and B have the same limit. The strong law of large numbers yields that B converges almost surely

to $E \left[\delta \frac{\epsilon_2(t)^2}{\epsilon(t)^2} \right]$

Therefore M converges in probability to:

$$M_0 = E \left[\delta \left\{ \frac{\frac{\partial h(\beta_0, z)}{\partial \beta}(\beta_0, z)^2}{h(\beta_0, z)^2} - \frac{\frac{\partial^2 h(\beta_0, z)}{\partial \beta^2}(\beta_0, z)}{h(\beta_0, z)} \right\} \right] + E \left[\frac{\delta E \left[1(y \geq t) \frac{\partial^2 h(\beta_0, z)}{\partial \beta^2}(\beta_0, z) \right]}{E \left[1(y \geq t) h(\beta_0, z) \right]} \right]$$

$$- E \left[\frac{\delta E \left[1(y \geq t) \frac{\partial h(\beta_0, z)}{\partial \beta}(\beta_0, z) \right]^2}{E \left[1(y \geq t) h(\beta_0, z) \right]^2} \right]$$

2nd step: $\left\{ -\frac{1}{N} \frac{\partial^2 \phi_n}{\partial \beta^2} \right\}$ are equicontinuous functions

of β in a neighbourhood of β_0 . In order to demonstrate this we prove that $\frac{\partial^2 \phi_n}{\partial \beta^2}$ is bounded in a neighbourhood of β_0 .

First let us compute $A = \frac{\partial^3 \Phi_n}{\partial \beta^3}$

$$A = \frac{1}{n} \sum_{i=1}^n \delta_i \left\{ \frac{\frac{1}{n} \sum_{j \in R_i} \frac{\partial^3 h}{\partial \beta^3}(\beta, z_j)}{\frac{1}{n} \sum_{j \in R_i} h(\beta, z_j)} - 3 \frac{\left[\frac{1}{n} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j) \right] \left[\frac{1}{n} \sum_{j \in R_i} \frac{\partial^2 h}{\partial \beta^2}(\beta, z_j) \right]}{\left[\frac{1}{n} \sum_{j \in R_i} h(\beta, z_j) \right]^2} + 2 \frac{\left[\frac{1}{n} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j) \right]^3}{\left[\frac{1}{n} \sum_{j \in R_i} h(\beta, z_j) \right]^3} \right\}$$

In the chapter on consistency we prove there was a neighbourhood of β_0 in which, for n large enough, $\frac{1}{n} \sum_{j \in R_i} h(\beta, z_j)$

was bounded for each risk set R_i by a constant m .

In this chapter on Asymptotic Normality, we assumed

that $\frac{\partial h}{\partial \beta}(\beta, z)$, $\frac{\partial^2 h}{\partial \beta^2}(\beta, z)$, $\frac{\partial^3 h}{\partial \beta_1 \partial \beta_2 \partial \beta_3}$

were monotone functions in β for z fixed for any l .

We therefore can prove by arguments similar to those

we used to prove that $\{\Phi_n(\beta)\}$ was an equicontinuous sequence for n large enough, that

$$\left\| \frac{1}{n} \sum_{j \in R_i} \frac{\partial h}{\partial \beta}(\beta, z_j) \right\|, \left\| \frac{1}{n} \sum_{j \in R_i} \frac{\partial^2 h}{\partial \beta_1 \partial \beta_2}(\beta, z_j) \right\|, \left\| \frac{1}{n} \sum_{j \in R_i} \frac{\partial^3 h}{\partial \beta_1 \partial \beta_2 \partial \beta_3}(\beta, z_j) \right\|$$

are

bounded in a neighbourhood of β_0 . This then implies that

$\left\| \frac{\partial^3 \Phi_n}{\partial \beta^3} \right\|$ is bounded in a neighbourhood of β_0 .

Therefore the sequence $\left\{ -\frac{1}{N} \frac{\partial^2 \phi_n}{\partial \beta^2} \right\}$ is an equicontinuous sequence.

This proves that the matrix $\hat{M} = -\frac{1}{N} \frac{\partial^2 \phi_n}{\partial \beta^2}$ converges to:

$$M_0 = E \left[\delta \left(\frac{\frac{\partial h(\beta_0, z)}{\partial \beta} (\beta_0, z)^2 - h(\beta_0, z) \frac{\partial^2 h(\beta_0, z)}{\partial \beta^2}}{h(\beta_0, z)^2} \right) \right]$$

$$+ E \left[\delta \left(\frac{E[1(y \geq t) h(\beta_0, z)] E[1(y \geq t) \frac{\partial^2 h(\beta_0, z)}{\partial \beta^2}] - E[1(y \geq t) \frac{\partial h(\beta_0, z)}{\partial \beta}]^2}{E[1(y \geq t) h(\beta_0, z)]^2} \right) \right]$$

C Covariance structure:

The variance-covariance matrix of $\frac{1}{\sqrt{n}} \left(\frac{\partial \phi_n}{\partial \beta} \right)_{\beta_0}$ is the

variance-covariance matrix \neq of the random vector x

where:

$$x = \frac{\delta}{h(\beta_0, z)} \frac{\partial h(\beta_0, z)}{\partial \beta} - \frac{\delta \varepsilon_z(0)}{\varepsilon(0)} - \delta \int_{[0, T_F]} 1(y \geq t) \frac{d\varepsilon_z}{\varepsilon}$$

$$+ \delta \int_{[0, T_F]} \frac{\varepsilon_z}{\varepsilon^2} 1(y \geq t) d\varepsilon + \frac{\partial h(\beta_0, z)}{\partial \beta} \int_{[0, T_F]} \frac{1(y \geq t)}{\varepsilon(t)} dQ$$

$$- h(\beta_0, z) \int_{[0, T_F]} \frac{\varepsilon_z}{\varepsilon^2} 1(y \geq t) d\varepsilon$$

But:

$$\int_{[0, T_F]} \frac{\varepsilon_z}{\varepsilon^2} 1(y \geq t) d\varepsilon - \int_{[0, T_F]} 1(y \geq t) \frac{d\varepsilon_z}{\varepsilon} = - \int_{[0, T_F]} 1(y \geq t) \frac{d\varepsilon_z}{\varepsilon}$$

$$= \frac{\varepsilon_z(0)}{\varepsilon(0)} - \frac{\varepsilon_z(y)}{\varepsilon(y)}$$

Therefore

$$X = \frac{\delta}{h(\beta_0, z)} \frac{\partial h}{\partial \beta}(\beta_0, z) + \frac{\delta \varepsilon_z(y)}{\varepsilon(y)} - h(\beta_0, z) \int_{[0, T_F]} \frac{\varepsilon_z}{\varepsilon^2} 1(y \geq t) d\varepsilon$$

$$+ \frac{\partial h}{\partial \beta}(\beta_0, z) \int_{[0, T_F]} \frac{1(y \geq t)}{\varepsilon(t)} dQ$$

Provided that M is regular, the variance-covariance matrix V of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is then:

$$V = (M_0^{-1})' \neq (M_0^{-1})$$

A consistent estimator of M is given by $\left(-\frac{1}{n} \frac{\partial^2 \Phi_n}{\partial \beta^2} \right) \Big|_{\hat{\beta}_n}$.

In the exponential case where z has only one dimension,

Tsiatis (1978) showed that $\text{Var } X = M_0$.

$$M_0 = \int_{[0, T_F]} (-dQ) \text{Var}(z | Q(t)) = \int_{[0, T_F]} (-dQ) \frac{E[z^2 e^{\beta z} 1(y \geq t)]}{E[e^{\beta z} 1(y \geq t)]} - \frac{E[z e^{\beta z} 1(y \geq t)]^2}{E[e^{\beta z} 1(y \geq t)]^2}$$

In this case $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution to a normal random variable with mean 0 and variance

equal to $\left[\int_{[0, T_F]} (-dQ) \text{Var}(z|Q(t)) \right]^{-1}$. As a consistent

estimator of that variance is available, the construction of confidence intervals is straightforward. Unfortunately in our multiparameter general case some further work is needed to find a consistent estimator of $\frac{1}{2}$.

Chapter 4: Extension and Problems:

In the first part of the chapter, we shall outline the proof of the asymptotic normality when the covariates are given constants. In the second part a problem encountered during the simulation studies will be discussed.

Part A: Consistency of Cox's estimate when the covariates are given constants.

In this case our arguments will be conditional on the covariates. z_1, z_2, \dots, z_n are given constants. Let us introduce the notation $E_i(f(t, z)) = E(f(t, z) | z_i)$. For reasons of simplicity, we shall only consider the exponential case. Hence the Cox likelihood can be written:

$$\phi_N(\beta) = \frac{1}{N} \sum_{i=1}^N \delta_i \beta' z_i - \frac{1}{N} \sum_{i=1}^N \delta_i \log \left[\frac{1}{N} \sum_{j=1}^N I(y_j \geq \tau_i) e^{\beta' z_j} \right]$$

Recall Kolmogorov's proposition: if the r.v.'s X_n are independent, then

$$\frac{1}{n^2} \sum \text{Var } X_n < +\infty$$

entails $\frac{1}{N} \left(\sum X_n - \sum E X_n \right) \xrightarrow{\text{a.s.}} 0$

Hence if $f_N(\tau) = \frac{1}{N} \sum_{j=1}^N I(y_j \geq \tau) e^{\beta' z_j}$, $f_N(\tau)$ converges

almost surely to $\tilde{f}_N(\tau) = \frac{1}{N} \sum_{j=1}^N E_j \left[I(y \geq \tau) e^{\beta' z} \right]$

where the index j means expectation conditional on z_j .

By a method similar to the one used in chapter 2, we hope to prove that we can replace $f_n(T_i)$ by $\tilde{f}_n(T_i)$ in $\Phi_N(\beta)$ under some assumptions. More precisely if

$$\Phi_N^*(\beta) = \frac{1}{n} \sum_{i=1}^n E_i [\delta \beta' z] - \frac{1}{n} \sum_{i=1}^n \delta_i \text{Log} \left[\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq T_i) e^{\beta z}] \right]$$

$\Phi_N^*(\beta) - \Phi_N(\beta)$ converges a.s. to 0. The

random variables $X_i^n = \frac{1}{n} \sum_{j=1}^n E_j [I(y \geq T_i) e^{\beta z}]$ are

independent but depend on n . Hopefully if:

$$\tilde{\Phi}_N(\beta) = \frac{1}{n} \sum_{j=1}^n E_i [\delta \beta' z] - \frac{1}{n} \sum_{i=1}^n E_{i,t} \left[\delta \text{Log} \left[\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq T_i) e^{\beta z}] \right] \right]$$

$\tilde{\Phi}_N(\beta) - \Phi_N^*(\beta)$ will converge a.s. to 0. The notation used is the same as in the preceding chapters. Therefore $\tilde{\Phi}_N(\beta) - \Phi(\beta)$ will converge a.s. to 0.

A) The true value β_0 is a maximum of $\tilde{\Phi}_N(\beta)$:

This paragraph is an actual proof of $\left(\frac{\partial \tilde{\Phi}_N}{\partial \beta} \right)_{\beta_0} = 0$

when β is of dimension 1. This proof can easily be extended to the case where β is a vector.

$$\left(\frac{\partial \tilde{\Phi}_N}{\partial \beta} \right)_{\beta_0} = \frac{1}{n} \sum_{i=1}^n E_i [\delta z] - \frac{1}{n} \sum_{i=1}^n E_{i,t} \left[\delta \frac{\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq T_i) z e^{\beta_0 z}]}{\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq T_i) e^{\beta_0 z}]} \right]$$

For reasons similar to the ones in chapter 2:

$$E_{i,t} \left[\frac{\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq T_i) z e^{\beta_0 z}]}{\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq T_i) e^{\beta_0 z}]} \right] = \frac{1}{n} \sum_{j=1}^n \int_{t=0}^{t=+\infty} \lambda_0(t) E_j [I(y \geq t) z e^{\beta_0 z}] E_i [I(y \geq t) e^{\beta_0 z}] dt$$

$$E_i [S|z] = \int_{t=0}^{t=+\infty} \lambda_0(t) E_i [z e^{\beta_0 z} I(y \geq t)] dt$$

$$\text{Hence: } \left(\frac{\partial \tilde{\Phi}_n}{\partial \beta} \right)_{\beta_0} = \int_{t=0}^{t=+\infty} \lambda_0(t) \left(\frac{1}{n} \sum_{i=1}^n E_i [z e^{\beta_0 z} I(y \geq t)] \right) dt$$

$$- \int_{t=0}^{t=+\infty} \lambda_0(t) \frac{\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq t) z e^{\beta_0 z}] \left(\frac{1}{n} \sum_{i=1}^n E_i [I(y \geq t) e^{\beta_0 z}] \right)}{\left(\frac{1}{n} \sum_{j=1}^n E_j [I(y \geq t) e^{\beta_0 z}] \right)^2} dt$$

Therefore $\left(\frac{\partial \tilde{\Phi}_n}{\partial \beta} \right)_{\beta_0} = 0$. The true value is a local

maximum of $\tilde{\Phi}_n$.

B) Concavity of Cox's likelihood:

The same arguments as when the covariates are i.i.d gives that:

$$-\frac{\partial^2 \Phi_n}{\partial \beta^2} = \sum_{i=1}^n \delta_i \text{ (Variance-Covariance matrix of } z \text{ given } R_i)$$

Therefore Φ_n is concave everywhere

C) Consistency of Cox's estimate:

Our lemma 8 does not apply here as the function $\tilde{\Phi}_n(\beta)$ depends on n . Nevertheless it is not too hard to modify this lemma to prove that point wise convergence of $(\Phi_n - \tilde{\Phi}_n)$ to 0 implies uniform convergence on compacts for $\Phi_n, \tilde{\Phi}_n$ concave. If $\tilde{\Phi}_n$ is asymptotically strictly concave it will be possible to prove that $\hat{\beta}_n$ converges to the true value β_0 in probability.

Part B: Infinite Cox estimate of β

In this short chapter we would like to emphasize the dependance of the efficiency of the estimate on the variance of the covariates. A problem we encountered frequently in our simulation studies was that the estimate was infinite, especially when the variance of the covariate z was large.

For simplicity let us consider the exponential case without censoring and the parameter having dimension 1. Let $Z(i)$ be the covariate of the i th dying subject. The Cox likelihood is built up by multiplying terms of

the form
$$\frac{e^{\beta Z(i)}}{\sum_{k \in R_i} e^{\beta Z_k}}$$

The derivative of the log of this term is:

$$\frac{\sum_{k \in R(t_i)} (Z(i) - Z_k) e^{\beta Z_k}}{\sum_{k \in R(t_i)} e^{\beta Z_k}}$$

Therefore if we have the ordering

$$Z(1) > Z(2) \dots\dots\dots > Z(n)$$

Cox's likelihood is monotone and the estimate is infinite.

The greater the difference between $Z_i = Z(k)$ and $Z_j = Z(1)$, is,

the most probable it is that $k < 1$. Therefore if the

variance of z is small the probability of the troublesome

ordering happening is low, but if the variance is large

this probability becomes close to 1, (at least for $\beta \neq 0$).

Our simulation studies confirmed this heuristic argument.

This gives an interesting insight on the result of Kalbfleisch

(1974) mentioned on page 8. He shows that an approximate

expression of the efficiency of the estimate is

$$\exp\{-\beta^2 \text{Var}(z)\} \quad \text{valid near } \beta = 0. \quad \text{The efficiency}$$

depends on $\text{Var}(Z)$.

APPENDIXA 1: Proof of lemma 1

$$E[S|z] = E_t[E[S|z,t]] \quad \text{but} \quad E[S|z,t] = P[t \leq S|z,t]$$

and $P[t \leq S|z,t] = G(t|z)$.

Hence

$$\underline{E[S|z] = E_t[G(t|z)]}$$

A 2: Proof of lemma 2

$$f(w) = E[h(\beta, z) | (Y \geq w)] = E[h(\beta, z) E[1(\text{Min}(t, s) \geq w) | z]]$$

meanwhile

$$E[1(\text{Min}(t, s) \geq w) | z] = \text{PROB}[\text{Min}(t, s) \geq w | z]$$

t and s are independant given z. Therefore:

$$P[\text{Min}(t, s) \geq w | z] = P[t \geq w | z] P[s \geq w | z] = G(w|z) \left(\exp - \int_0^w \lambda_0(u) h(\beta, z) du \right)$$

Hence:

$$\underline{f(w) = E_z \left[h(\beta, z) G(w|z) \left(\exp - \int_0^w \lambda_0(u) h(\beta, z) du \right) \right]}$$

A 3: Proof of lemma 3

$$E[S \text{Log} f(t)] = E_{z,t,s} [S \text{Log} f(t)] = E_{z,t} [E_s [S \text{Log} f(t) | t, z]]$$

$$E[S \text{Log} f(t)] = E_{z,t} [\text{Log} f(t) E_s [S | t, z]]$$

meanwhile $E_s [S | t, z] = P[s \geq t | t, z] = G(t|z)$. Hence:

$$\underline{E[S \text{Log} f(t)] = E_{z,t} [\text{Log} f(t) G(t|z)]}$$

A 4: Proof of lemma 4

$$g(w) = E[S | (Y \geq w) | z] = P[T \leq S, Y \geq w | z]$$

As $Y = \min(T, S) = T$ when $T \leq S$, $g(w) = P[T \leq S, T \geq w | z]$.

Meanwhile

$$P[T \leq S, T \geq w | z] = E_t [P[t \leq S, t \geq w | z, t]]$$

Hence:

$$E[S | (Y \geq w) | z] = \int_{t=w}^{t=\infty} \lambda_0(t) h(\beta, z) G(t|z) \left(\exp - \int_0^t \lambda_0(u) h(\beta, z) du \right) dt$$

A 5: Proof of

$$M(t) = \int_{x=-\infty}^{x=+\infty} \int_{z=x}^{z=+\infty} \left[\frac{\frac{\partial h(\beta, z)}{\partial \beta}}{h(\beta, z)} - \frac{\frac{\partial h(\beta, x)}{\partial \beta}}{h(\beta, x)} \right]^2 d\nu(t, z, x)$$

We have $M(t) = A_1(t) - A_2(t)$ where

$$A_1(t) = E_z \left[h(\beta, z) \left(\exp - \int_0^t \lambda_0(u) h(\beta, z) du \right) G(t|z) \right] \cdot E_z \left[\frac{\frac{\partial h^2}{\partial \beta} G(t|z)}{h(\beta, z)} \exp - \int_0^t \lambda_0(u) h(\beta, z) du \right]$$

$$A_2(t) = E_z \left[\frac{\partial h}{\partial \beta} G(t|z) \exp - \int_0^t \lambda_0(u) h(\beta, z) du \right]^2$$

Recall $\kappa(z)$ is the density of z .

$$\text{Let } a(t, x) = G(t|x) \left(\exp - \int_0^t \lambda_0(u) h(\beta, x) du \right) \kappa(x)$$

Hence:

$$A_1(t) = \int_{x=-\infty}^{x=+\infty} \int_{z=-\infty}^{z=+\infty} \left\{ \frac{\frac{\partial h(\beta, z)}{\partial \beta}}{h(\beta, z)} \right\}^2 h(\beta, x) a(t, x) h(\beta, z) a(t, z) dx dz.$$

Let $dp(t, z, x) = a(t, x) h(\beta, x) a(t, z) h(\beta, z) dx dz$

$A_1(t)$ can be written as $A_1(t) = K_1(t) + K_2(t)$ where

$$K_1(t) = \int_{x=-\infty}^{x=+\infty} \int_{z=x}^{z=+\infty} \left\{ \frac{\frac{\partial h(\beta, z)}{\partial \beta}}{h(\beta, z)} \right\}^2 dp(t, z, x) \quad \text{and}$$

$$K_2(t) = \int_{x=-\infty}^{x=+\infty} \int_{z=-\infty}^{z=x} \left\{ \frac{\frac{\partial h}{\partial \beta}}{h(\beta, z)} \right\}^2 dp(t, z, x)$$

But $K_2(t)$ can be written as:

$$K_2(t) = \int_{z=-\infty}^{z=+\infty} \int_{x=z}^{x=+\infty} \left\{ \frac{\frac{\partial h}{\partial \beta}}{h(\beta, z)} \right\}^2 dp(t, z, x)$$

As $dp(t, z, x) = dp(t, x, z)$,

$$A_1(t) = \int_{x=-\infty}^{x=+\infty} \int_{z=x}^{z=+\infty} \left\{ \left[\frac{\frac{\partial h(\beta, z)}{\partial \beta}}{h(\beta, z)} \right]^2 + \left[\frac{\frac{\partial h(\beta, x)}{\partial \beta}}{h(\beta, x)} \right]^2 \right\} dp(t, z, x)$$

Similarly it can be proved that:

$$A_2(t) = \int_{x=-\infty}^{x=+\infty} \int_{z=x}^{z=+\infty} \left\{ \left[\frac{\frac{\partial h(\beta, z)}{\partial \beta}}{h(\beta, z)} \right] \left[\frac{\frac{\partial h(\beta, x)}{\partial \beta}}{h(\beta, x)} \right] + \left[\frac{\frac{\partial h(\beta, x)}{\partial \beta}}{h(\beta, x)} \right] \left[\frac{\frac{\partial h(\beta, z)}{\partial \beta}}{h(\beta, z)} \right] \right\} dp(t, z, x)$$

A 6: Proof of

$$E\left[\left(Z_n(t_1) - Z_n(t_2)\right)^2\right] \leq c_4 \left[F(t_2) - F(t_1)\right]^{4/3}$$

where

$$F(t) = P[Y < t] \quad \text{and} \quad 0 \leq t_1 \leq t \leq t_2.$$

Let $p_1 = F(t) - F(t_1)$ and $p_2 = F(t_2) - F(t)$, it yields $F(t_2) - F(t_1) = p_1 + p_2$

As F is monotone, $0 \leq p_1^{2/3} p_2^{2/3} \leq (p_1 + p_2)^{4/3}$

$$Z_n(t_1) - Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ h(\beta, z_j) \left[I(y_j \geq t_1) - I(y_j \geq t) \right] - E \left[\left\{ I(y \geq t_1) - I(y \geq t) \right\} h(\beta, z) \right] \right\}$$

But $I(y_j \geq t_1) - I(y_j \geq t) = I(y_j \in [t_1, t[)$.

Consider

$$U_{1i} = h(\beta, z_i) I(y_i \in [t_1, t[) - E \left[I(y \in [t_1, t[) h(\beta, z) \right]$$

$$U_{2i} = h(\beta, z_i) I(y_i \in [t, t_2[) - E \left[I(y \in [t, t_2[) h(\beta, z) \right]$$

Therefore $Z_n(t_1) - Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n U_{1j}$, $Z_n(t) - Z_n(t_2) = \frac{1}{\sqrt{n}} \sum_{j=1}^n U_{2j}$

It yields that:

$$E \left[\left| Z_n(t_1) - Z_n(t) \right|^2 \mid Z_n(t_1) - Z_n(t) \right] = \frac{1}{n^2} E \left[\left(\sum_{i=1}^n U_{1i} \right)^2 \left(\sum_{i=1}^n U_{2i} \right)^2 \right]$$

The random vectors (U_{1i}, U_{2i}) are independent and

$E[U_{ii}] = 0$. It can be proved that

$$E \left[\left(\sum_{i=1}^n U_{1i} \right)^2 \left(\sum_{i=1}^n U_{2i} \right)^2 \right] = n E[U_1^2 U_2^2] + n(n-1) E[U_1^2] E[U_2^2] + 2n(n-1) E[U_1 U_2]^2$$

let $I_1 = 1(y_i \in [t_1, t_2])$ and $I_2 = 1(y_i \in [t_2, t_3])$,

$$U_{i,1}^2 = [h(\beta, z_i) I_1 - E[h(\beta, z_i) I_1]]^2 = [h(\beta, z_i) - E(h(\beta, z) I_1)]^2 I_1 + E[h(\beta, z) I_1]^2 \bar{I}_1$$

Recall Holder's inequality: $E|XY| \leq E^r |X|^r \cdot E^s |Y|^s$ where $\frac{1}{r} + \frac{1}{s} = 1$. Here let $r = 3/2$ and $s = 3$. Then

$$E[h(\beta, z) I_1] \leq E[I_1^{3/2}]^{2/3} E[h(\beta, z)^3]^{1/3}$$

but $I_1^{3/2} = I_1$. Let $p_1 = E[I_1^{3/2}]$ and $c_1 = E[h(\beta, z)^3]$, then

$$E[h(\beta, z) I_1] \leq c_1 p_1^{2/3}$$

$$E[h(\beta, z) I_2] \leq c_1 p_2^{2/3}$$

Now

$$E[U_{i,1}^2 U_{i,2}^2] \leq c_1^4 p_1^{4/3} p_2^{4/3} E[\bar{I}_1 \bar{I}_2] + c_1^2 p_1^{4/3} E[(h(\beta, z) - E[h(\beta, z) I_2])^2 I_2 \bar{I}_1]$$

$$+ c_1^2 p_2^{4/3} E[I_1 \bar{I}_2 \{h(\beta, z) - E[h(\beta, z) I_1]\}^2]$$

$$E[\bar{I}_1 \bar{I}_2] = 1 - p_1 - p_2$$

$$\bar{I}_1 I_2 = I_2$$

$$E[I_2 (h(\beta, z) - E[I_2 h(\beta, z)])^2] \leq E[(h(\beta, z) I_2 - E[I_2 h(\beta, z)])^2]$$

$$\leq E[h(\beta, z)^2 I_2] \leq c_2 p_2^{2/3}$$

Hence

$$E[U_{i,1}^2 U_{i,2}^2] \leq m p_1^{2/3} p_2^{2/3} \quad m \text{ being a suitable constant}$$

Similarly:

$$E[U_1^2] E[U_2^2] \leq c_3^2 p_1^{2/3} p_2^{2/3} \quad m \text{ being a suitable}$$

constant

$$E[U_1 U_2] = E[(h(\beta, z)I_1 - E[h(\beta, z)I_1])(h(\beta, z)I_2 - E[h(\beta, z)I_2])]$$

Therefore

$$E[U_1 U_2] = -E[h(\beta, z)I_1] E[h(\beta, z)I_2] \leq c_1^2 p_1^{2/3} p_2^{2/3}$$

But:

$$E[(Z_n(t) - Z_n(t_1))^2 (Z_n(t_2) - Z_n(t))^2] \leq E[U_1^2 U_2^2] + E[U_1^4] E[U_2^4] + 2 E[U_1 U_2]^2$$

Hence for a suitable constant C_4

$$E[(Z_n(t) - Z_n(t_1))^2 (Z_n(t_2) - Z_n(t))^2] \leq C_4 [F(t_2) - F(t_1)]^{4/3}$$

A 7: Proof of lemma 7:

First step: for N large enough, $(\Phi_n(\beta))$

have a common lower-bound on a hypercube

included in the compact K .

Let us consider a hypercube C included in K :

$$C = \{ \beta \in K; (\forall k=1, \dots, p) | \beta^k - \beta_0^k | \leq d \}$$

where β^k is the k th coordinate of β . Define the hull

$$\bar{C} \text{ of } C : \bar{C} = \{ \beta; (\forall k=1, \dots, p) | \beta^k - \beta_0^k | = d \}$$

Let S_e be the set of extreme points of \bar{C} . S_e has 2^p elements. Let ϵ be fixed. For each point x_i belonging

to see $(\exists N_i)(\forall N > N_i) |\phi_N(x_i) - \phi(x_i)| < \epsilon$. Then for $N > N_0 = \text{Sup } N_i$, for each point x_i of S_ϵ , $\phi_N(x_i)$ has a lower-bound m independent of x_i .

Now let us consider $y \in (x_i, x_j)$ where x_i and x_j are two extreme points. Let A be the set of all such points y .

$\phi_N(x)$ is concave on the line (x_i, x_j) . Therefore

$$\phi_N(y) \geq \text{Inf} \{ \phi_N(x_i), \phi_N(x_j) \} \geq m$$

For each point y of A , $\phi_N(y) \geq m$.

Consider $x \in \bar{C}$, x is on line joining y_k , point of A , to y_l , another point of A . ϕ_N is concave on the line (y_k, y_l) . Therefore

$$\phi_N(x) \geq \text{Inf} \{ \phi_N(y_k), \phi_N(y_l) \} \geq m$$

For each point x of \bar{C} , $\phi_N(x) \geq m$.

Now consider $x \in C$: x is on line joining $z_i \in \bar{C}$ to $z_k \in \bar{C}$. then by the reasons as before,

$\phi_N(x)$ is lower-bounded by m for $N > N_0$

As ϕ_N is concave on C , ϕ_N has one local maximum on C : \hat{x}_N

(maybe on the frontier). We shall prove \hat{x}_N converges

to β_0 . Then for N large enough, \hat{x}_N is inside C , \hat{x}_N

is not on the frontier \bar{C} . It yields that \hat{x}_N is the local

maximum of ϕ_N not only on C but on K . We shall then

have proved that $\phi_N(\beta)$ has a maximum inside K and that

that maximum converges to β_0 .

2nd step: \hat{x}_N converges to β_0 :

Let us prove it by contradiction: let us assume that

\hat{x}_N does not converge to β_0 .

Let ϵ be fixed.

There is a subsequence $\{\hat{x}_{n'}\}$ such $(\forall n') \|\hat{x}_{n'} - \beta_0\| > \varepsilon$

Let us define $y_{n'}$ such that:

$$\begin{cases} y_{n'} = \lambda_{n'} \hat{x}_{n'} + (1 - \lambda_{n'}) \beta_0 \\ \|y_{n'} - \beta_0\| = \varepsilon \end{cases}$$

C is a compact. Then the sequence $\{y_{n'}\}$ has an accumulation point y_0 in the compact C . Let us define a subsequence

$\{y_{n_k}\}$ of $\{y_{n'}\}$ such that y_{n_k} converges to y_0 .

As β_0 is a local maximum of ϕ , $\phi(\beta_0) > \phi(y_0)$

let $\phi(\beta_0) = \phi(y_0) + \delta$. As $\phi_{n'}(y_0)$ converges to $\phi(y_0)$

and $\phi_{n'}(\beta_0)$ converges to $\phi(\beta_0)$, we have

$$(\exists N^*) (\forall N > N^*) \begin{cases} \phi(y_0) - \frac{\delta}{4} \leq \phi_N(y_0) \leq \phi(y_0) + \frac{\delta}{4} \\ \phi(\beta_0) - \frac{\delta}{4} \leq \phi_N(\beta_0) \leq \phi(\beta_0) + \frac{\delta}{4} \end{cases}$$

$(\forall N) \phi_N(y_n) \geq \phi_N(\beta_0)$ as \hat{x}_n is maximum of ϕ_N concave

and $y_n = \lambda_n \hat{x}_n + (1 - \lambda_n) \beta_0$ Then

Let $A = \{u; \|y_0 - u\| = d_2\}$. We can choose d_2 such that

ACC, as $(\forall n') \|\hat{x}_{n'} - \beta_0\| = \varepsilon$. Let $u_n \in A$ such that

$$u_n - y_0 = \rho_n (y_0 - y_n) \quad \text{with } \rho_n \in \mathbb{R}^+$$

$$\|u_n - y_0\| = d_2 = \rho_n \|y_0 - y_n\| \quad \text{then } \rho_n = \frac{d_2}{\|y_0 - y_n\|} \rightarrow +\infty$$

$$y_0 = \frac{u_n}{1 + \rho_n} + \frac{\rho_n}{1 + \rho_n} y_n$$

But

$$y_0 \in [u_n, y_n]$$

then from the concavity of ϕ_N :

$$\frac{1}{1+\mu_n} \Phi_n(u_n) + \frac{\mu_n}{1+\mu_n} \Phi_n(y_n) \leq \Phi_n(y_0) \leq \Phi(y_0) + \frac{\delta}{4}$$

Then

$$\Phi_n(u_n) \leq (1+\mu_n) \left[\Phi(y_0) + \frac{\delta}{4} \right] - \mu_n \Phi_n(y_n)$$

$$\Phi_n(u_n) \leq \Phi(y_0) + \frac{\delta}{4} + \mu_n \left[\Phi(y_0) - \Phi(\beta_0) + \frac{\delta}{2} \right]$$

$$\Phi_n(u_n) \leq \Phi(y_0) + \frac{\delta}{4} - \mu_n \frac{\delta}{2}$$

But $\mu_n \rightarrow +\infty$; we then have a contradiction as $\Phi_n(u_n)$ is bounded by m . Then:

\hat{x}_n converges to β_0

A8 Decomposition of C_n (page 62)

$$A = \int_{[0, T_F]} d\hat{Q} \frac{\hat{\varepsilon}_z}{\hat{\varepsilon}} = \int_{[0, T_F]} d(\hat{Q} - Q) \frac{\hat{\varepsilon}_z}{\hat{\varepsilon}} + \int_{[0, T_F]} dQ \frac{\hat{\varepsilon}_z}{\hat{\varepsilon}}$$

We have:

$$\frac{\hat{\varepsilon}_z}{\hat{\varepsilon}} = \frac{\varepsilon_z}{\varepsilon} + \frac{\hat{\varepsilon}_z - \varepsilon_z}{\hat{\varepsilon}} + \frac{\varepsilon_z [\varepsilon - \hat{\varepsilon}]}{\hat{\varepsilon} \varepsilon}$$

Hence:

$$A = \int_{[0, T_F]} \frac{\varepsilon_z}{\varepsilon} d(\hat{Q} - Q) + \int_{[0, T_F]} \frac{\hat{\varepsilon}_z - \varepsilon_z}{\hat{\varepsilon}} d(\hat{Q} - Q) + \int_{[0, T_F]} \frac{\varepsilon_z (\varepsilon - \hat{\varepsilon})}{\hat{\varepsilon} \varepsilon} d(\hat{Q} - Q) + \int_{[0, T_F]} dQ \frac{\hat{\varepsilon}_z}{\hat{\varepsilon}}$$

Now:

$$\int_{[0, T_F]} \frac{\varepsilon_2}{\varepsilon} d(\hat{Q} - Q) = \left[\frac{\varepsilon_2}{\varepsilon} (\hat{Q} - Q) \right]_0^{T_F} - \int_{[0, T_F]} (\hat{Q} - Q) d\left(\frac{\varepsilon_2}{\varepsilon}\right)$$

$$= - \frac{\varepsilon_2(0)}{\varepsilon(0)} [\hat{Q}(0) - Q(0)] - \int_{[0, T_F]} \frac{(\hat{Q} - Q)}{\varepsilon} d\varepsilon_2 + \int_{[0, T_F]} \frac{(\hat{Q} - Q)}{\varepsilon^2} \varepsilon_2 d\varepsilon$$

as $\hat{Q}(T_F) = Q(T_F) = 0$

Moreover

$$\int_{[0, T_F]} dQ \frac{\hat{\varepsilon}_2}{\hat{\varepsilon}} = \int_{[0, T_F]} dQ \frac{\varepsilon_2}{\varepsilon} + \int_{[0, T_F]} \frac{(\hat{\varepsilon}_2 - \varepsilon_2)}{\hat{\varepsilon}} dQ + \int_{[0, T_F]} \frac{\varepsilon_2 [\varepsilon - \hat{\varepsilon}]}{\hat{\varepsilon} \varepsilon} dQ$$

where

$$\int_{[0, T_F]} \frac{\hat{\varepsilon}_2 - \varepsilon_2}{\hat{\varepsilon}} dQ = \int_{[0, T_F]} \frac{\hat{\varepsilon}_2 - \varepsilon_2}{\varepsilon} dQ + \int_{[0, T_F]} \frac{(\hat{\varepsilon}_2 - \varepsilon_2)(\varepsilon - \hat{\varepsilon})}{\varepsilon \hat{\varepsilon}} dQ$$

and

$$\int_{[0, T_F]} \frac{\varepsilon_2 [\varepsilon - \hat{\varepsilon}]}{\hat{\varepsilon} \varepsilon} dQ = \int_{[0, T_F]} \frac{\varepsilon_2 [\varepsilon - \hat{\varepsilon}]}{\varepsilon^2} dQ + \int_{[0, T_F]} \frac{\varepsilon_2 [\varepsilon - \hat{\varepsilon}]^2}{\varepsilon^2 \hat{\varepsilon}} dQ$$

It therefore implies:

$$C_n = C_{1n} + C_{2n} + C_{3n} + C_{4n} + C_{5n} + C_{6n} + R_{1n} + R_{2n} + R_{3n} + R_{4n}$$

BIBLIOGRAPHY

- (1) Armitage, P. and Doll, R., (1957). A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *British Journal of Cancer* XI, 161-169.
- (2) Armitage, P. and Doll, R., (1961). Stochastic Models for carcinogenesis. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, Le Cam and Neyman eds., 19-38.
- (3) Armitage, P., (1974). Multistage carcinogenesis models. Presented at a "Workshop meeting" sponsored by the national Institute of environmental Sciences. Wrights Ville Beach, North Carolina.
- (4) Barnard G.A., (1963). some aspects of the fiducial argument *J. R. Statistic Society*, B25, 111-4.
- (5) Billingsley, P., (1968). *Convergence of probability measures*. Wiley. New York.
- (6) Breslow, N. and Crowley J., (1974). "A large sample study of the life table and product limit estimates under random Censorship". *Annals of Statistics* 2., 437-53.
- (7) Breslow N., (1974). "Covariance analysis of Censored data". *Biometrics*, 30., 89-99.
- (8) Cea, J. *Optimisation: theories et Algorithmes*, (Dunod 1971).
- (9) Cox, D.R., (1972). "Regression Models and Life Tables". (*Journal of the royal statistical society*), Ser. B. 34., 187-220.
- (10) Cox, D.R., (1975). "Partial Likelihood". *Biometrika*, 62., 269-79.
- (11) Efron, B., (1977). the efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.*, 72., 557-565.
- (12) Farewell, V.T., and Prentice, R.L., (1977). A study of distributional shape in life testing. *Technometrics*, 18., 69-76.
- (13) Feigl, P. and Zelen, M., (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 21., 826-838.

BIBLIOGRAPHY

- (1) Armitage, P. and Doll, R., (1957). A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *British Journal of Cancer* XI, 161-169.
- (2) Armitage, P. and Doll, R., (1961). Stochastic Models for carcinogenesis. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4, Le Cam and Neyman eds., 19-38.
- (3) Armitage, P., (1974). Multistage carcinogenesis models. Presented at a "Workshop meeting" sponsored by the national Institute of environmental Sciences. Wrights Ville Beach, North Carolina.
- (4) Barnard G.A., (1963). some aspects of the fiducial argument *J. R. Statistic Society*, B25, 111-4.
- (5) Billingsley, P., (1968). *Convergence of probability measures*. Wiley. New York.
- (6) Breslow, N. and Crowley J., (1974). "A large sample study of the life table and product limit estimates under random Censorship". *Annals of Statistics* 2., 437-53.
- (7) Breslow N., (1974). "Covariance analysis of Censored data". *Biometrics*, 30., 89-99.
- (8) Cea, J. *Optimisation: theories et Algorithmes*, (Dunod 1971).
- (9) Cox, D.R., (1972). "Regression Models and Life Tables". (*Journal of the royal statistical society*), Ser. B. 34., 187-220.
- (10) Cox, D.R., (1975). "Partial Likelihood". *Biometrika*, 62., 269-79.
- (11) Efron, B., (1977). the efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.*, 72., 557-565.
- (12) Farewell, V.T., and Prentice, R.L., (1977). A study of distributional shape in life testing. *Technometrics*, 18., 69-76.
- (13) Feigl, P. and Zelen, M., (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 21., 826-838.

- (14) Fraser, D.A.S., (1968). The structure of inference. New York. Witney.
- (15) Hartley and Sielken, (1975 b). A "non-parametric" for "safety" testing of carcinogenic agents. FDA Technical Report No. 2, Institute of Statistics, Texas A. & M. University.
- (16) Kalbfleisch, J.D. and Sprott, D.A., (1970). Application of likelihood methods to models involving a large number of parameters, (with discussion). J.R. Statistic. Soc. B.32., 175-208.
- (17) Kalbfleisch, J. and Prentice, R., (1973). "Marginal Likelihoods based on Cox's Regression and Life Model". Biometrika, 60., 267-79.
- (18) Kalbfleisch, J., (1974). "Some efficiency Calculations for survival distributions". Biometrika, 61., 31-8.
- (19) Kendall and Stuart. Advanced theory of statistics. (Hafner). New York.
- (20) Liu, P.Y., (1978). Large sample theory of the M.L.E. based on Cox's regression model for survival data. Unpublished PhD dissertation. University of Wisconsin. Madison.
- (21) Liu, P.Y. and Crowley, J. (to appear) Large sample theory of the M.L.E. based on Cox's regression model for survival data.
- (22) Loeve, M., (1963). Probability theory. Princeton.
- (23) Mantel, N. and Bryan, W.R., (1961). "Safety" Testing of Carcinogenic Agents. Journal of the National Cancer Institute, 27., 455-470.
- (24) Mantel and Myers, (1971). Problems of convergence of maximum likelihood iterative procedures in multiparameter situations. J.A.S.A., 66., 484-491.
- (25) Oakes, D., (1977). "The asymptotic information in censored survival data". Biometrika (1977), 64., 3., 441-48.
- (26) Peto, R. and Peto, J., (1972). Asymptotically efficient rank invariant test procedures. J.R. Statis. Soc., A 135., 185-206.
- (27) Prentice, R.L., (1973). Exponential survival with censoring and explanatory variables. Biometrika, 60., 279-88.

- (28) Prentice, R. and Kalbfleisch, J.D., (1979). Hazard Rate Models with Covariates. *Biometrics*, 35., 25-39.
- (29) Pyke, R. and Shorack, G.R., (1968). "Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems". *Annals of Mathematical Statistics* (1968). Vol. 39., No. 3., 755-771.
- (30) Saguez, M. Optimisation et controle optimal des systemes. Ecole centrale Paris. 1978.
- (31) Skorokhod, (1956). "Limit Theorems for stochastic processes". *Theory of Probability and its applications*. Vol. 1., No. 3., 261-280.
- (32) Tsiastis, A., (1978). A large sample study of the estimate for the integrated hazard function in Cox's Regression model for survival data. To appear in the *Annals of Statistics* 1980.
- (33) Zippin, C. and Armitage, P., (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. *Biometrics*, 22., 665-672.