This is to certify that the

dissertation entitled

REGRESSION MODELS WITH (CASE 2) INTERVAL CENSORING

presented by

Vasilis Katsikiotis

has been accepted towards fulfillment
of the requirements for

the Doctor of Philosophy degree in Statistics and Probability

Joseph C. Gardiner
Major professor

Date August 2, 1995

REGRESSION MODELS WITH (CASE 2) INTERVAL CENSORING

By

Vasilis Katsikiotis

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1995

ABSTRACT

REGRESSION MODELS WITH (CASE 2) INTERVAL CENSORING

By

Vasilis Katsikiotis

Interval censoring occurs frequently in longitudinal studies with periodic follow-up. The outcome of interest is not directly observed but its occurrence can be ascertained within an interval of successive inspection times.

The Accelerated Failure Time (AFT) and the Proportional Hazards (PH) are two of the regression models used widely in survival analysis and reliability theory. Maximum likelihood estimation is pursued in both models in a semiparametric framework. Existence of the estimators is established along the lines of Groeneboom and Wellner (1992). Strong consistency is proved and necessary conditions are given under which the information for the finite dimensional parameter is positive. The importance of the information calculation is illustrated in two ways. A lower bound for the asymptotic variance of regular estimators is derived first. Moreover, the benefit of scheduling two inspections instead of a single one is measured explicitly by the anticipated gain in the information measure. Estimates of this measure are also provided.

Lack of smoothness in the function $\theta \mapsto F_n(\cdot,\theta)$ motivates the search for alternative estimators in the AFT model. Asymptotically Generalized M-Estimators (AGME) are considered and a few of the conditions of a general theorem due to Bickel, Klaassen, Ritov and Wellner (1993) are established. A simulation study evaluates the performance of the MLE in the AFT model.

DEDICATION

To my parents

ACKNOWLEDGMENTS

I would like to thank all members of my guidance committee for their encouragement and support during my years at M.S.U. Special thanks to Professor R.V. Ramamoorthi for some very valuable suggestions. His comments have greatly improved early versions of this manuscript. My deep appreciation to Professor D. Gilliland for his critical support during my first difficult year in the U.S. To Professor P. Groeneboom, special thanks for making his computer programs available to me. Finally but not ultimately, my sincere appreciation to two people that mean so much to me. To my advisor, Professor Joseph Gardiner, whose guidance and continuous support was vital for the completion of this work. To Professor Nigel Paneth whose research, analytic thinking and approach to problems will be so influential to me now and in the years to come.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

1. Regression Models in Survival Analysis: One of the principal goals in survival analysis is to make inference about the *time to a specific response* or *event*, in relation to the *risk factors* that influence its occurrence. In most applications the identification of important risk factors is a challenging problem itself, sometimes with significant statistical input. Regression models establish a relationship between the outcome of interest and a vector of covariates. Although such models merely approximate the true relationship between two groups of variables, they become important analytic tools, especially when they build upon the characteristics of the variables involved.

Exploratory statistical methods often provide a very useful insight to the relationship between the lifetime of interest and the covariates involved. Models based on the *monotonicity of failure rates* or the *time-invariant relative risks for failure* or the *proportionality of odds*, have been used in a variety of situations with considerable success as far as goodness of fit and interpretation of results. On the other hand, general regression models often demonstrate a trade off between adaptivity to specific applications and mathematical complexity.

The *Cox Proportional Hazards Model (PHM)*, Cox (1972), is one of the most frequently used models to express the relationship between a lifetime and a vector of covariates. There is an enormous literature that refers to this model and statistical methods associated with it have been examined in a variety of situations. *Multiplicative Intensity Models (MIM),* Aalen (1978), make up another important category of regression models used in survival analysis. They are based on a product factorization of the *rate of failure* (intensity process) into a component that describes the *risk set* at a given time t and a component that describes the *risk of failing at t, given the covariates* z. Such models have been used in a variety of scientific fields like medical research, econometrics, reliability and engineering. Models with *increasing (decreasing) failure rate* are often used to incorporate some prior knowledge on the distribution of the failure time, while others that assume *proportionality of odds-ratios* are applicable to situations more general than those described by the classical PHM.

A *general regression model* establishes the relationship between a failure time and covariates through a regression function g. When g is linear (nonlinear) and known up to a finite dimensional parameter θ, then we obtain familiar formats of regression models, like the linear regression and the accelerated failure time models. When g remains unspecified to a large extent, then the problem of nonparametric regression - one of the most outstanding ones - provides a wide open field for further research, while at the same time it allows the broadest range of applications.

In this dissertation we consider estimation methods in regression models that represent the two major categories of models described earlier. In particular we analyze the Accelerated Failure Time (AFT) and the PH models under a censoring scheme that occurs frequently in longitudinal studies and studies with periodic follow up.

2. Random Censoring : One of the characteristic features in the analysis of survival data is the partial loss of information on the main variable of interest. Follow up studies have usually a finite horizon over which the outcome of interest might occur. In addition, subjects may withdraw from the study for a number of reasons or simply miss scheduled examinations. Moreover there are situations where continuous flow of information with respect to the variable of interest is virtually impossible. In such cases, an inspection scheme is needed to provide the necessary information. If X denotes *the time to event* i.e. 'failure time', then there are situations where X might be : 1) fully observed, 2) partially observed, 3) not observed at all. Usually there is a competing variable Y called the *censoring variable* and an inspection scheme that provide information about X in cases 2 and 3 respectively. This variable can be either fixed or random.

Non-random censoring is common in economics related research, where variables are observed or become of interest, whenever they fall above or below a fixed threshold. A simple example is that of an employee who plans to achieve a certain goal during his/her tenure with a company. The time at which he/she achieves that goal is observed as long as it occurs prior to the termination of the employee's service.

Random censoring occurs naturally in a variety of situations. In toxicological experiments animals injected with suspected carcinogens are monitored for tumor development (Hoel and Walburg (1972)). The presence or absence of tumor is assessed at random times of sacrifice for each animal. Depending on the lethality of the injection, the time X of tumor onset is randomly censored at the time of sacrifice Y.

Other observational studies provide a *window* that allows information on X. To assess motor development of pre-school children, a study was planned to test the skills of participating children (Leiderman et. al. (1973)). If a child had the skill prior to the initiation of the study, then the time to event is left censored at the beginning of the study. If the child develops the skill during the study period, then the time of the event is directly observed. If at the end of the study a child lacks the specific skill, then the event of interest is said to be censored to the right, at the study's termination period. The window over which X is observed consists of the random time of entry and exit for each subject. This scheme is known in the literature as *double censoring* (Turnbull (1974)).

Another form of random censorship is the one where X is never observed but known to belong to a time interval, consisting of the last negative and the first positive assessment of the event's occurrence. The time of mail delivery illustrates this form of censoring. One is interested in the time the mailman delivers a piece of mail. However only a sequence of mailbox inspections provides information about the time of delivery, which is known up to an interval. This scheme is known as *interval censoring*, (Peto and Peto (1972), Turnbull (1976)).

3. Interval Censoring - A General Scheme: Interval censoring occurs frequently in studies in which information about an event of interest is obtained by an inspection process that assesses at each inspection time, the occurrence or not of the target event. The most appealing applications of interval censoring, appear in cancer and AIDS related research. An important measure of effectiveness of a treatment therapy in cancer research is the length of time that a patient remains in remission *(remission duration)* - Rücker and Messerer (1988). Remission duration is defined as the time period between complete remission after treatment and tumor relapse. It is clear that both the initial and terminal events that define remission duration are subject to interval censoring because they can only be assessed by a sequence of inspections. A similar problem occurs in situations where one is interested in the length of the incubation period of AIDS. By *incubation period* we mean the time elapsed from infection to the onset of clinical AIDS. Evidently at least the initial event (time of infection) is in almost every situation subject to interval censoring.

The following general model that describes the interval censoring mechanism, has been proposed by Wang, Gardiner, Ramamoorthi (1994). Let $\{W_k : k \geq 1\}$ be a sequence of ordered positive random variables that represent the potential examination times for a subject. At each time W_k an assessment is made of whether or not the event of interest has occurred. Let t denote current time measured from the beginning of the study and define $N(t) = \min\{k \geq 1\colon X \leq W_k \leq t\}$, if such k exists, with $N(t) = \infty$ otherwise. Also let $M(t) = \max\{k \geq 1\colon W_k \leq t\}$, $W_0 = 0$ and $W_\infty = W_M$. Assuming at least one examination is made, N when finite, marks the first assessment at which a positive diagnosis of the event of interest, occurring at the unobserved time X is made, by time t. On the other hand if $N = +\infty$, we only have knowledge that $X > W_M$. In their paper these authors prove the identifiability of the distribution of X from the datum (W_{N-1}, W_N, N, M). The *time of diagnosis* of the event of interest is defined as $Z = W_N$, if $N < \infty$, but we only know that $Z > W_M$, when $N = +\infty$.

In what follows we will consider a rather simple interval censoring model consisting of two inspections. Groeneboom and Wellner (1992) call this scheme, case-2 interval censoring, to distinguish it from the case of a single inspection (case-1) which can be viewed as a degenerate interval censoring situation.

The manuscript is organized as follows. In chapter 2 we discuss the existence of maximum likelihood estimators in two semiparametric models. In chapter 3 we prove the strong consistency of the estimators. Chapter 4 examines the optimality in the estimation. We compute efficient scores and information lower bounds for the finite dimensional parameters and provide estimates of their asymptotic variances. In chapter 5 we consider generalized M-estimation in the AFT model. We review some fundamental results from the modern theory of empirical processes and establish a few of the necessary conditions needed to obtain the asymptotic distribution of our estimators. Finally in chapter 6 we present some simulation studies to illustrate the performance of our estimators.

Chapter 2

MAXIMUM LIKELIHOOD ESTIMATION IN TWO SEMIPARAMETRIC MODELS

1. The Cox PHM is one of the most popular regression models used extensively in survival analysis. It assumes that the conditional hazard function given a vector of covariates Z, factorizes as

$$\lambda(t|z) = \lambda_0(t)g(z) \tag{1.1}$$

with λ_0 a nonnegative function (baseline hazard) of time t and g a function independent of time. When time dependent covariates are considered, g depends on t only through the covariates Z. The choice of g marks the degree of generalization of the original model proposed by Cox (1972), in which he considered functions of the form $g(z) = e^{\theta' z}$. There are two broad categories of papers addressing (1.1). One class considers functions g known up to a finite dimensional parameter θ and the second regards g as a general nonegative function. The following is a brief historical account of the subject.

O'Sullivan (1986) considers PH models of the form

$$\lambda(t|z) = \lambda_0(t)e^{r(z)} \tag{1.2}$$

with $r : R^k \to R$. For k large, "partly linear" forms of the r function were introduced and studied by Green (1985) and Heckman (1986). An interesting version of Cox's model considers hazard functions of the form

(1.3) $$\lambda(t|z) = \exp(\theta_1 z_1)\lambda(t|z_2) \quad \textit{for} \quad z = (z_1, z_2).$$

In case Z_2 has finite support, (1.3) is called the *stratified Cox model.*

A different generalization of (1.1) was studied by Aalen (1978,1980). In Aalen's model

(1.4) $$\lambda(t|z) = \sum_{i=1}^{k} z_i \lambda_i(t)$$

where the λ_i's are unknown functions. To avoid the nonegativity constraints in (1.4) Zucker (1986) and Zucker and Karr (1987). have considered estimation when the conditional hazard function is given by

(1.5) $$\lambda(t|z) = \exp\left(\sum_{j=1}^{k} z_j \lambda_j(t)\right).$$

In the present and subsequent chapters, we will restrict our attention to the classical model (1.1) with $g(z) = e^{\theta' z}$ and we will comment on any generalizations of our results to other models.

2. The Accelerated Failure Model is widely used in reliability studies and industrial life testing. It assumes a log-linear relationship between the failure time T and the covariates Z, namely

(2.1) $$T = T_0 \exp(\theta' Z), \quad T > 0$$

with T_0 a baseline "failure time". The model's name illustrates that the covariates Z have an accelerated (decelerated) effect on the survival function of T, compared to the corresponding function of T_0. The log-transformation reduces (2.1) to the familiar linear regression model

(2.2) $$X = \theta' Z + \varepsilon$$

with $X = \log T$ and $\varepsilon = \log T_0$. This model has been studied in a variety of situations including random (right-left) censoring. Important work on the subject can be found in Buckley, James (1979), Koul, Susarla, Van Ryzin (1981), Ritov (1990), Schick (1993).

3. Maximum Likelihood in Semiparametric Models: Consider a family of distributions $\mathcal{P} = \{P_{\theta,F} : (\theta, F) \in \Theta \times \mathcal{F}\}$ on a measurable space $(\mathcal{X}, \mathcal{B})$. Let μ be a measure on $\mathcal{B}$ and $p_{\theta,F}(\cdot)$ a density of $P_{\theta,F}$ with respect to μ. Suppose that $\Theta \subset R^d$ while $\mathcal{F}$ is an infinite dimensional space.

Let (θ_0, F_0) be the true parameter and suppose that $X_1, X_2, \ldots, X_n$ is a random sample from P_{θ_0,F_0}. Define the maximum likelihood estimator $(\hat{\theta}_n, \hat{F}_n)$ by $(\hat{\theta}_n, \hat{F}_n) \equiv \underset{\Theta \times \mathrm{F}}{\arg\max} \int \log p_{\theta,F}(x) d\boldsymbol{P}_n(x)$ with $\boldsymbol{P}_n$ the empirical measure based on $X_1, X_2, \ldots, X_n$, under the assumption that such maximizer exists. Let $\xi = (\theta, F) \in \Xi \equiv \Theta \times \mathcal{F}$.

4. Interval Censoring with two inspections: Let A be a random variable having distribution F_0, (T,U) random variables with joint distribution H, Z a random vector with distribution W. Denote by J the joint distribution of (T,U,Z). Suppose that A is independent of (T,U) conditional on Z with distribution $A|Z \sim G_{\xi_0}$, $\xi_0 \in \Xi$. We will refer to (T,U) as the censoring variables and will assume throughout that $\Pr\{T < U\} = 1$.

Consider the measure space $(R \times R^2 \times R^d, \mathcal{B}, Q_\xi)$ with Q_ξ a probability measure on the Borel σ–algebra $\mathcal{B}$. Denote by $Y^0 \equiv (A,T,U,Z)$ a typical element from this space and let ϕ be the measurable transformation $Y = \phi(Y^0) = (\delta, \gamma, T, U, Z)$ where $\delta = 1_{\{A \le T\}}$, $\gamma = 1_{\{T < A \le U\}}$. Let $P_\xi = Q_\xi \circ \phi^{-1}$ be the probability measure induced by ϕ. Then P_ξ and Q_ξ relate to each other by

(4.1) $$P_\xi\{\delta, \gamma, A \times B, E\} = \int_{R \times A \times B \times E} 1^{\delta}{}_{\{a \le t\}}\, 1^{\gamma}{}_{\{t < a \le u\}}\, 1^{1-\gamma-\delta}{}_{\{a > u\}}\, dQ_\xi(a,t,u,z)$$

with $A \times B$, E Borel sets in R^2 and R^d respectively.

Our problem is to estimate $\xi_0 = (\theta_0, F_0)$ on the basis of a sample $\{Y_1, Y_2, \dots, Y_n\}$ of independent and identically distributed observations. In what follows, we consider maximum likelihood estimation in regression models under the censoring scheme described above. We will call this situation *interval censoring* without any further reference.

4a. The Cox model with interval censoring: Suppose that the hazard function λ, associated with a nonnegative random variable X, conditional on a vector of covariates

Z, is given by

(4.2) $$\lambda(x|z) = \lambda(x)e^{\theta' z}$$

We maintain all the notations and assumptions introduced earlier in the section, with the exception that $A \equiv X \geq 0 \quad w.p.\,1$. In addition we will assume that (T,U) and Z have densities h and w with respect to Lebesgue measure, which do not depend on $\theta \in \Theta$. Based on the observable $y = (\delta, \gamma, t, u, z)$, we have that for $\xi = (\theta, F) \in \Xi$

$$P_\xi\{\delta = 1 | T = t, U = u, Z = z\} = P_\xi(X \leq t | Z = z) \equiv F_\theta(t|z) = 1 - \left[\overline{F}(t)\right]^{e^{\theta z}}.$$

Using the definition of the conditional cumulative hazard function $\Lambda(t|z) \equiv \int_0^t \lambda(s|z)ds$ and (4.2), we can write the density of Y with respect to $\mu = \nu_2 \times \tau_{d+2}$

(4.3) $$p_\xi(y) = \left(1 - \overline{F}(t)^{\exp(\theta' z)}\right)^\delta \left(\overline{F}(t)^{\exp(\theta' z)} - \overline{F}(u)^{\exp(\theta' z)}\right)^\gamma \left(\overline{F}(u)^{\exp(\theta' z)}\right)^{1-\gamma-\delta} h(t, u/z) w(z),$$

where $\nu_2 =$ counting measure on $\{0,1\}^{\otimes 2}$ and $\tau_d =$ Lebesgue measure on R^d.

The transformation $\overline{F} = e^{-\Lambda}$ allows an equivalent form of the density (4.3), namely

(4.4) $$p_\xi(y) = \left(1 - \exp[-\Lambda(t)e^{\theta' z}]\right)^\delta \times \left(\exp[-\Lambda(t)e^{\theta' z}] - \exp[-\Lambda(u)e^{\theta' z}]\right)^\gamma \times$$
$$\left(\exp[-\Lambda(u)e^{\theta' z}]\right)^{1-\gamma-\delta} \times h(t, u|z) \times w(z).$$

Very often we will switch between (4.3) and (4.4) depending on the circumstances. Since we have assumed that h and w do not contain any information about θ and since our primary goal is to do inference about ξ, we can proceed safely considering h and w known. The log-likelihood function based on n independent and identically distributed observations $\{Y_1, Y_2, \ldots, Y_n\}$ is (up to an additive constant) given by

$$l_n(\theta, F; y) = \sum_{i=1}^{n} \delta_i \log\left(1 - \overline{F}(t_i)^{\exp(\theta' z_i)}\right) + \gamma_i \log\left(\overline{F}(t_i)^{\exp(\theta' z_i)} - \overline{F}(u_i)^{\exp(\theta' z_i)}\right) +$$
$$\sum_{i=1}^{n} (1 - \gamma_i - \delta_i) \log \overline{F}(u_i)^{\exp(\theta' z_i)}$$

and

$$l_n(\theta, \Lambda; y) = \sum_{i=1}^{n} \delta_i \log\left(1 - e^{-\Lambda(t_i) e^{\theta' z_i}}\right) + \gamma_i \log\left(e^{-\Lambda(t_i) e^{\theta' z_i}} - e^{-\Lambda(u_i) e^{\theta' z_i}}\right) - .$$
$$-(1 - \gamma_i - \delta_i) \Lambda(u_i) e^{\theta' z_i}$$

Early attempts to estimate the parameters in the Cox model under interval censoring were confined to purely parametric methods. Finkelstein (1986) has considered maximum likelihood estimation under the assumption that the baseline hazard has finite support. Recently Huang (1994) has completed a thesis on efficient estimation in the Cox model with case 1 interval censoring. He has proved that the MLE of the finite dimensional parameter is asymptotically normal and efficient. Although from one point of view our results can be taken as a natural generalization of Huang's results, case 2 interval censoring, still in its infancy, displays difficulties that do not appear in case 1. The biggest of all is the potential "nearness" of the two inspections. This is not a mere technicality that one has to address in a way or another but an integral part of the problem. Although there are complete results that describe the asymptotic distribution of the NPMLE in case 1, (Groeneboom (1989)), no asymptotic theory for case 2 has been developed as of today, to the best of our knowledge. In fact, not even the rate of convergence of the NPMLE - fundamental tool in efficiency considerations- is known for case 2 as opposed to case 1.

4b. The linear regression model with interval censoring: Consider the model (2.2) with ε having distribution F_0. In addition to the basic assumptions for the interval censored models that we considered earlier, we assume here that ε is independent of the covariates Z. Then for $A \equiv \varepsilon$, a density of Y with respect to μ is given (up to a multiplicative constant) by

$$p_{\xi}(y) = \{F(t-\theta'z)\}^{\delta}\{F(u-\theta'z)-F(t-\theta'z)\}^{\gamma}\{1-F(u-\theta'z)\}^{1-\gamma-\delta} \tag{4.5}$$

and the log-likelihood function by

$$\begin{aligned} l_n(\theta, F; y) = &\sum_{i=1}^{n}\delta_i \log\{F(t_i-\theta'z_i)\} + \gamma_i \log\{F(u_i-\theta'z_i)-F(t_i-\theta'z_i)\} + \\ &\sum_{i=1}^{n}(1-\gamma_i-\delta_i)\log\{1-F(u_i-\theta'z_i)\}. \end{aligned} \tag{4.6}$$

Finkelstein and Wolfe (1985) were the first to consider this model under interval censoring. To model the joint distribution of (X,Z), they introduced a parametric formulation of the conditional distribution of Z given X and they estimated the distribution of X, using the "self-consistent" nonparametric estimator of Turnbull (1976). Although they argue that their estimators are maximum likelihood in nature, it is well known today that the self-consistent equations do not always yield maximum likelihood estimators.

5. Profile Likelihood Estimation: We consider here a three step procedure that yields a maximum likelihood estimator for the models we have introduced. This is a standard approach for M-estimation in semiparametric problems and has been used among others by Anderson and Gill (1982), Whittemor and Keller (1986), Leblanc and Crowley (1995). It can be summarized in the following three steps.

<u>Step 1</u>: For $\theta \in \Theta$ fixed, consider $F_n(\cdot,\theta) = \arg\max_{F \in \mathfrak{I}} l_n(\theta, F)$.

<u>Step 2</u>: Replace F by $F_n(\cdot,\theta)$ and consider the *profile likelihood* function

$$\theta \mapsto l_n(\theta, F_n(\cdot,\theta)).$$

<u>Step 3</u>: Let $\tilde{\theta}_n = \arg\max_{\theta \in \Theta} l_n(\theta, F_n(\cdot,\theta))$.

$$\text{Set } \hat{\theta}_n \equiv \tilde{\theta}_n \quad \textit{and} \quad \hat{F}_n(\cdot) = F_n(\cdot, \tilde{\theta}_n).$$

There are a number of issues that need to be clarified before we proceed to the properties of the maximum likelihood estimator. In Step 1 we need to justify the existence of the maximizer. Moreover a practical method for the computation of the maximizer might be a priority. In the next section, we provide the arguments for the legitimacy of step 1 in the two regression models that we consider. In that regard, the work of Groeneboom and Wellner (1992) on the existence of the NPMLE is the basis for our arguments. Details on the computational aspects of the MLE from interval censored data and algorithms to carry out step 1, are given in Groeneboom and Wellner (1992). In relation to step 3, there seems to be an ad-hock assumption that $\hat{\theta}_n \in \Theta \quad \forall n$. This is not generally the case. We will only need $\hat{\theta}_n \in \Theta$ eventually, with high probability and this result will be established in Theorem 3.1. Finally, with all three steps substantiated we obtain

$$l_n(\hat{\theta}_n, \hat{F}_n) \geq l_n(\widetilde{\theta}_n, F_n(\cdot, \widetilde{\theta}_n)) \text{ with } \left(\hat{\theta}_n, \hat{F}_n\right) = \arg\max_{\Theta \times \mathcal{F}} l_n(\theta, F).$$

Moreover $l_n(\widetilde{\theta}_n, F_n(\cdot, \widetilde{\theta}_n) \geq l_n(\theta, F_n(\cdot, \theta)) \geq l_n(\theta, F) \quad \forall\ (\theta, F) \in \Theta \times \mathcal{F}$, by steps 3,1 respectively. It follows that $l_n(\hat{\theta}_n, \hat{F}_n) = l_n(\widetilde{\theta}_n, F_n(\cdot, \widetilde{\theta}_n))$. This proves that the three step procedure described above yields a maximum likelihood estimator. In the remainder of this chapter we will fix a $\theta \in \Theta$. We will also use the abbreviation $q_i = \theta' z_i$ for $i \in \{1, 2, \ldots, n\}$.

6. Characterization of the NPMLE: In this section we state necessary and sufficient conditions for an estimate of F to be a maximum likelihood estimate. Our focus is in the Cox model for which we provide the full details. In Theorem 6.2 we state similar results for the linear regression model without any proofs, since this would be a duplication of arguments to a large extent.

We consider the mapping $\Lambda \mapsto l_n(\Lambda)$ based on (4.4),

$$(6.1) \quad l_n(\Lambda; y) = \sum_{i=1}^{n} \delta_i \log\left(1 - e^{-\Lambda(t_i)e^{q_i}}\right) + \gamma_i \log\left(e^{-\Lambda(t_i)e^{q_i}} - e^{-\Lambda(u_i)e^{q_i}}\right) - (1 - \gamma_i - \delta_i)\Lambda(u_i)e^{q_i}.$$

Let $\mathcal{J}_n^{(1)} = \{T_i:\ \delta_i = 1 \text{ or } \gamma_i = 1 \text{ for } \mathrm{i} = 1, 2, \ldots, \mathrm{n}\}$,

$\mathcal{J}_n^{(2)} = \{U_i:\ \gamma_i = 1 \text{ or } 1 - \delta_i - \gamma_i = 1 \quad \text{for } \mathrm{i} = 1, 2, \ldots, \mathrm{n}\}$ and $J_n = \mathcal{J}_n^{(1)} \bigcup \mathcal{J}_n^{(2)}$.

Notice that $J_n \subset \{T_i \ ; i=1,2,\dots,n\} \bigcup \{U_i \ ; i=1,2,\dots,n\}$ marks the set of relevant observations which contributes to the likelihood function. Let $0 \le \eta_{(1)} \le \dots \le \eta_{(m)}$ be the order statistics of the elements of J_n. Write $\Lambda_j \equiv \Lambda(\eta_{(j)})$ and notice that $0 \le \Lambda_1 \le \Lambda_2 \le \dots \le \Lambda_m$ due to the monotonicity of Λ. We abuse notation when we write $\delta_{(j)}$, $\gamma_{(j)}$ or $Z_{(j)}$ referring to $\delta's$, $\gamma's$ and Z's associated with $\eta_{(j)}$. The MLE $\hat{\Lambda}_n$ of Λ_0 can be chosen to be a right continuous, nondecreasing, step function, with jumps at J_n. Set $\eta_{(0)} = 0$ and $\hat{\Lambda}_n(0) = 0$. Then $\hat{\Lambda}_n$ will have the form

$$\begin{array}{lll} & 0 & 0 \le t < \eta_{(1)} \\ \hat{\Lambda}_n(t) = & \hat{\Lambda}_n(\eta_{(j)}) & \eta_{(j)} \le t < \eta_{(j+1)} \quad j = 1,\dots,m-1 \\ & \text{unspecified} & t > \eta_{(m)}. \end{array}$$

It is worth noting that if $\delta_{(1)} = 0$ or $(\delta_{(m)} = 1$ or $\gamma_{(m)} = 1)$ then $\hat{\Lambda}_n(\eta_{(1)}) = 0$ or $\hat{\Lambda}_n(\eta_{(m)}) = \infty$ respectively. So without loss of generality we will assume that $\delta_{(1)} = 1$ and $\delta_{(m)} = \gamma_{(m)} = 0$. In this way we can restrict ourselves to functions from

$$\mathcal{L} = \left\{\Lambda : \Lambda(\eta_{(1)}) > 0,\ \Lambda(\eta_{(m)}) < \infty,\ \Lambda(\eta_{(j)}) - \Lambda(\eta_{(j-1)}) > 0 \ \ \forall j \in \{1,2,\dots,m\}\right\}.$$

In this way we can avoid pathological situations with maximizers of the form $l_n(\hat{\Lambda}_n, Y) = -\infty$. Before we state the main theorems in this section, we introduce some additional notations. Let

$$W_{\Lambda,q}(t) = \sum_{i:T_i \le t} \left(\frac{\delta_i}{1-e^{-\Lambda(T_i)e^{q_i}}} - \frac{\gamma_i}{e^{-\Lambda(T_i)e^{q_i}} - e^{-\Lambda(U_i)e^{q_i}}} \right) e^{q_i - \Lambda(T_i)e^{q_i}}$$

$$+ \sum_{i:U_i \le t} \left(\frac{\gamma_i}{e^{-\Lambda(T_i)e^{q_i}} - e^{-\Lambda(U_i)e^{q_i}}} - \frac{1-\gamma_i-\delta_i}{e^{-\Lambda(U_i)e^{q_i}}} \right) e^{q_i - \Lambda(U_i)e^{q_i}}$$

$$= n \int_{s \le t} \left(\frac{\delta}{1-e^{-\Lambda(s)e^{q}}} - \frac{\gamma}{e^{-\Lambda(s)e^{q}} - e^{-\Lambda(u)e^{q}}} \right) e^{q-\Lambda(s)e^{q}} d\boldsymbol{P}_n(\delta,\gamma,s,u,z) +$$

$$n \int_{u \le t} \left(\frac{\gamma}{e^{-\Lambda(s)e^{q}} - e^{-\Lambda(u)e^{q}}} - \frac{1-\gamma-\delta}{e^{-\Lambda(u)e^{q}}} \right) e^{q-\Lambda(u)e^{q}} d\boldsymbol{P}_n(\delta,\gamma,s,u,z).$$

Note that the process W has jumps at points of J_n.

THEOREM 6.1. *For fixed* $\theta \in \Theta$, *let* $q_i = \theta' z_i$ for $i \in \{1,2,\ldots,n\}$. *Suppose that* $\delta_{(1)} = 1$ and $\delta_{(m)} = \gamma_{(m)} = 0$. *The following conditions are sufficient and necessary for a function* $\widetilde{\Lambda}_n(\cdot,\theta)$ *to maximize (6.1) over* $\mathcal{L}$.

$$i) \quad \int_t^\infty dW_{\widetilde{\Lambda}_n,q}(s) \le 0 \quad \forall t \ge 0 \qquad \text{and} \qquad ii) \quad \int_0^\infty \widetilde{\Lambda}_n(s,\theta)\, dW_{\widetilde{\Lambda}_n,q}(s) = 0.$$

Moreover $\widetilde{\Lambda}_n(\cdot,\theta)$ *is uniquely determined by i) and ii).*

Proof. Define $S = \left\{ \widetilde{x} = (\Lambda_1,\Lambda_2,\ldots,\Lambda_m) \ : \ \Lambda_j = \Lambda\left(\eta_{(j)}\right), \quad \Lambda \in \mathcal{L} \right\}$ and let $\varphi : S \to R$ be the function $\varphi(\widetilde{x}) = l_n(\Lambda)$ if $\widetilde{x} = (\Lambda_1,\Lambda_2,\ldots,\Lambda_m)$. We supress the dependence of the likelihood function from θ. Notice that φ is non positive on S. If $\hat{x}_n = \arg\max_{\widetilde{x} \in S} \varphi(\widetilde{x})$ then we set $\left(\hat{\Lambda}_1,\hat{\Lambda}_2,\ldots,\hat{\Lambda}_m\right) \equiv \hat{x}_n$. It is easy to see that

$$\frac{\partial\varphi}{\partial\tilde{x}_i}(\tilde{x}) = W_{\Lambda,q}(\eta_i) - W_{\Lambda,q}(\eta_{i-1}) \quad \forall i \in \{1,\dots,m\}.$$

Suppose that (i), (ii) hold and let $\Lambda \in \mathcal{L}$. Set $\tilde{y} = (\tilde{\Lambda}_1,\dots,\tilde{\Lambda}_m)$ and write $\tilde{\Lambda}$ instead of $\tilde{\Lambda}_n$. In the Appendix we prove that φ is concave. Exploring this property of φ and the structure of S, we obtain for arbitrary $\pi \in (0,1)$

$$\begin{aligned}\varphi(\tilde{x}) - \varphi(\tilde{y}) &\le \frac{1}{\pi}\big(\varphi(\pi\tilde{x} + (1-\pi)\tilde{y}) - \varphi(\tilde{y})\big) \\ &= \frac{\varphi(\pi(\tilde{x}-\tilde{y}) + \tilde{y}) - \varphi(\tilde{y})}{\pi(\tilde{x}-\tilde{y})}(\tilde{x}-\tilde{y}) \xrightarrow[\pi\to 0]{} \langle\nabla\varphi(\tilde{y}),\, \tilde{x}-\tilde{y}\rangle.\end{aligned}$$

Thus $\displaystyle l_n(\Lambda) - l_n(\tilde{\Lambda}) = \varphi(\tilde{x}) - \varphi(\tilde{y}) \le \langle\nabla\varphi(\tilde{y}),\, \tilde{x}-\tilde{y}\rangle = \sum_{j=1}^{m} \Delta W_{\tilde{\Lambda},q}(\eta_{(j)})(\Lambda_j - \tilde{\Lambda}_j)$

$$= \int_0^\infty \big(\Lambda(s) - \tilde{\Lambda}(s)\big)dW_{\tilde{\Lambda},q}(s) \overset{(ii)}{=} \sum_{j=1}^{m} \Lambda_j \Delta W_{\tilde{\Lambda},q}(\eta_{(j)}) = \langle\tilde{x}, \nabla\varphi(\tilde{y})\rangle.$$

Now we can write $\tilde{x}$ as a telescopic sum

$$\tilde{x} = \sum_{j=1}^{m} a_j \mathbf{1}_j, \quad \mathbf{1}_j = (0,\dots,0,1,\dots,1)$$

with the last j components equal to one and $a_j = \Lambda_{m-j+1} - \Lambda_{m-j}$, $1 \le j \le m$. It follows that

$$l_n(\Lambda) - l_n(\tilde{\Lambda}) \le \sum_{j=1}^{m} a_j \langle\mathbf{1}_j, \nabla\varphi(\tilde{y})\rangle = \sum_{j=1}^{m} a_j \sum_{l \ge j} \Delta W_{\tilde{\Lambda},q}(\eta_{(l)})$$

$$= \sum_{j=1}^{m} a_j \int_{\eta_{(j)}}^{\infty} dW_{\tilde{\Lambda},q}(s) \overset{(i)}{\le} 0.$$

So if $\tilde{\Lambda}_n$ satisfies (i) and (ii) then it maximizes (6.1) over $\mathcal{L}$. Now suppose that

$\widetilde{\Lambda}_n = \arg\max_{\Lambda \in \mathcal{L}} l_n(\Lambda)$. For $\widetilde{y} = \left(\widetilde{\Lambda}_1, \ldots, \widetilde{\Lambda}_m\right) \in S$, we can express the maximizer of the log-likelihood function with respect to $\widetilde{y} = \arg\max_{\widetilde{x} \in S} \varphi(\widetilde{x})$.

For an $\varepsilon > 0$ we have $\left(\widetilde{y} + \varepsilon \mathbf{1}_j\right) \in S \quad \forall\, 1 \le j \le m$.

Thus

$$\begin{aligned}
\lim_{\varepsilon \to 0} \frac{\varphi\left(\widetilde{y} + \varepsilon \mathbf{1}_j\right) - \varphi(\widetilde{y})}{\varepsilon} &= \left\langle \mathbf{1}_j, \nabla\varphi(\widetilde{y})\right\rangle = W_{\widetilde{\Lambda},q}\left(\eta_{(m)}\right) - W_{\widetilde{\Lambda},q}\left(\eta_{(m-j+1)}\right) \\
&= \int_{\eta_{(m-j+1)}}^{\infty} dW_{\widetilde{\Lambda},q}(s) \le 0 \quad , \qquad \forall\, 1 \le j \le m.
\end{aligned}$$

Since $(\widetilde{y} + h\widetilde{y}) \in S \quad \forall h > 0$,

$$\lim_{h \to 0} \frac{\varphi(\widetilde{y} + h\widetilde{y}) - \varphi(\widetilde{y})}{h} = \left\langle \widetilde{y}, \nabla\varphi(\widetilde{y})\right\rangle = \int_0^{\infty} \widetilde{\Lambda}(s)\, dW_{\widetilde{\Lambda},q}(s) = 0,$$

because $\widetilde{y}$ is the maximizer of φ. Thus (i) and (ii) hold.

To prove uniqueness of the estimator we need to preview Proposition 2.1 in the appendix. It provides an alternative characterization of the maximum likelihood estimator as the left derivative of the convex minorant of a self induced cumulative sum diagram. Since two consecutive vertices of the convex minorant always include a point that corresponds to a left $(\delta = 1)$ or interval $(\gamma = 1)$ censored observation, it is enough to prove uniqueness of the M.L.E. on this set of the observation points only.

Suppose that besides $\widetilde{y} = \left(\widetilde{\Lambda}_1, \ldots, \widetilde{\Lambda}_m\right) \in S$, there exists another maximizer of (6.1), call it $\widetilde{y}^* = \left(\Lambda^*_1, \ldots, \Lambda^*_m\right) \in S$. Then a second order Taylor approximation will give

$$0 = \varphi(\tilde{y}^{*}) - \varphi(\tilde{y}) = (\tilde{y}^{*} - \tilde{y})' \nabla\varphi(\tilde{y}) - \frac{1}{2}(\tilde{y}^{*} - \tilde{y})' M(p)\,(\tilde{y}^{*} - \tilde{y})$$

$$= -\frac{1}{2}(\tilde{y}^{*} - \tilde{y})' M(p)\,(\tilde{y}^{*} - \tilde{y}),$$

where $M(p) = -\left[\nabla^2_{i,j}\varphi(p)\right]^{m}_{i,j=1}$ with $|\tilde{y} - p| \le |\tilde{y} - \tilde{y}^{*}|$.

The matrix M of second derivatives has entries of the form,

$$\nabla^2_{i,j}\varphi(\tilde{x}) = \frac{\partial^2\varphi(\tilde{x})}{\partial x_i \partial x_j} = -\left[\delta_{(i)} \frac{R^2_{(i)} e^{-\Lambda_i R_{(i)}}}{\left(1 - e^{-\Lambda_i R_{(i)}}\right)^2} 1_{\{T_k = \eta_{(i)}\}} + \gamma_{(i)} \frac{R^2_{(i)} e^{-\Delta_k R_{(i)}}}{\left(1 - e^{-\Delta_k R_{(i)}}\right)^2} 1_{\{U_k = \eta_{(i)}\}}\right], \quad \textit{if} \quad i = j$$

$$= 0 \quad , \text{ else}$$

for some $1 \le k \le n$ *and* $\Delta_k = \Lambda(U_k) - \Lambda(T_k)$.

It follows that $\sum_{i:\, \delta_{(i)}=1 \text{ OR } \gamma_{(i)}=1} (y_i^{*} - y_i)^2 \nabla^2_{ii}\varphi(\tilde{y}) = 0$ from which we obtain $y_i^{*} = y_i \ \forall i$ such that $\delta_{(i)} = 1$ or $\gamma_{(i)} = 1$. This establishes the uniqueness of our estimator. Theorem (6.1) is now proved.

REMARK 6.2. As a result of Proposition 2.1 in the appendix, conditions (i) and (ii) always hold. This along with the consistency result ($\theta_0 \in \mathring{\Theta}$, $\hat{\theta}_n \xrightarrow{w.p.1} \theta_0$) confirms that the M.L.E. $\hat{\Lambda}_n(\cdot) = \tilde{\Lambda}_n(\hat{\theta}_n, \cdot)$ always exists for n sufficiently large.

REMARK 6.3. If we set $\hat{F}_n(\cdot) = 1 - e^{-\hat{\Lambda}_n(\cdot)}$ then we obtain a nondecreasing, right continuous step function, satisfying $\hat{F}_n(0)=0, \quad 0<\hat{F}_n\left(\eta_{(i)}\right)<1 \quad \forall \ 1\le i\le m$. This function is defined explicitly as a result of our Theorem 6.1 and has jumps at the same points as $\hat{\Lambda}_n$ does. The monotonicity of our transformation $\Lambda \to 1-e^{-\Lambda}$ implies that $\hat{F}_n = \arg\max_{F\in\mathcal{F}} l_n(F)$, with $\mathcal{F}$ = {subdistribution function F : $F(0)=0, \ 0<F\left(\eta_{(i)}\right)<1$}.

REMARK 6.4. We can treat the problem of maximum likelihood estimation in the linear regression model (2.2) in the same way. To justify step 1 in the profile likelihood approach, we need to consider

$$T_i^\theta = T_i - \theta' z,\ U_i^\theta = U_i - \theta' z,\ \delta_i^\theta = \mathbf{1}_{\{\varepsilon_i \le T_i^\theta\}},\ \gamma_i^\theta = \mathbf{1}_{\{T_i^\theta < \varepsilon_i \le U_i^\theta\}}$$

The following theorem is the direct analog of Theorem 6.1 for the regression model (2.2). We state it without a proof.

THEOREM 6.2. *For a fixed* $\theta \in \Theta$, *suppose that* $\delta_{(1)}^\theta = 1$ and $\delta_{(m)}^\theta = \gamma_{(m)}^\theta = 0$. *The following conditions are sufficient and necessary for a function* $\tilde{F}_n(\cdot,\theta)$ *to maximize (4.6) over* $\mathcal{F}$.

i) $\int_t^\infty dW_{\tilde{F}_n,\theta}(s) \le 0 \quad \forall t\ge 0$ and *ii)* $\int_0^\infty \tilde{F}_n(s,\theta)\, dW_{\tilde{F}_n,\theta}(s) = 0$, *with*

$$W_{F,\theta}(t) = \sum_{i=1}^n \mathbf{1}_{\{T_i^\theta \le t\}}\left(\frac{\delta_i^\theta}{F(T_i^\theta)} - \frac{\gamma_i^\theta}{F(U_i^\theta)-F(T_i^\theta)}\right) + \sum_{i=1}^n \mathbf{1}_{\{U_i^\theta \le t\}}\left(\frac{\gamma_i^\theta}{F(U_i^\theta)-F(T_i^\theta)} - \frac{1-\delta_i^\theta-\gamma_i^\theta}{1-F(U_i^\theta)}\right).$$

Moreover $\tilde{F}_n(\cdot,\theta)$ *is uniquely determined by i) and ii).*

Chapter 3

STRONG CONSISTENCY OF M.L.E.

In chapter 2 we have defined the maximum likelihood estimator $\left(\hat{\theta}_n, \hat{\Lambda}_n\right)$ and $\left(\hat{\theta}_n, \hat{F}_n\right)$ in the Cox and the linear regression models respectively. In the next two theorems we prove its consistency under a suitable topology on the parameter space. Pfanzagl (1988) proves consistency of the NPMLE under a global condition -compactness of Θ - and a local condition - continuity of $\theta \to f_\theta$ in the neighborhood of θ_0. Van de Geer (1993) obtains consistency of the NPMLE with respect to the Hellinger distance using some entropy calculations. Although her results cover a much wider range of applications and in certain cases lead to rates of convergence, we adapt Pfanzagl's approach here, since it is more direct and suitable to the demands of the semiparametric nature of our problem. Theorem 1 presents the consistency result for the Cox model, while Theorem 2 is its direct counterpart for the linear regression.

For the Cox model described by (2.4.2), a density of P_ξ with respect to $\mu = \nu_2 \times \tau_{2+d}$ is given by

$$p_\xi(y) = \left(1 - \bar{F}(t)^{\exp(\theta' z)}\right)^{\delta} \left(\bar{F}(t)^{\exp(\theta' z)} - \bar{F}(u)^{\exp(\theta' z)}\right)^{\gamma} \left(\bar{F}(u)^{\exp(\theta' z)}\right)^{1-\gamma-\delta} h(t, u / z) w(z)$$

with $\xi = (\theta, F) \in \Xi \equiv \Theta \times \mathcal{F}$, ν_2=counting measure on $\{0,1\}^{\otimes 2}$ and τ_d Lebesgue measure on R^d . Define

$$f_\xi(y) \equiv \delta\left(1-\overline{F}(t)^{exp(\theta' z)}\right)+\gamma\left(\overline{F}(t)^{exp(\theta' z)}-\overline{F}(u)^{exp(\theta' z)}\right)+(1-\gamma-\delta)\left(\overline{F}(u)^{exp(\theta' z)}\right).$$

Let S_H be the support of H and define $a_0 \equiv \sup\{x: F_0(x)=0\}$, $b_0 \equiv \inf\{x: F_0(x)=1\}$.

THEOREM 1. (Consistency in the Cox model)

Suppose that (i) Θ *is a subset of* R^d *with bounded closure* .

(ii) $\theta_0 \in \mathring{\Theta}$, *the interior of* Θ.

(iii) $\forall\ \theta \neq \theta_0 \quad Pr\{\theta Z \neq \theta_0 Z\} > 0$

(iv) $\forall\ (t,u) \in S_H, \quad 0 \le a_0 < t < u < b_0$.

Then $\hat{\theta}_n \xrightarrow{as} \theta_0$ *and* $\hat{F}_n(y) \xrightarrow{as} F_0(y) \quad \forall y \in E \cap C_{F_0}$, *under*

$P_0 \equiv P_{(\theta_0, F_0)}$, *where* C_{F_0} *denotes the set of continuity points of* F_0 *and* $E=(a_0, b_0)$. *In the special case that* F_0 *is continuous on E, the above convergence is uniform with probability one. i.e.* $\sup_{y\in E}\left|\hat{F}_n(y)-F_0(y)\right| \xrightarrow{as} 0$.

REMARK 1. Assumptions *(i)-(iii)* are essentially the same as in Pfanzagl (1988). While assumption *(iii)* safeguards against non-identifiability problems, *(iv)* is naturally satisfied in almost all problems with interval censored observations. Moreover *(iv)* is essential here since the MLE $\hat{F}_n$ is uniquely defined on S_H and it is a step function on $[0, b_0)$. Finally Pfanzagl's assumption of a concave density with respect to the unknown parameter, seems unnecessary here. He used it in order to verify condition (2.6), page

140, in his Lemma 2.5 -originally due to Wang (1985). We show that a more direct way is indeed possible.

Proof of Theorem 1. Consider the measurable space $(E,\mathcal{B})$ with $\mathcal{B}$ the Borel σ-field. Let $\mathcal{M} = \{$ measures F on $(E,\mathcal{B})$: $F(E) \le 1$ $\}$. We have defined earlier $\mathcal{F}$ to be the set

$\mathcal{F} = \{F = \text{distr. function: } F(0) = 0,\ 0 < F(\eta_{(i)}) < 1, F(\eta_i) - F(\eta_{i-1}) > 0 \quad \forall i \in \{1,2,\ldots,m\}\}$.

Note that $\mathcal{F} \subset \mathcal{M}$. Equip $\mathcal{M}$ with the topology $\mathcal{T}_1$ of vague convergence, i.e. the smallest topology that makes functions of the form $F \in \mathcal{M} \mapsto \int f dF$ continuous $\forall f \in C^c(E)$, the space of continuous functions with compact support. Helly's theorem asserts that ($\mathcal{M}$, $\mathcal{T}_1$) is a vaguely compact topological space. Equip Θ with the usual Euclidean topology $\mathcal{T}_2$. Then $(\overline{\Theta} \times \mathcal{M}, \mathcal{T})$ is compact in the product topology $\mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2$. We say that $(\theta_n, F_n) \xrightarrow{in\ \mathcal{T}} (\theta, F)$ if and only if $\theta_n \to \theta$ and $F_n \xrightarrow{v} F$, with the latter denoting vague convergence.

Fix an arbitrary $a \in (0,1)$ and let $F \in \mathcal{F}$. For $\xi = (\theta, F) \neq (\theta_0, F_0) = \xi_0$, $E \equiv E_{\xi_0}$, $f \equiv f_\xi$, $f_0 \equiv f_{\xi_0}$, Jensen's inequality applied to the strictly concave function $\varphi(y) = \log[1 + a(y-1)]$ $y \ge 0$, along with *(iii)* gives

$$E\varphi\left(\frac{f(Y)}{f_0(Y)}\right) < \varphi\left(E\frac{f(Y)}{f_0(Y)}\right) \tag{1.1}$$
$$= \varphi\left(\int f(y)\mu(dy)\right) \le \log 1 = 0.$$

By the concavity of the log function and the fact that $\hat{\xi}_n$ is the MLE of ξ , we obtain

$$l_n(\hat{\xi}_n) \geq l_n(\xi_0) \quad \text{which implies} \quad \sum_{i=1}^{n} \log \frac{\hat{f}_n(Y_i)}{f_0(Y_i)} \geq 0 \quad \text{and}$$

$$(1.2) \qquad \sum_{i=1}^{n} \varphi\left(\frac{\hat{f}_n(Y_i)}{f_0(Y_i)}\right) = \sum_{i=1}^{n} \log\left[1 + a\left(\frac{\hat{f}_n(Y_i)}{f_0(Y_i)} - 1\right)\right]$$

$$= \sum_{i=1}^{n} \log\left[a\left(\frac{\hat{f}_n(Y_i)}{f_0(Y_i)}\right) + (1-a)\right] \geq a \sum_{i=1}^{n} \log \frac{\hat{f}_n(Y_i)}{f_0(Y_i)} \geq 0.$$

Consider now a collection $\{\mathcal{N}_\varepsilon(\xi)\colon \varepsilon > 0\}$ of nested neighborhoods of ξ. Let $\mathcal{N}_\varepsilon \equiv \mathcal{N}_\varepsilon(\xi)$ and $\tilde{f}_\varepsilon(y) = \sup_{\xi' \in \mathcal{N}_\varepsilon} f_{\xi'}(y)$. Notice that $\xi \to f_\xi(y)$ is continuous under $\mathcal{T}$ for μ-*a.e* y, and bounded above by 1. Thus $\tilde{f}_\varepsilon(y) \downarrow_{\varepsilon \downarrow 0} f(y)$ for μ-*a.e* y.

We want to prove the measurability of $y \mapsto \tilde{f}_\varepsilon(y)$. Notice that for $\overline{\mathcal{N}}_\varepsilon$ denoting the closure of $\mathcal{N}_\varepsilon$ under $\mathcal{T}$, we can find $\hat{\xi} \in \overline{\mathcal{N}}_\varepsilon$ such that $f_{\hat{\xi}}(\cdot) = \sup_{\xi' \in \overline{\mathcal{N}}_\varepsilon} f_{\xi'}(\cdot)$. If $\hat{\xi} \in \mathcal{N}_\varepsilon$ then measurability of $y \mapsto f_{\hat{\xi}}(y)$ holds and there is nothing to prove. However if $\hat{\xi} \in \overline{\mathcal{N}}_\varepsilon \setminus \mathcal{N}_\varepsilon$ then $\exists \langle \xi_n \rangle_{n \in N}$ in $\mathcal{N}_\varepsilon$ such that $\xi_n \xrightarrow{\mathcal{T}} \hat{\xi}$. Thus $f_{\xi_n}(y) \xrightarrow{as-\mu} f_{\hat{\xi}}(y)$ and by the completeness of μ, $f_{\hat{\xi}}$ is $\mathcal{B}$-measurable. It follows that

$$\varphi\left(\frac{\tilde{f}_\varepsilon(y)}{f_0(y)}\right) \xrightarrow{as-\mu} \varphi\left(\frac{f(y)}{f_0(y)}\right) \quad as \quad \varepsilon \to 0.$$

If $\langle M_k : k \in N \rangle$ with $M_k > 0 \ \forall k$ and $M_k \uparrow \infty$ then by the Lebesgue dominated convergence theorem we obtain

$$\lim_{\varepsilon \downarrow 0} E\left(\varphi\left(\frac{\tilde{f}_\varepsilon(y)}{f_0(y)}\right) \wedge M_k\right) = E\left(\varphi\left(\frac{f(y)}{f_0(y)}\right) \wedge M_k\right) \quad \forall k .$$

Moreover since $\varphi\left(\frac{\tilde{f}_\varepsilon(y)}{f_0(y)}\right) \wedge M_k \geq \log(1-a) \quad \forall k$, an application of the monotone convergence theorem and (1.1) give

$$\lim_{k \uparrow \infty} \lim_{\varepsilon \downarrow 0} E\left(\varphi\left(\frac{\tilde{f}_\varepsilon(y)}{f_0(y)}\right) \wedge M_k\right) = \lim_{k \uparrow \infty} E\left(\varphi\left(\frac{f(y)}{f_0(y)}\right) \wedge M_k\right)$$

$$= E\left(\varphi\left(\frac{f(y)}{f_0(y)}\right)\right) < 0.$$

Thus for $\varepsilon > 0$ small $\exists\, k \equiv k(\varepsilon)$:

$$E\left(\varphi\left(\frac{\tilde{f}_\varepsilon(Y)}{f_0(Y)}\right) \wedge M_k\right) < 0. \tag{1.3}$$

Let $\mathcal{B}$ be an arbitrary neighborhood of ξ_0 and set $\Xi_* = \overline{\Theta} \times \mathcal{M}$. Since $\mathcal{B}' = \Xi_* \setminus \mathcal{B}$ is compact in $\mathcal{J}$, any open cover of $\mathcal{B}'$ contains a finite subcover. Let $\{U_\xi : \ \xi \notin \mathcal{B}\}$ be one such cover of $\mathcal{B}'$. Then we can find $\{\xi_1, \xi_2, \ldots, \xi_m\} \in \mathcal{B}'$ such that $\mathcal{B}' \subset \bigcup_{j=1}^{m} U_j$ for $U_j \equiv U_{\xi_j}$. Therefore from (1.2) we obtain,

$$\begin{aligned} \left\{\hat{\xi}_n \notin \mathcal{B}\right\} &\subset \bigcup_{j=1}^{m} \left\{\hat{\xi}_n \in U_j\right\} \\ &\subset \bigcup_{j=1}^{m} \left\{\sum_{i=1}^{n} \varphi\left(\frac{\tilde{f}_j(Y_i)}{f_0(Y_i)}\right) \geq 0\right\} \subset \bigcup_{j=1}^{m} \left\{\sum_{i=1}^{n} \varphi\left(\frac{\tilde{f}_j(Y_i)}{f_0(Y_i)}\right) \wedge M_j \geq 0\right\} \end{aligned} \tag{1.4}$$

for some $M_j > 0$, with $\tilde{f}_j = \sup_{\xi \in U_j} f_\xi$.

Set $Y'_{j,i} = \varphi\left(\frac{\tilde{f}_j(Y_i)}{f_0(Y_i)}\right) \wedge M_j$. The random variables $\{Y'_{j,i}: i \geq 1\}$ are i.i.d $\forall j \in \{1,2,\ldots m\}$.

Moreover $\left|Y'_{j,i} - EY'_{j,i}\right| \leq M_j - \log(1-a)$ and as we have shown in (1.3), $EY'_{j,i} < 0$.

Applying Hoeffding's inequality to the sum of centered, bounded random variables $Y'_{j,i} - EY'_{j,i}$, we obtain

$$
\begin{aligned}
P_0\left\{\hat{\xi}_n \notin \mathcal{B}\right\} &\leq \sum_{j=1}^{m} P_0\left\{\sum_{i=1}^{n}(Y'_{j,i} - EY'_{j,i}) \geq -nEY'_{j,1}\right\} \\
&\leq \sum_{j=1}^{m} \exp\left(-2\left(-nEY'_{j,1}\right)^2 / 4n\left(M_j - \log(1-a)\right)^2\right) \\
&\leq \sum_{j=1}^{m} \exp\left(-n\beta_j{}^2 / 2\left(M_0 - \log(1-a)\right)^2\right) \leq m\exp\left(-n\beta_0{}^2 / 2\left(M_0 - \log(1-a)\right)^2\right),
\end{aligned}
$$

where $\beta_j = EY'_{j,1} < 0 \quad \forall j \in \{1,\ldots,m\}$, $\beta_0 = \bigvee_{j=1}^{m} \beta_j$, $M_0 = \bigwedge_{j=1}^{m} M_j$.

Thus $\sum_{n=1}^{\infty} P_0\left\{\hat{\xi}_n \notin \mathcal{B}\right\} < \infty$ and Borel-Cantelli gives

$$
P_0\left\{\hat{\xi}_n \in \mathcal{B} \quad \text{eventually}\right\} = 1 .
$$

Althought the above argument seems to depend heavily on the choice of $\mathcal{B}$, it holds in a straightforward manner for any such $\mathcal{B}$. This is true because the parameter space has a countable base. It follows that $\hat{\xi}_n = \left(\hat{\theta}_n, \hat{F}_n\right) \xrightarrow{in\ \mathcal{T}} (\theta_0, F_0)$ a.s. P_0^{∞} and the theorem is now proved.

REMARK 2. One of the advantages of Pfanzagl's ideas incorporated in the above proof is the lack of dependence from the exact form of the MLE. This is extremely usefull particularlly in cases where a close form solution of the MLE is very difficult or practically impossible to obtain. The models that we consider here contribute in the above proof in three aspects. Firstly a topology on the parameter space is chosen that makes sense from a statistical point of view. Secondly, both the global and local conditions, necessary for the existence of the MLE and the verification of Pfanzagl's assumptions, hold true in our models. Finally these conditions are easily verified in real life applications with interval censored data.

COROLLARY 1. *Under the assumptions of Theorem 1 and for* $\hat{\Lambda}_n \equiv -\log(1-\hat{F}_n)$ *we obtain*

$$\hat{\Lambda}_n(y) \xrightarrow{as} \Lambda_0(y) \quad \forall y \in E \cap C_{F_0}$$

and in case that F_0 *is continuous on* E, $\sup_{y \in E}\left|\hat{\Lambda}_n(y) - \Lambda_0(y)\right| \xrightarrow{as} 0$.

We now turn to the linear regression model (2.2.2). A density of P_ξ with respect to $\mu = \nu_2 \times \tau_{2+d}$ is given by

$$p_\xi(y) = (F(t-\theta z))^\delta (F(u-\theta z) - F(t-\theta z))^\gamma (\bar{F}(u-\theta z))^{1-\gamma-\delta} h(t,u|z)w(z),$$

with $\xi = (\theta, F)$, ν_2=counting measure on $\{0,1\}^{\otimes 2}$ and τ_d Lebesgue measure on R^d.

Let $f_\xi(y) = \delta F(t-\theta z) + \gamma[F(u-\theta z) - F(t-\theta z)] + (1-\gamma-\delta)\bar{F}(u-\theta z)$, with $T_0 = T - \theta_0 Z$, $U_0 = U - \theta_0 Z$. Denote by V, J the joint distributions of (T_0, U_0) and

(T,U,Z) respectively and by S_V the support of V.

THEOREM 2. (Consistency in the Linear Regression model)

Suppose that: *(i)* Θ *is a subset of* R^d *with bounded closure*

(ii) $\theta_0 \in \mathring{\Theta}$, *the interior of* Θ

(iii) $\forall\ \theta \neq \theta_0 \quad Pr\{\theta Z \neq \theta_0 Z\} > 0$

(iv) $\forall\ (t,u) \in S_V,\ \ 0 \leq a_0 < t < u < b_0,$

with $a_0 = \sup\{x : F_0(x) = 0\}$ and $b_0 = \inf\{x : F_0(x) = 1\}$.

Then

$$\hat{\theta}_n \xrightarrow{as} \theta_0 \ \ and \ \ \hat{F}_n(y) \xrightarrow{as} F_0(y) \quad \forall y \in E \cap C_{F_0},$$

under $P_0 \equiv P_{(\theta_0, F_0)}$, *where* C_{F_0} *denotes the set of continuity points of* F_0 *and* $E = (a_0, b_0)$. *In the special case that* F_0 *is continuous on* E, *the above convergence is uniform with probability one, i.e* $\sup_{y \in E} \left| \hat{F}_n(y) - F_0(y) \right| \xrightarrow{as} 0$.

Proof of Theorem 2. As in Theorem 1, we endow the parameter space with the product topology $\mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2$. Continuity of $\xi \mapsto f_\xi(y)$ holds for μ-*a.e* *y*, under $\mathcal{T}$ and the proof goes through as in Theorem 1.

REMARK 3. In his thesis, Huang (1994) gives a proof of the consistency of M.L.E. with case 1 interval censoring, based on a uniform law of large numbers (generalized Glivenco-Cantelli) for V.C. subgraph, classes of functions. This nice proof is based on the specific form of the model under consideration - as opposed to the proof we gave earlier-. Moreover it introduces some useful technics - largely due to the powerful results from the theory of empirical processes. This proof can also work in case 2 interval censoring, with some modifications. We present it here for our linear regression model.

Second Proof of Theorem 2. Using the same argument as in Theorem 1, we will treat $(\overline{\Theta}\times\mathcal{M},\mathcal{T})$ as a compact topological space with $\mathcal{T}$ the product topology. Let $\Omega=\{\omega=\langle y_n\rangle_{n\in N}:\ y_j=(\delta_j,\gamma_j,t_j,u_j,z_j)\}$, $\mathcal{B}$ the Borel σ-algebra on Ω, $P_0\equiv P_{\xi_0}$ the true underlined distribution, $(\Omega,\mathcal{B},P_0^\infty)$ the corresponding probability space. Let $\boldsymbol{P}_n$ be the empirical distribution based on i.i.d. observations $Y_1,Y_2,\ldots,Y_n$. From the strong law of large numbers (S.L.L.N) and a separability argument

$$P_0^\infty\{\omega:\ \mathrm{P}_n(y;\omega)\xrightarrow[n]{}P_0(y)\}=1\quad\forall\,y=(\delta,\gamma,t,u,z).$$

Now fix an ω from the above set and denote $\hat{\theta}_n\equiv\hat{\theta}_n(\omega)$, $\hat{F}_n(\cdot)\equiv\hat{F}_n(\cdot,\omega)$.

For every subsequence $(n')\subset(n)$, $\exists\ (n'')\subset(n')$ and a (θ_*,F_*) $\overline{\Theta}\times\mathcal{M}$, such that

$$\hat{\theta}_{n''}\to\theta_*\ \text{ and }\ \hat{F}_{n''}\xrightarrow{v}F_*. \tag{2.0}$$

We want to show that $\theta_*\equiv\theta_0$ and $F_*\equiv F_0$. Define

$$\hat{a}_n=\inf\{t-\hat{\theta}_nz:(t,u,z)\in S_J\},\ \hat{b}_n=\sup\{u-\hat{\theta}_nz:(t,u,z)\in S_J\}$$

and let a_*,b_* denote their counterparts when θ_* replaces $\hat{\theta}_n$. At first we would like to prove

(2.1) $\forall\, a,b \in (a_*,b_*)$ with $a<b$, $\exists\, M_1, M_2$: $0 < M_1 \le \hat{F}_n(a) \le \hat{F}_n(b) \le M_2 <$

for large n.

The maximum likelihood property and the definition of $\hat{F}_n$ give

$$\text{(2.2)}\quad \forall n \quad l_n(\xi_0) \le l_n(\hat{\xi}_n) \le \int I_{[\hat{a}_n,a]}(t-\hat{\theta}_n z)\delta \log \hat{F}_n(t-\hat{\theta}_n z) d\mathcal{P}_n(y)$$

$$\le \log \hat{F}_n(a) \int 1_{[\hat{a}_n,a]}(t-\hat{\theta}_n z)\delta\, d\boldsymbol{P}_n(y)$$

while the S.L.L.N asserts that

$$\text{2.3)}\quad l_n(\xi_0) \xrightarrow{as-P_0} E_J \,\{\, F_0(T_0)\log[F_0(T_0)] + [F_0(U_0)-F_0(T_0)]\log[F_0(U_0)-F_0(T_0)] + \\ + \bar{F}_0(U_0)\log[\bar{F}_0(U_0)] \,\} \equiv C.$$

Now using the convention $0\cdot(\pm\infty)=0$, assumption *(iv)* and the relation $x\log x \ge -\dfrac{1}{e}$

for $x \in (0,1)$, we obtain

$$\text{(2.4)}\qquad \log \hat{F}_n(a) \int \mathbf{1}_{[\hat{a}_n,a]}(t-\hat{\theta}_n z)\delta\, d\boldsymbol{P}_n(y) \ge -2 \quad \text{for large n, as-}P_0.$$

Similarly

$$\text{(2.5)}\qquad \log\left[1-\hat{F}_n(b)\right] \int \mathbf{1}_{[b,\hat{b}_n]}(u-\hat{\theta}_n z)(1-\delta-\gamma)\, d\boldsymbol{P}_n(y) \ge -2 \quad \text{for large n, as-}P_0.$$

Now notice that the sequences $f_n(y) = 1_{[\hat{a}_n,a]}(t-\hat{\theta}_n z)$, $g_n(y) = 1_{[b,\hat{b}_n]}(u-\hat{\theta}_n z)$ with

$y=(t,u,z)$, belong to a VC-subgraph, class of functions. Apply the Generalized

Glivenko-Cantelli Theorem (G.G.C.T) for half-spaces to obtain

$$\text{(2.6)}\qquad (\boldsymbol{P}_n - P_0)\mathbf{1}_{[\hat{a}_n,a]}(t-\hat{\theta}_n z) \xrightarrow{as} 0,\quad (\boldsymbol{P}_n - P_0)\mathbf{1}_{[b,\hat{b}_n]}(u-\hat{\theta}_n z) \xrightarrow{as} 0 \quad \text{and}$$

$$(\boldsymbol{P}_n - P_0)\mathbf{1}_{[b,\hat{b}_n]}(u-\hat{\theta}_n z)\mathbf{1}_{[a,\hat{a}_n]}(t-\hat{\theta}_n z) \xrightarrow{as} 0.$$

Bounded convergence theorem implies,

$$(2.7)\quad P_0\,\gamma\,\mathbf{1}_{[b,\hat{b}_n]}(u-\hat{\theta}_n z)\mathbf{1}_{[a_*,\hat{a}_n]}(t-\hat{\theta}_n z) \xrightarrow[n]{} P_0\,\gamma\,\mathbf{1}_{[b,b_*]}(u-\theta_* z)\mathbf{1}_{[a_*,a_*]}(t-\theta_* z)$$

$$P_0(1-\gamma-\delta)\mathbf{1}_{[b,\hat{b}_n]}(u-\hat{\theta}_n z) \xrightarrow[n]{} P_0\,(1-\gamma-\delta)\mathbf{1}_{[b,b_*]}(u-\theta_* z) = E_J F_0(U_0)\mathbf{1}_{[b,b_*]}(U_*) > 0$$

and $P_0\,\delta\mathbf{1}_{[a_*,\hat{a}_n]}(t-\hat{\theta}_n z) \xrightarrow[n]{} P_0\,\delta\mathbf{1}_{[a_*,a_*]}(t-\theta_* z) = E_J F_*(T_0)\mathbf{1}_{[a_*,a_*]}(T_*) > 0$.

This is true by assumption *(iv)* of our theorem, the definition of a_*, b_* and the fact that $a_* < a < b < b_*$. Notice the use of some extra notation, $T_* \equiv T - \theta_* Z$, $U_* \equiv U - \theta_* Z$.

The combination of (2.4) and (2.5) with (2.6) and (2.7) proves (2.1).

Now let's consider the following empirical processes

$$\boldsymbol{B}_n^1 = \int \mathbf{1}_{[a,b]}(t-\hat{\theta}_n z)\delta \log \hat{F}_n(t-\hat{\theta}_n z) d\boldsymbol{P}_n(y) \;,$$

$$\boldsymbol{B}_n^2 = \int \mathbf{1}_{[a,b]}(t-\hat{\theta}_n z)\mathbf{1}_{[a,b]}(u-\hat{\theta}_n z)\,\gamma\, \log\left[\hat{F}_n(u-\hat{\theta}_n z) - \hat{F}_n(t-\hat{\theta}_n z)\right] d\boldsymbol{P}_n(y)$$

$$\boldsymbol{B}_n^3 = \int \mathbf{1}_{[a,b]}(u-\hat{\theta}_n z)(1-\delta-\gamma)\log\left[1-\hat{F}_n(u-\hat{\theta}_n z)\right] d\boldsymbol{P}_n(y) .$$

Denote by $\mathbf{B}_*^1, \mathbf{B}_*^2, \mathbf{B}_*^3$ the corresponding integrals obtained by substituting $\hat{\theta}_n \leftrightarrow \theta_*$, $\hat{F}_n \leftrightarrow F_*$, $\boldsymbol{P}_n \leftrightarrow P_0$. The combination of G.G.C.T, (2.0), (2.1) and bounded convergence theorem gives

$$\mathbf{B}_n^j \xrightarrow{as} \mathbf{B}_*^j \quad j = 1,2,3 .$$

Although this convergence is straightforward for $j = 1,3$, it deserves a closer look when j=2. We need to check how the G.G.C.T. applies. Notice that if $\mathbf{1}_{[a,b]}(t-\hat{\theta}_n z)\mathbf{1}_{[a,b]}(u-\hat{\theta}_n z) = 0$ eventually, then $\boldsymbol{B}_n^2 \xrightarrow{as} \boldsymbol{B}_*^2 \equiv 0$. However if $\mathbf{1}_{[a,b]}(t-\hat{\theta}_n z)\mathbf{1}_{[a,b]}(u-\hat{\theta}_n z) = 1$ eventually, then $\boldsymbol{B}_n^2$ as a sum over the "relevant set" of observations $J_n^{\hat{\theta}_n}$, involves only terms for which $\log\left[\hat{F}_n(\eta_i^{\hat{\theta}_n}) - \hat{F}_n(\eta_j^{\hat{\theta}_n})\right] > -\infty$ for some $i > j$, (REMARK 2.6.3 and 2.6.4). Therefore in this case, $\hat{F}_n(b) > \hat{F}_n(a)$ since $[a,b]$ contains at least one element from $J_n^{\hat{\theta}_n}$, eventually. It follows that

$$\boldsymbol{B}_n^2 \leq \log(M_2 - M_1)\int \mathbf{1}_{[a,b]}(t-\hat{\theta}_n z)\mathbf{1}_{[a,b]}(u-\hat{\theta}_n z)\,\gamma\;\; d(\boldsymbol{P}_n - P_0)(y)$$

$$+\int \mathbf{1}_{[a,b]}(t-\hat{\theta}_n z)\mathbf{1}_{[a,b]}(u-\hat{\theta}_n z)\,\gamma\,\log\left[\hat{F}_n(u-\hat{\theta}_n z) - \hat{F}_n(t-\hat{\theta}_n z)\right] dP_0(y),$$

which implies that $\boldsymbol{B}_n^2 \xrightarrow{as} \boldsymbol{B}_\bullet^2$.

Since $l_n(\xi_0) \leq l_n(\hat{\xi}_n) \leq \sum_{j=1}^{3} \boldsymbol{B}_n^j \quad \forall n, \quad \sum_{j=1}^{3} \boldsymbol{B}_n^j \xrightarrow{as} \sum_{j=1}^{3} \boldsymbol{B}_\bullet^j, \quad l_n(\xi_0) \xrightarrow{as} C$, it follows that $\sum_{j=1}^{3} \boldsymbol{B}_\bullet^j \geq C \quad \forall\, a_\bullet < a < b < b_\bullet$. Taking $a \downarrow a_\bullet$, $b \uparrow b_\bullet$ in the last relation, we obtain

$E_0 \log \dfrac{p_{\xi_\bullet}(Y)}{p_{\xi_0}(Y)} \equiv -K\left(P_0, P_{\xi_\bullet}\right) \geq 0$, with $K(P,Q) \geq 0$ denoting the Kullback-Leibler distance between two measures. So $K\left(P_0, P_{\xi_\bullet}\right) = 0$ which in turn implies $P_0 \equiv P_{\xi_\bullet}$.

Therefore $F_\bullet(t_\bullet) = F_0(t_0)$ and $F_\bullet(u_\bullet) = F_0(u_0)$ a.e. J, while assumption *(iii)* gives $\theta_\bullet \equiv \theta_0$ and $F_\bullet \equiv F_0$ a.e V. This concludes our proof.

Chapter 4

INFORMATION THEORY

1. Efficient Scores and Information Bounds: In this chapter we calculate the efficient score and the information lower bound for the finite dimensional parameter in the Cox and the linear regression models with interval censoring. This calculation is of imperative importance since it asseses the degree of difficulty of the estimation problem. Positive information is a necessary condition for existence of efficient estimators. In most parametric and semiparametric models including the Cox model with right censoring , estimators of the finite dimensional parameter were 'easily' obtained at the usual $\sqrt{n}$ rate - see the excellent treatments by Fleming and Harrington (1991), Andersen, Borgan, Keiding and Gill (1993) - However it is not obvious that results obtained under the familiar right censoring scheme, could be extended to interval censoring models in a straightforward manner. The problem is that the infinite dimensional parameter in the interval censoring model cannot be estimated at the usual $\sqrt{n}$-rate. Although in the current context we consider the infinite dimensional parameter as 'nuisance', its influence in the overall estimation problem remains strong. The information calculation sheds some light in the prospects of efficient estimation and in the following theorems we prove that the information for θ in the two models under consideration is indeed positive under reasonable assumptions. For many interesting results and more details on the estimation of the finite dimensional parameter in Semiparametric models, we found the recent monograph by Bickel, Klaasen, Ritov and Wellner (1994) to be a valuable source.

Before we proceed to the main results in this chapter, we state a few definitions and auxilliary results that will be used in the following sections.

2. Preliminaries: Let $\Theta \subset R^d$ be an open set. We call *model*, a family of probability measures $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ on a measurable space $(\mathcal{X}, \mathcal{U})$ and *experiment* the triplet $\mathcal{E} = \langle \mathcal{X}, \mathcal{U}, \mathcal{P} \rangle$. Let μ be a measure on $(\mathcal{X}, \mathcal{U})$ and suppose that the family $\mathcal{P}$ is dominated with μ-densities $p_\theta(\cdot)$. The model $\mathcal{P}$ is said to be *Hellinger-* $L_2(\mu)$ *- differentiable* at $\theta \in \Theta$ if there is a function $\dot{l}_\theta : \mathcal{X} \to R^d$, $\int |\dot{l}_\theta(x)|^2 dP_\theta(x) < \infty$ such that

$$\int \left(\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h' \dot{l}_\theta \sqrt{p_\theta} \right)^2 d\mu = o(|h|^2), \quad h \to 0 .$$

The *Fisher information matrix* is defined as $I(\theta) = \int \dot{l}_\theta \dot{l}_\theta' dP_\theta$. We say that $\mathcal{E}$ is *regular* in Θ, if $\mathcal{P}$ is: (i) dominated with μ-densities $p_\theta(\cdot)$, (ii) Hellinger differentiable with derivative $\dot{l}_\theta$, (iii) the function $\theta \mapsto p_\theta(x)$ is differentiable for μ-a.e. x.

Let's fix a measure $P_0 \in \mathcal{P}$ and let $a \in L_2^0(P_0) = \left\{ h \in L_2(P_0) : \int h dP_0 = 0 \right\}$ be an arbitrary function. Without loss of generality we might assume that a is bounded.

Otherwise for some $M>0$ we can work with $a_M(x) = a(x)\mathbf{1}_{\{|a|\leq M\}} - \int a(x)\mathbf{1}_{\{|a|\leq M\}} dP_0(x)$.

For Θ one dimensional, consider the parametric family,

$$f_\theta(x) = p_0(x)\exp(\theta a(x) - b(\theta)),$$

with $b(\theta) = \log\left[\int e^{\theta a(x)} dP_0(x)\right] < \infty \quad \forall \theta \in \Theta$, since a is a bounded function. Notice that $\{f_\theta : \theta \in \Theta\}$ is an exponential family -thus regular- passing thru p_0 with score function at $\theta = 0$ given by $a(x) \equiv \frac{\partial}{\partial\theta} \log f_\theta(x)\big|_{\theta=0}$. In this way we establish an association between $L_2^0(P_0)$ and regular parametric models. This association plays a fundamental role in the computation of efficient scores and influence functions for semi or non parametric models.

3. Information lower bound: In this section we establish the conditions under which the information for the finite dimensional parameter in the Cox and the linear regression models is positive. In addition to the notation introduced in earlier chapters, we add some more here to accomodate the needs of the current section. We start with the PH model.

$$\varphi(t,u) = \frac{E\{Ze^{2\theta' Z}O_t(Z) | T=t, U=u\}}{E\{e^{2\theta' Z}O_t(Z) | T=t, U=u\}}$$

$$O_t(z) = \frac{\bar{F}(t|z)}{1-\bar{F}(t|z)} \quad O_u(z) = \frac{\bar{F}(u|z)}{1-\bar{F}(u|z)} \quad R(z) = \frac{O_t(z)}{O_t(z) - O_u(z)}$$

$$\psi(t,u)=\frac{E\{Ze^{2\theta'Z}R(Z)O_u(Z)|\,T=t,U=u\}}{E\{e^{2\theta'Z}R(Z)O_u(Z)|\,T=t,U=u\}}$$

$$A_{t,u}(z)=R(z)(1+O_u(z))\quad Q_1(y)=e^{\theta'z}\{\delta O_t(z)-\gamma A_{t,u}(z)\}\quad Q_2(y)=e^{\theta'z}\{\gamma A_{t,u}(z)-(1-\delta)\}.$$

THEOREM 1. *Suppose that*

(i) $\forall\ (t,u)\in S_H \quad 0\le a_0<\alpha<t<u<\beta<b_0$ *for some* $\alpha<\beta$.

(ii) $\exists\ M>0\ :\ \Pr\{|Z|\le M\}=1$.

Then

(a) The efficient score function for θ *is*

$$\dot{l}_\theta^*(y)=[z-\varphi(t,u)][\Lambda(t)Q_1(y)+\Lambda(u)Q_2(y)]+[\varphi(t,u)-\psi(t,u)][\Lambda(u)-\Lambda(t)]Q_2(y).$$

(b) The information for θ *is*

$$I(\theta)=E\left[\dot{l}_\theta^*\right]^{\otimes 2}=$$

$$E_J(Z-\psi)^{\otimes 2}\left[\Lambda(U|Z)-\Lambda(T|Z)\right]^2 O_U(Z)R(Z)+E_J(Z-\varphi)^{\otimes 2}\Lambda(T|Z)^2O_T(Z)$$

which is positive definite, unless the distribution of Z *is degenerate at some point* z_0.

REMARK 1. With $\Pr\{T=U\}=1$ and the convention $0\times\infty=0$, our Theorem 1.2 gives Theorem 2.3 of Huang and Wellner (1993). In their paper, these two authors provide a survey of regression models, under a censoring scheme consisting of a single inspection, (a situation that Groeneboom and Wellner (1992) call *case 1* interval censoring).

Proof of Theorem 1. Without loss of generality we might assume that d=1. Let $y=(\delta,\gamma,t,u,z)\in\{0,1\}^2\times R_+^2\times R$, be fixed throughout. Without causing confusion, we do not exhibit the argument when using the functions φ, ψ, O_t, O_u, R, A, Q_1, Q_2. The log-likelihood function up to an additive constant is

$$l(\theta,\Lambda;y)=\delta\log\left(1-e^{-\Lambda(t)e^{\theta z}}\right)+\gamma\log\left(e^{-\Lambda(t)e^{\theta z}}-e^{-\Lambda(u)e^{\theta z}}\right)-(1-\gamma-\delta)\Lambda(u)e^{\theta z}.$$

The derivative of $l(\theta,\Lambda;y)$ wrt θ gives the score function for θ,

$$\dot{l}_\theta(y)=z\Lambda(t)Q_1+z\Lambda(u)Q_2.$$

Fix a distribution F with Lebesgue density f and for $a\in L_2^0(F)$ consider a regular parametric family $\mathcal{H}=\{f_\eta:|\eta|<1\}\subset\mathcal{F}_t=\{f\geq 0:\int f d\tau=1\}$, passing through f. Then

$$a(t)=\frac{\partial}{\partial\eta}\log f_\eta(t)\Big|_{\eta=0}\quad\text{and}\quad\frac{\partial}{\partial\eta}\bar{F}_\eta(t)\Big|_{\eta=0}=\int_t^\infty adF.$$

The score operator for f is obtained as the directional derivative of the function $\eta\to l(\theta,F_\eta;\cdot)$ at $\eta=0$. It is given by

$$(1.1)\quad \dot{l}_f\,a(y)=-\delta\frac{O_t e^{\theta z}\int_t^\infty adF}{\bar{F}(t)}+\gamma\left(\frac{\mathrm{A}e^{\theta z}\int_t^\infty adF}{\bar{F}(t)}-\frac{(A-1)e^{\theta z}\int_u^\infty adF}{\bar{F}(u)}\right)+(1-\gamma-\delta)\frac{e^{\theta z}\int_u^\infty adF}{\bar{F}(u)}$$

$$=-Q_1\frac{\int_t^\infty adF}{\bar{F}(t)}-Q_2\frac{\int_u^\infty adF}{\bar{F}(u)}.$$

Moreover the assumptions of the theorem guarantee that both $\dot{l}_\theta$, $\dot{l}_f a$ are P_0-square integrable.

The efficient score function for θ is defined by

$$(1.2)\qquad \dot{l}_\theta^*(y)=\dot{l}_\theta(y)-\dot{l}_f\, a_*(y)$$

with $\dot{l}_f\, a_*$ being the projection of $\dot{l}_\theta$ onto the closure of the space spanned by $\{\dot{l}_f a : a\in L_2^0(F)\}$.

Let $V_1(a_*)=z\Lambda(t)+\dfrac{\int_t^\infty a_* dF}{\bar{F}(t)}$, $\mathrm{V}_2(a_*)=z\Lambda(u)+\dfrac{\int_u^\infty a_* dF}{\bar{F}(u)}$.

Then

$$-\left(\dot{l}_\theta(y)-\dot{l}_f\, a_*(y)\right)\dot{l}_f\, a(y)=\frac{\int_t^\infty adF}{\bar{F}(t)}\left[Q_1^2V_1(a_*)+Q_1Q_2V_2(a_*)\right]+\frac{\int_u^\infty adF}{\bar{F}(u)}\left[Q_2^2V_2(a_*)+Q_1Q_2V_1(a_*)\right].$$

Notice that

$$E\left(Q_1^2|T,U,Z\right)=e^{2\theta Z}O_TR,\quad E\left(Q_1Q_2|T,U,Z\right)=-e^{2\theta Z}O_UR,\quad E\left(Q_2^2|T,U,Z\right)=e^{2\theta Z}O_UR.$$

Thus

$$-E\left(\dot{l}_\theta(Y)-\dot{l}_f\, a_*(Y)\right)\dot{l}_f\, a(Y)=E_H\left\{\frac{\int_T^\infty adF}{\bar{F}(T)}E\left[e^{2\theta Z}R\left(O_T\, V_1(a_*)-O_U\, V_2(a_*)\right)|T,U\right]\right\}+$$
$$E_H\left\{\frac{\int_U^\infty adF}{\bar{F}(U)}E\left[e^{2\theta Z}RO_U\left(V_2(a_*)-V_1(a_*)\right)|T,U\right]\right\}$$

We want to have $\quad E\left(\dot{l}_\theta(y)-\dot{l}_f\, a_*(y)\right)\dot{l}_f\, a(y)=0\quad \forall a\in L_2^0(F)$.

Let
$$r_1(t,u)=E\left[e^{2\theta Z}R\left(O_t\, V_1(a_*)-O_u\, V_2(a_*)\right)|T=t,U=u\right]$$
$$r_2(t,u)=E\left[e^{2\theta Z}RO_u\left(V_2(a_*)-V_1(a_*)\right)|T=t,U=u\right]$$

Set $r_1(t,u)=r_2(t,u)=0$.

It follows that

$$(1.3)\qquad g_1\frac{\int_t^{\infty} a_* dF}{\bar{F}(t)} - g_2\frac{\int_u^{\infty} a_* dF}{\bar{F}(u)} = f_2\Lambda(u) - f_1\Lambda(t)$$

$$(1.4)\qquad g_2\left[\frac{\int_u^{\infty} a_* dF}{\bar{F}(u)} - \frac{\int_t^{\infty} a_* dF}{\bar{F}(t)}\right] = f_2[\Lambda(t) - \Lambda(u)],$$

with $f_1(t,u) = E\{Ze^{2\theta Z}RO_t|T=t,U=u\}$ $g_1(t,u) = E\{e^{2\theta Z}RO_t|T=t,U=u\}$

$f_2(t,u) = E\{Ze^{2\theta Z}RO_u|T=t,U=u\}$ $g_2(t,u) = E\{e^{2\theta Z}RO_u|T=t,U=u\}$.

Solving (1.3) and (1.4) we obtain

$$(1.5)\qquad \int_t^{\infty} a_* dF = -\varphi\Lambda(t)\bar{F}(t) \text{ and } \int_u^{\infty} a_* dF = [(\psi-\varphi)\Lambda(t) - \psi\Lambda(u)]\bar{F}(u).$$

Now by assumption (*ii*) we can easily verify that $\left|\int_t^{\infty} a_* dF\right| \le M\Lambda(t)\bar{F}(t) \to 0$ as $t \to 0$.

Moreover $\int_t^{\infty} a_* dF \to \int_0^{\infty} a_* dF$ *as* $t \to 0$. Thus $a_* \in L_2^0(F)$. Now we can use (1.1) and (1.5) to obtain from (1.2) the efficient score function for θ,

$$\dot{l}_\theta^*(y) = \dot{l}_\theta(y) - \dot{l}_f\, a_*(y)$$

$$=(z-\varphi)[\Lambda(t)Q_1 + \Lambda(u)Q_2] + (\varphi-\psi)[\Lambda(u) - \Lambda(t)]Q_2.$$

Finally the information for θ is

$$E\dot{l}_\theta^{*2}(Y) = E_G(Z-\varphi)^2\left[\Lambda(T|Z)^2 O_T R + \Lambda(U|Z)^2 O_U R - 2\Lambda(T|Z)\Lambda(U|Z)O_U R\right] +$$

$$E_G(\varphi-\psi)^2[\Lambda(U|Z) - \Lambda(T|Z)]^2 O_U R +$$

$$2E_G(Z-\varphi)(\varphi-\psi)[\Lambda(U|Z) - \Lambda(T|Z)]^2 O_U R =$$

$$E_G(Z-\psi)^2[\Lambda(U|Z) - \Lambda(T|Z)]^2 O_U R + E_G(Z-\varphi)^2[\Lambda(T|Z)]^2 O_T.$$

The theorem is now proved.

Before we turn to the linear regression model, we add here some more notations.

$$Q_1(y)=\frac{\delta}{F(t-\theta z)}-\frac{\gamma}{F(u-\theta z)-F(t-\theta z)}\ ,\quad Q_2(y)=\frac{\gamma}{F(u-\theta z)-F(t-\theta z)}-\frac{1-\delta-\gamma}{\overline{F}(u-\theta z)}$$

$$V(\cdot,z)=zf(\cdot-\theta z)+\int_{-\infty}^{\cdot-\theta z}a_* dF\ ,\quad k(r,s)=E\left(Z\middle|T-\theta Z=r,\ U-\theta Z=s\right)$$

$$A_1(r,s)=\frac{1}{F(r)}+\frac{1}{F(s)-F(r)}\ ,\ A_{12}(r,s)=\frac{1}{F(s)-F(r)}\ ,\ A_2(r,s)=\frac{1}{F(s)-F(r)}+\frac{1}{\overline{F}(s)}\ ,\ r<s$$

$$B_1(r)=\frac{1}{F(r)}+\frac{1}{\overline{F}(r)}\quad ,\quad B_2(r,s)=\frac{1}{F(s)-F(r)}-\frac{1}{\overline{F}(r)}\,.$$

For a fixed distribution function F and $\theta\in\Theta$, let

$$a_0=\sup\{x\colon F(x)=0\},\quad b_0=\inf\{x\colon F(x)=1\},\quad (T,U,Z)\sim J,\quad (T-\theta Z,U-\theta Z)\sim \mathcal{L}.$$

THEOREM 2. *Suppose that*

(i) *F is a strictly increasing distribution with Lebesgue density f satisfying*

$$f(s)\xrightarrow[s\to\pm\infty]{}0.$$

(ii) $\exists\ M>0\ :\ \Pr\{|Z|\le M\}=1$.

(iii) $\forall\ (t,u)\in S_{\mathcal{L}}\quad 0\le a_0<\alpha<t<u<\beta<b_0$ *for some* $\alpha<\beta$.

Then

(a) *The efficient score function* for θ *is*

$$l_\theta^*(y)=-[z-k(t-\theta z,u-\theta z)][Q_1(y)f(t-\theta z)+Q_2(y)f(u-\theta z)].$$

(b) *The information* for θ *is*

$$I(\theta)=E\left[l_\theta^*(Y)\right]^{\otimes 2}=E_J\left[\left[Z-E\left(Z\middle|T_\#,U_\#\right)\right]^{\otimes 2}\left[f^2(T_\#)B_1+\left\{f(U_\#)\sqrt{A_2}-f(T_\#)\sqrt{B_2}\right\}^2\right]\right],$$

with $T_\#=T-\theta Z,\ U_\#=U-\theta Z,\ A_j\equiv A_j(T_\#,U_\#)\quad\forall j$.

which is positive definite, unless $Z=E\left(Z\middle|T_\#,U_\#\right)$ a.e. $\mathcal{L}$.

Proof of Theorem 2. We follow the pattern established in Theorem 1. Let d=1. Arguing as in section 2, for a given (bounded) function $a \in L_2^0(F)$, we can construct a regular parametric family $\mathcal{H}$ passing through f. Let

$\mathcal{H} = \{f_\eta : |\eta| < 1\} \subset \mathcal{F}_t = \{f \geq 0 : \int f d\tau = 1\}$. We obtain $a(t) = \frac{\partial}{\partial\eta} \log f_\eta(t)\big|_{\eta=0}$,

$\frac{\partial}{\partial\eta} \overline{F}_\eta(t)\big|_{\eta=0} = \int_t^\infty adF$ and since $a \in L_2^0(F)$ we also have $\frac{\partial}{\partial\eta} F_\eta(t)\big|_{\eta=0} = \int_{-\infty}^{t} adF$.

The log-likelihood function up to an additive constant is

$$l(\theta, F; y) = \delta \log F(t-\theta z) + \gamma \log[F(u-\theta z) - F(t-\theta z)] + (1-\delta-\gamma)\log \overline{F}(u-\theta z)$$

and the associated score function for θ,

$$\dot{l}_\theta(y) = -zf(t-\theta z)Q_1(y) - zf(u-\theta z)Q_2(y).$$

The score operator for f is obtained as the directional derivative of the function

$$\eta \to l(\theta, F_\eta; \cdot) \text{ at } \eta = 0.$$

(2.1) $$i_f a(y) \equiv E_f(a(X) | Y = y)$$

$$= \delta \frac{\int_{-\infty}^{t-\theta z} adF}{F(t-\theta z)} + \gamma \frac{\int_{t-\theta z}^{u-\theta z} adF}{F(u-\theta z) - F(t-\theta z)} + (1-\delta-\gamma)\frac{\int_{u-\theta z}^{\infty} adF}{\overline{F}(u-\theta z)}$$

$$= Q_1(y)\int_{-\infty}^{t-\theta z} adF + Q_2(y)\int_{-\infty}^{u-\theta z} adF = \frac{\partial}{\partial\eta} l(\theta, F_\eta; y)\big|_{\eta=0}.$$

Then

$$-(\dot{l}_\theta(y) - i_f a_*(y))i_f a(y) = [Q_1^2(y)V(t,z) + Q_1(y)Q_2(y)V(u,z)]\int_{-\infty}^{t-\theta z} adF$$

$$+\left[Q_2^2(y)V(u,z)+Q_1(y)Q_2(y)V(t,z)\right]\int_{-\infty}^{u-\theta z} adF.$$

Easy computations show that

$$E\{Q_1^2(Y)|T,U,Z\}=A_1(T_\#,U_\#)\,,\ E\{Q_2^2(Y)|T,U,Z\}=A_2(T_\#,U_\#)$$

$$\text{and}\quad E\{Q_1(Y)Q_2(Y)|T,U,Z\}=-A_{12}(T_\#,U_\#).$$

Thus

$$-E\left(\dot{l}_\theta-i_f a_\bullet\right)i_f a=E_J\int_{-\infty}^{T-\theta Z} adF\left[V(T,Z)A_1-V(U,Z)A_{12}\right]+E_J\int_{-\infty}^{U-\theta Z} adF\left[V(U,Z)A_2-V(T,Z)A_{12}\right]$$

$$=E_{\mathcal{L}}\int_{-\infty}^{r} adF\left[A_1(r,s)E\left\{V(T,Z)\middle|\begin{matrix}T-\theta Z=r\\U-\theta Z=s\end{matrix}\right\}-A_{12}(r,s)E\left\{V(U,Z)\middle|\begin{matrix}T-\theta Z=r\\U-\theta Z=s\end{matrix}\right\}\right]$$

$$+\ E_{\mathcal{L}}\int_{-\infty}^{s} adF\left[A_2(r,s)E\left\{V(U,Z)\middle|\begin{matrix}T-\theta Z=r\\U-\theta Z=s\end{matrix}\right\}-A_{12}(r,s)E\left\{V(T,Z)\middle|\begin{matrix}T-\theta Z=r\\U-\theta Z=s\end{matrix}\right\}\right],$$

with $\mathcal{L}$ the joint distribution of $(T-\theta Z,\, U-\theta Z)$. To satisfy the orthogonality condition

$$E\left(\dot{l}_\theta-i_f a_\bullet\right)i_f a=0\quad \forall a\in L_2^0(F),$$

set the two brackets in the preceding equation equal to zero and solve the system of equations to obtain

$$\text{(2.2)}\qquad \int_{-\infty}^{r} a_\bullet dF=-f(r)k(r,s)\ \text{ and }\ \int_{-\infty}^{u} a_\bullet dF=-f(u)k(r,s).$$

From *(i)* and *(ii)* we obtain

$\left|\int_{-\infty}^{r} a_\bullet dF\right|\le Mf(r)\xrightarrow[r\to\infty]{}0$. Since $\int_{-\infty}^{r} a_\bullet dF\to\int_{-\infty}^{\infty} a_\bullet dF$ *as* $r\to\infty$, we conclude that

$a_\bullet\in L_2^0(F)$. Now from (2.1) and (2.2), the efficient score function for θ is

$$\dot{l}^{*}_{\theta}(y) = \dot{l}_{\theta}(y) - i_f a_*(y)$$

$$= -[z - k(t - \theta z, u - \theta z)][Q_1(y)f(t - \theta z) + Q_2(y)f(u - \theta z)]$$

and the information at θ,

$$I(\theta) = E[\dot{l}^{*}_{\theta}(Y)]^2 = E_J\Big[[Z - E(Z|T_\#, U_\#)]^2[f^2(T_\#)A_1 + f^2(U_\#)A_2 - 2f(T_\#)f(U_\#)A_{12}]\Big],$$

with $T_\# = T - \theta Z$, $U_\# = U - \theta Z$, $A_j \equiv A_j(T_\#, U_\#)$ $\forall j$.

Notice that strict monotonicity of F and assumption *(iii)* imply that $t_\# \to O(t_\#)$ is a strictly decreasing function. Moreover

$$f^2(T_\#)B_1 + \{f(U_\#)\sqrt{A_2} - f(T_\#)\sqrt{B_2}\}^2 = f^2(T_\#)A_1 + f^2(U_\#)A_2 - 2f(T_\#)f(U_\#)A_{12} > 0,$$

if and only if $O(t_\#) > O(u_\#)$ for $t_\# < u_\#$. This shows that the information for θ is positive unless $Z = E(Z / T_\#, U_\#)$ a.e. ℓ. The theorem is now proved.

REMARK 2. We want to emphasize here that both theorems that we have considered have immediate generalizations to situations where interval censoring is a result of an inspection process with finitely many observations-inspections. It will be a matter of future research the case of a random number of inspections as it is described in Wang et.al (1994).

4. Estimation of $I(\theta_0)$**:** In this section we provide an estimator of $I(\theta_0)$ in both models that we have analyzed in this thesis. This estimator provides at least an approximation of the lower bound for the variance of $\hat{\theta}_n$ in the Neonatal Brain

Hemorrhage application that we consider in Chapter 6. Let $(\hat{\theta}_n, \hat{\Lambda}_n)$, $(\hat{\theta}_n, \hat{F}_n)$ denote MLE's of (θ, Λ) and (θ, F) in the two models respectively. We make use of the notation introduced in previous sections with obvious modifications when the MLE replaces the true parameter. To make sure that there is no confusion about the model to which we refer, we denote by $\hat{I}_n^C$ and $\hat{I}_n^L$ the estimates of the information in the Cox and the Linear regression model respectively. The major difficulty that we encounter here is the estimation of the conditional expectation that appears in the expression for $I(\theta_0)$. The following special cases are particularly useful in the Cox model due to its special structure and the form of the information measure. No simple expression for $I(\theta_0)$ is available in the linear regression model.

i) <u>*Z independent of (T,U)*</u>.

Then

$$Z - \psi(t,u) = Z - \frac{E_W\{Ze^{2\theta Z} R_{t,u}(Z) O_u(Z)\}}{E_W\{e^{2\theta Z} R_{t,u}(Z) O_u(Z)\}} \quad \text{and} \quad Z - \varphi(t,u) = Z - \frac{E_W\{Ze^{2\theta Z} O_t(Z)\}}{E_W\{e^{2\theta Z} O_t(Z)\}}.$$

Let $\hat{\psi}_{n,i} = \dfrac{\sum_{j=1}^n Z_j e^{2\hat{\theta}_n Z_j} \hat{R}_{T_i,U_i}(Z_j) \hat{O}_{n,U_i}(Z_j)}{\sum_{j=1}^n e^{2\hat{\theta}_n Z_j} \hat{R}_{T_i,U_i}(Z_j) \hat{O}_{n,U_i}(Z_j)}$ and $\hat{\varphi}_{n,i} = \dfrac{\sum_{j=1}^n Z_j e^{2\hat{\theta}_n Z_j} \hat{O}_{n,T_i}(Z_j)}{\sum_{j=1}^n e^{2\hat{\theta}_n Z_j} \hat{O}_{n,T_i}(Z_j)}$.

Now we can estimate $I^C(\theta_0)$ by

$$(4.1) \qquad \hat{I}_n^C = \frac{1}{n}\sum_{i=1}^n (Z_i - \hat{\psi}_{n,i})^2 \left[\hat{\Lambda}_n(U_i|Z_i) - \hat{\Lambda}_n(T_i|Z_i)\right]^2 \hat{O}_{n,U_i}(Z_i)\hat{R}_i(Z_i) +$$
$$\frac{1}{n}\sum_{i=1}^n (Z_i - \hat{\varphi}_{n,i})^2 \left[\hat{\Lambda}_n(T_i|Z_i)\right]^2 \hat{O}_{n,T_i}(Z_i).$$

ii) ***Suppose that Pr{Z=1}=1-a and Pr{Z=0}=a***

Let $\Gamma_{t,u}(Z)=e^{2\theta Z}R_{t,u}(Z)O_u(Z)$ and $\Delta_t(Z)=e^{2\theta Z}O_t(Z)$.

Then

$$E_W\{\Delta_T(Z)|T,U\}=\Delta_T(0)\Pr\{Z=0|T,U\}+\Delta_T(1)\Pr\{Z=1|T,U\}$$

$$=\Delta_T(0)\frac{ah_0(T,U)}{h(T,U)}+\Delta_T(1)\frac{(1-a)h_1(T,U)}{h(T,U)}\ ,$$

$$E_W\{Z\Delta_T(Z)|T,U\}=\Delta_T(1)\frac{(1-a)h_1(T,U)}{h(T,U)}\ .$$

Similarly

$$E_W\{\Gamma_{T,U}(Z)|T,U\}=\Gamma_{T,U}(0)\frac{ah_0(T,U)}{h(T,U)}+\Gamma_{T,U}(1)\frac{(1-a)h_1(T,U)}{h(T,U)}$$

and
$$E_W\{Z\Gamma_{T,U}(Z)|T,U\}=\Gamma_{T,U}(1)\frac{(1-a)h_1(T,U)}{h(T,U)}$$

with $h_i(t,u)=h(t,u|Z=i)$ $i=0,1$. Now we can estimate φ,ψ by

$$\tilde{\varphi}_n(t,u)=\frac{(1-a)\hat{\Delta}_n(1)h_{n,1}(t,u)}{a\hat{\Delta}_n(0)h_{n,0}(t,u)+(1-a)\hat{\Delta}_n(1)h_{n,1}(t,u)}$$

and

$$\tilde{\psi}_n(t,u)=\frac{(1-a)\hat{\Gamma}_n(1)h_{n,1}(t,u)}{a\hat{\Gamma}_n(0)h_{n,0}(t,u)+(1-a)\hat{\Gamma}_n(1)h_{n,1}(t,u)}$$

respectively, where $h_{n,i}(t,u)=h_n(t,u|Z=i)$ $i=0,$ is a kernel type estimator of $h_i(t,u)=h(t,u|Z=i)$ $i=0,1$. By appropriate choice of kernel and bandwidth, $h_{n,i}$ can be a consistent estimator of h_i - see Silverman (1986). An estimate of $I^C(\theta)$ is given by (4.1) using $\tilde{\psi}_n,\tilde{\varphi}_n$.

To estimate $I(\theta)$ in other cases than the previously noted, one has to estimate the conditional expectations that appear in the information for the models that we consider. In addition to that, the linear regression model requires an estimate of the density f. Kernel density estimation might be a solution here, leaving us the burden to choose the kernel and the appropriate bandwidth. In this case the estimator will have the form

$$\hat{f}_n(t) = \frac{1}{h_n}\int K\left(\frac{t-u}{h_n}\right) d\hat{F}_n(u). \tag{4.2}$$

To estimate the conditional expectations that appear in the definition of (Ψ, φ) and k in the Cox and the linear regression model, we proceed in two steps. If the conditional expectation has the form $E\big(\varsigma_\theta(T,U,Z,F) \big| \pi_\theta(T,U)\big) = g\big(\pi_\theta(T,U)\big)$, for some known functions $\varsigma_\theta, \pi_\theta$, then

i) Approximate $\varsigma_\theta(T,U,Z,F)$ by $\varsigma_{\hat{\theta}_n}(T,U,Z,\hat{F}_n)$ and π_θ by $\pi_{\hat{\theta}_n}$.

ii) Employ tools from nonparametric regression to estimate g in

$$\varsigma_{\hat{\theta}_n}(T,U,Z,\hat{F}_n) = g\big(\pi_{\hat{\theta}_n}(T,U)\big) + \varepsilon, \text{ with } E\varepsilon = 0.$$

Call the resulting estimator $\hat{g}_n(T,U)$. Now a meaningful estimator of $I(\theta)$ in the Cox model is given by (4.1), with $\hat{\psi}_n, \hat{\varphi}_n$ functions of $\hat{g}_n(T,U)$. The corresponding estimator in the linear regression is

$$\hat{I}_n^L = \frac{1}{n}\sum_{j=1}^{n}\left[Z_j - \hat{g}_n(T_j,U_j)\right]^2 \left[\hat{f}_n^2(\hat{T}_{j,\#})\hat{A}_1 + \hat{f}_n^2(\hat{U}_{j,\#})\hat{A}_2 - 2\hat{f}_n(\hat{T}_{j,\#})\hat{f}_n(\hat{U}_{j,\#})\hat{A}_{12}\right]. \tag{4.3}$$

5. Comparison of information measures: It is expected that there is significant loss of information in the models that we have considered due to interval censoring. In this section we try to quantify this loss for a number of distributions that the information measure is relatively easy to compute. We do it for the Cox model only, since there is no easy way to compute $I(\theta)$ in the linear regression. Assume that Z is a binary covariate, independent of (T, U) and for $\gamma > 0$, $\Pr\{Z=0\} = 1-\gamma$, $\Pr\{Z=1\} = \gamma$.

For $\Gamma_{t,u}(Z) = e^{2\theta Z} R_{t,u}(Z) O_u(Z)$, $\Delta_t(Z) = e^{2\theta Z} O_t(Z)$ as in the previous section, we obtain

$$\psi(t,u) = \frac{E_Z\{Z\Gamma_{t,u}(Z)\}}{E_Z\{\Gamma_{t,u}(Z)\}} = \frac{e^{2\theta} R_{t,u}(1) O_u(1)\gamma}{e^{2\theta} R_{t,u}(1) O_u(1)\gamma + R_{t,u}(0) O_u(0)(1-\gamma)}$$

and $$\varphi(t,u) = \frac{E_Z\{Z\Delta_t(Z)\}}{E_Z\{\Delta_t(Z)\}} = \frac{\gamma e^{2\theta} O_t(1)}{\gamma e^{2\theta} O_t(1) + (1-\gamma) O_t(0)}.$$

We will also assume that the censoring distribution is discrete uniform on the lattice $\mathcal{L}_6 = \{(i,j) : i < j,\ i,j \in \{1,2,3,4\}\}$. Tables 5.1 and 5.2 contain the information measure for θ, for a few selected distributions under interval censoring and no censoring. In the latter case, Ritov and Wellner (1988) prove that $I^u(\theta) = E_X\{\mathrm{var}(Z|X)\}$. In the current setup, the information for θ without any censoring is given by

$$I^u(\theta) = E_X\{\xi(X) - \xi^2(X)\} \quad \text{with} \quad \xi(x) = \frac{\gamma e^{\theta}[\bar{F}(x)]^{e^{\theta}}}{\gamma e^{\theta}[\bar{F}(x)]^{e^{\theta}} + (1-\gamma)\bar{F}(x)}.$$

Notice that if $X \sim F$, F a continuous distribution, then $F(X) \sim U(0,1)$, thus making the information independent of F. This is indeed the case as the next two tables show. For interval censoring, we present the information measure in case that two inspections are available as well as in case of a single inspection (see Remark 3.1). In the following tables the numbers in parenthesis correspond to case 1 interval censoring.

Table 5.1: Information measure in the Cox model $(\gamma = .2)$

Hazard	θ	**-2**	**-1**	**0**	**1**	**2**
exp(.5)	$I(\theta)$	.037 (.0235)	.077 (.052)	.117 (.086)	.096 (.076)	.031 (.021)
U(0,5)	$I(\theta)$	.028 (.0135)	.060 (.032)	.101 (.061)	.105 (.079)	.061 (.046)
$f(x)=.08x$ $0 \le x \le 5$	$I(\theta)$	.017 (.006)	.039 (.014)	.072 (.030)	.091 (.049)	.065 (.049)
no censoring	$I^{u}(\theta)$	.076	.125	.160	.131	.081

Table 5.2: Information measure in the Cox model $(\gamma = .5)$

Hazard	θ	**-2**	**-1**	**0**	**1**	**2**
exp(.5)	$I(\theta)$	.077 (.050)	.136 (.095)	.183 (.134)	.150 (.119)	.052 (.036)
U(0,5)	$I(\theta)$	.059 (.030)	.110 (.060)	.158 (.096)	.152 (.110)	.082 (.063)
$f(x)=.08x$ $0 \le x \le 5$	$I(\theta)$	.037 (.013)	.073 (.027)	.113 (.046)	.126 (.061)	.080 (.059)
no censoring	$I^{*}(\theta)$	.165	.223	.250	.187	.097

We have chosen the three distributions on the basis of severity in censoring. Clearly an exponential distribution for the variable of interest creates many left censored observations, while the distribution specified by the density f puts most of its mass towards the right endpoint of the interval $[0,5]$, thus causing an excess of right censored observations.

Chapter 5

GENERALIZED M-ESTIMATION IN THE ACCELERATED FAILURE TIME MODEL

In Chapter 2 we defined maximum (profile) likelihood estimators for the Cox and the linear regression models and established sufficient and necessary conditions for step 1 in the profile likelihood algorithm to be well defined. Although in the Cox model the entire algorithm seems to be easily implemented, largely due to the smoothness of the function $\theta \mapsto e^{\theta z}$, the same result might not hold in linear regression (in step 2 we might not get a maximizer) . The problem arises from the unpleasant situation of nonsmoothness in the function $\theta \mapsto F_n(\cdot;\theta)$, with $F_n(\cdot;\theta)$ the maximum likelihood estimator of F_0 for a fixed θ. Any attempt to develop the asymptotic theory for this class of estimators will have to confront problems of this nature.

To avoid artificial assumptions which are often impossible to verify in practice and make the problem unnatural, we turn to a smaller class of M-estimators which hopefully provides asymptotic theory under reasonable assumptions on the underline model. Our efforts follow closely the theory of Asymptotic Generalized M-Estimators (AGME) of Bickel, Klaassen, Ritov, Wellner (1993)- hereafter abbreviated to BKRW. In their master theorem 7.3.1, page 312, these authors establish the conditions for an AGM estimator to be asymptotically normal. However their requirement for a consistent AGM estimator is extremely difficult to verify, especially in models where a close form of such an estimator is not available or very hard to obtain. We manage to obtain the result of their theorem by relaxing the consistency assumption to the expence of more smoothness in the objective function. To verify the rest of the conditions in their theorem, we exploit the modern arsenal in the theory of empirical processes. Sufficient conditions for convergence of stochastic processes and results of the form

$$\sup\left\{\sqrt{n}\,(\mathbf{P}_n - P)f \ : \ f \in \mathcal{F}_n \subset \mathcal{F}\right\} = o_p(1),$$

are given in Pollard (1989) for certain classes of functions $\mathcal{F}$. In the appendix of this thesis we put together the necessary machinery that will enable us to verify the assumptions of our modified version of Theorem 7.3.1 of BKRW. Most of this work will be the subject of future research.

1. The master theorem: Let $\mathcal{P}$ be a model and $\mathcal{P} \subset \mathcal{M}_0$ with $\mathcal{M}_0$ containing all measures with finite support. Let $W_n, W\colon R^m \times \mathcal{M}_0 \to R^m$ and $v\colon \mathcal{P} \to R^m$. Suppose $W(\nu(P), P) = 0 \quad \forall P \in \mathcal{P}$ and $W_n(\nu, P_0) = W(\nu, P_0) + o(1)$. Introduce the notation $\boldsymbol{W}_n(\nu) \equiv W_n(\nu, \boldsymbol{P}_n)$ with $\boldsymbol{P}_n$ denoting empirical distribution. We say that $\hat{\nu}_n$ is *generalized M-estimator* (GME) of $\nu(P)$ if $\boldsymbol{W}_n(\hat{\nu}_n) = 0$ and *asymptotically generalized M-estimator* (AGME) if $\boldsymbol{W}_n(\hat{\nu}_n) = o_p(n^{-1/2})$.

Let $V_n(\nu) \equiv \sqrt{n}\left(\boldsymbol{W}_n(\nu) - W(\nu, P_0)\right)$, $\nu_0 \equiv \nu(P_0)$.

$$\text{(GM0)} \quad \forall \varepsilon_n \downarrow 0 \quad \sup\left\{\frac{|V_n(\nu) - V_n(\nu_0)|}{1 + \sqrt{n}\,|\nu - \nu_0|} : |\nu - \nu_0| \le \varepsilon_n\right\} = o_p(1)$$

(GM0′) $\forall\, M<\infty$, $\sup\left\{\left|V_n(\nu)-V_n(\nu_0)\right| : |\nu-\nu_0|\le Mn^{-1/2}\right\}=o_p(1)$.

(GM1) $\exists\ \nu: \mathcal{M}_0 \to R^m$ such that $W(\nu(P),P)=0\quad \forall P\in\mathcal{M}_0$.

(GM2) $\exists\ \psi: \mathcal{X}\times\mathcal{M}_0 \to R^m$, $\psi\in\left\{L_2^0(P_0)\right\}^m$ such that

$$W_n(\nu_0)=n^{-1}\sum_{i=1}^{n}\psi(Y_i,P_0)+o_p(n^{-1/2}).$$

(GM3) $W(\cdot,P_0)$ is differentiable at ν_0 and $\dot W(P_0)\equiv \dot W(\nu_0,P_0)$ is nonsingular.

THEOREM (BKRW-1993). *Suppose* $P_0\in\mathcal{P}$. *Let* $\hat\nu_n$ *be an AGM estimator of* ν_0. *If* $\hat\nu_n$ *is consistent and (GM0)-(GM3) hold or* $\hat\nu_n$ *is* $\sqrt{n}$*-consistent and* $(GM0')-(GM3)$ *hold, then*

$$\sqrt{n}(\hat\nu_n-\nu_0)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot W^{-1}(P_0)\psi(Y_i,P_0)+o_p(1).$$

2. Generalized M-estimation under interval censoring: From our computations in Theorem 4.3.2, the efficient score function for θ is

$$l_\theta^*(y)=-\left[z-k(t_\#,u_\#)\right]\left[Q_1(y)f(t_\#)+Q_2(y)f(u_\#)\right] \tag{2.0}$$

with $a_\#\equiv a-\theta z\quad \forall a$. We will derive a generalized estimator of θ by appropriately modifying (2.0). Notice (although not exhibited) that Q_1, Q_2 depend on both θ and F. We replace f, k in (2.0) by known functions $\tilde f, \tilde k$. In addition we replace $F(\cdot-\theta z)$ by its maximum likelihood estimator $F_n(\cdot-\theta z;\theta)$. Denote the corresponding *generalized score function* by $\tilde l_n(y;\theta)$ and let $W_n(\theta)\equiv \mathbb{P}_n\,\tilde l_n(\theta;Y)$, $W_n(\theta,P)\equiv P\tilde l_n(\theta;Y)$, with the functional representation of the integral $Pf=\int f\,dP$, used everywhere in this chapter. Finally let

$X-\theta Z \sim F(\cdot;\theta)\ \forall\theta\in\Theta$. From the consistency property of the NPMLE of F_0, (Groeneboom and Wellner (1992)), we obtain $F_n(\cdot-\theta z;\theta)\xrightarrow{as} F(\cdot-\theta z;\theta)$ with $F(\cdot-\theta_0 z;\theta_0)\equiv F_0(\cdot-\theta_0 z)$. It follows that

$$\text{(2.1)}\quad \tilde{l}_n(y;\theta)\xrightarrow{as}\tilde{l}(y;\theta)\equiv-\left[z-\tilde{k}(t_\#,u_\#)\right]\left\{\left[\frac{\delta}{F(t_\#;\theta)}-\frac{\gamma}{F(u_\#;\theta)-F(t_\#;\theta)}\right]\tilde{f}(t_\#)\right.$$

$$\left.+\left[\frac{\gamma}{F(u_\#;\theta)-F(t_\#;\theta)}-\frac{1-\delta-\gamma}{\bar{F}(u_\#;\theta)}\right]\tilde{f}(u_\#)\right\}.$$

Choosing carefully $\tilde{f},\tilde{k}$, we will prove that

$$\text{(2.2)}\qquad P_0\tilde{l}_n(\theta_0,Y)\to P_0\tilde{l}(\theta_0,Y)\quad\text{and}$$

$$\text{(2.3)}\qquad \boldsymbol{W}_n(\theta_0)\equiv(\boldsymbol{P}_n-P_0)\tilde{l}_n(\theta_0;Y)+P_0\tilde{l}_n(\theta_0;Y)\xrightarrow{as}P_0\tilde{l}(\theta_0;Y)\equiv W(\theta_0,P_0).$$

Moreover

$$W(\theta;P_0)\equiv P_0\tilde{l}(\theta,Y)=-\int\left[z-\tilde{k}\right]\left\{\left[\frac{F_0(t-\theta_0 z)}{F(t_\#;\theta)}-\frac{F_0(u-\theta_0 z)-F_0(t-\theta_0 z)}{F(u_\#;\theta)-F(t_\#;\theta)}\right]\tilde{f}(t_\#)+\right.$$

$$\left.\left[\frac{F_0(u-\theta_0 z)-F_0(t-\theta_0 z)}{F(u_\#;\theta)-F(t_\#;\theta)}-\frac{\bar{F}_0(u-\theta_0 z)}{\bar{F}(u_\#;\theta)}\right]\tilde{f}(u_\#)\right\}dJ(t,u,z),$$

from which it clearly follows that

$$\text{(2.4)}\qquad W(\theta_0;P_0)\equiv 0.$$

The following theorem is our modification of the master theorem presented earlier. It shows that we can relax the consistency assumption and still manage to obtain the same asymptotic result at the expense of more smoothness in the limit of the objective function. Consider the conditions

(C1) $\sup\left\{\left|V_n(\theta)-V_n(\theta_0)\right|:\left|\theta-\theta_0\right|\le n^{-a}\right\}=o_p(1)$, *for some* $a>0$.

(C2) $\exists\ \psi: R_+^2\times R^d\times\{0,1\}^2\times\Theta\to R^m,\ \psi\in\{L_2^0(P_0)\}^m$ *such that*

$$W_n(\theta_0) = P_n\psi(Y,\theta_0) + o_p(n^{-1/2}).$$

(C3) $\theta \mapsto W(\theta, P_0)$ *is differentiable at* θ_0 with $\dot{W}(\theta_0)$ *nonsingular. Moreover*

$$W(\theta,P_0) = W(\theta_0,P_0) + \dot{W}(\theta_0)(\theta-\theta_0) + |\theta-\theta_0|^{\gamma} \quad \textit{with} \quad \gamma a > 1/2.$$

(C4) $W(\theta_0, P_0) = 0$.

THEOREM 1. *Let* $\theta_0 \in \overset{\circ}{\Theta}$, $V_n(\theta) \equiv \sqrt{n}[W_n(\theta) - W(\theta, P_0)]$, $\theta \in \Theta$. Suppose *that* $(\mathrm{C1\text{-}C4})$ *hold. Then*

(i) *For large* n $\exists\, \tilde{\theta}_n : \left|\tilde{\theta}_n - \theta_0\right| \le n^{-a}$ *with* $\gamma a > 1/2$ *such that* $W_n(\tilde{\theta}_n) = o_p(n^{-1/2})$ *almost surely* P_0.

(ii) $\dot{W}(\theta_0)\sqrt{n}(\tilde{\theta}_n - \theta_0) = -\sqrt{n}\, P_n\psi(Y,\theta_0) + o_p(1)$,

from which it follows that

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma_0)$$

with $\Sigma_0 = \dot{W}(\theta_0)^{-1}[E_0\psi\psi']\left[\dot{W}(\theta_0)^{-1}\right]'$.

REMARK 1. As we have mentioned earlier, the verification of the conditions of Theorem 1 in our interval censoring problem, will be the subject of future work. However in Theorem 2 we prove a uniform law of large numbers, tailored for the needs of our linear regression model, see display (2.3). This result is used in the proof of Theorem 1.

REMARK 2. Notice the connection between the modulo of continuity in the objective function (C1) and the smoothness of the model (C3). If $\theta \mapsto W(\theta, P_0)$ is twice differentiable and (C3) holds with γ=2, then (C1) needs to be established for $a > 1/4$. In this case, it might be appropriate to follow the approach of Manski (1975), (1985), seeking "maximum scored" estimators of θ_0. Huang (1993), proved that the "maximum scored estimator" of θ_0 with case 1 interval censored data is $n^{1/3}$-consistent. If we manage to establish such a result with case 2 interval censoring, then all is left to do in the present context is to solve $W_n(\theta) = 0$ in a $|\theta - \tilde{\theta}_n| = O_p(n^{-1/3})$ neighborhood of $\tilde{\theta}_n$.

Proof of Theorem 1. Let $\theta \in B_n(a) = \{\theta : |\theta - \theta_0| \le n^{-a}\}$. In view of (C4) ,write (C1) as

$$W_n(\theta) = W(\theta) + W_n(\theta_0) + o_p(n^{-1/2})$$

$$\overset{(C3)}{=} W_n(\theta_0) + \dot{W}(\theta_0)(\theta - \theta_0) + |\theta - \theta_0|^{\gamma} + o_p(n^{-1/2})$$

$$= W_n(\theta_0) + \dot{W}(\theta_0)(\theta - \theta_0) + o_p(n^{-1/2}) \quad \text{since} \quad \gamma a > 1/2.$$

Now from (2.3), (2.4) and the definition of $B_n(a)$, (i) follows. To obtain (ii) write

$$W_n(\tilde{\theta}_n) = W_n(\theta_0) + \dot{W}(\theta_0)(\tilde{\theta}_n - \theta_0) + o_p(n^{-1/2}),$$

from which it follows that

$$\tilde{\theta}_n - \theta_0 = -[\dot{W}(\theta_0)]^{-1} W_n(\theta_0) + o_p(n^{-1/2}).$$

In view of (C2), the theorem is now proved.

The following lemma is used in the proof of Theorem 2. Recall that H, J, V denote the joint distributions of (T,U), (T,U,Z), $(T - \theta_0 Z, U - \theta_0 Z)$ respectively, while $[a_0, b_0]$ indicates the support of F_0.

LEMMA 1. *Suppose that*

(i) F_0 *is strictly increasing.*

(ii) J *is a continuous distribution,* (T,U) *and* Z *have densities with respect to Lebesgue measure.*

(iii) (T,U) *and* Z *are bounded. Moreover* $\forall\ (\tau,\rho) \in$ *support*$(V) \equiv S_V$

$$0 < F_0(\tau) < F_0(\rho) < 1.$$

(iv) $\exists\, \zeta > 0 : H\{(t,u):\ u \geq t + \zeta\} = 1.$

Then there exists a neighborhood $N(\theta_0)$ *of* θ_0 *and an* $M > 0$ *such that*

$$\sup_{\theta \in N(\theta_0)} \sup_{t,u,z} \left[\frac{1}{F_n(t-\theta z;\theta;\omega)} + \frac{1}{F_n(u-\theta z;\theta;\omega) - F_n(t-\theta z;\theta;\omega)} + \frac{1}{\bar{F}_n(u-\theta z;\theta;\omega)} \right] \leq M$$

for n large enough and $\omega \in B$ *with* $P_0^\infty(B) = 1$.

Proof of Lemma 1. According to our definition in chapter 2 and for $\theta \in \Theta$ fixed,

$$F_n(\cdot,\theta) = \arg\max_{F \in \mathfrak{I}} l_n(\theta, F).$$

Let $\varepsilon \in (0,1)$. For simplicity we denote by $F_n(\cdot) \equiv F_n(\cdot,\theta)$. Then

$$l_n[(1-\varepsilon)F_n + \varepsilon F_0] - l_n(F_n) \leq 0 \quad \forall n.$$

It follows that

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}\{l_n[(1-\varepsilon)F_n + \varepsilon F_0] - l_n(F_n)\} \leq 0 \quad \forall n.$$

If we evaluate this limit explicitly we obtain

$$\text{(2.5)} \quad \int \left[\delta\, \frac{F_0(t-\theta z)}{F_n(t-\theta z)} + \gamma\, \frac{F_0(u-\theta z) - F_0(t-\theta z)}{F_n(u-\theta z) - F_0(t-\theta z)} + (1-\gamma-\delta)\, \frac{\bar{F}_0(u-\theta z)}{\bar{F}_n(u-\theta z)} \right] d\boldsymbol{P}_n(\delta,\gamma,t,u,z) \leq 1.$$

In view of (ii) and (iii), we can find $[a,b] \subset [a_0,b_0]$ such that for a properly chosen neighborhood $N(\theta_0)$ of θ_0, we have, $\forall\, \theta \in N(\theta_0)$ and $\forall\, (\tau,\rho) \in S_{(T-\theta Z, U-\theta Z)}$, where

$S_{(T-\theta Z, U-\theta Z)}$ is the support of $(T-\theta Z, U-\theta Z)$,

(2.6) $$0 < F_0(a) \le F_0(\tau) < F_0(\rho) \le F_0(b) < 1.$$

From (2.5) and (2.6) we obtain

(2.7) $$F_0(a) \int \frac{\delta}{F_n(t-\theta z)} d\boldsymbol{P}_n \le 1.$$

(2.8) $$\eta \int_{A_\eta} \frac{\gamma}{F_n(u-\theta z) - F_n(t-\theta z)} d\boldsymbol{P}_n \le 1$$

with $A_\eta = \{(t,u) \in [a,b]^{\otimes 2} : F_0(u) - F_0(t) > \eta\}$ and some $\eta > 0$.

(2.9) $$\bar{F}_0(b) \int \frac{1-\gamma-\delta}{\bar{F}_n(u-\theta z)} d\boldsymbol{P}_n \le 1.$$

From the strong law of large numbers we obtain the following weak convergence result

$$\boldsymbol{P}_n(\cdot;\omega) \xrightarrow{w} P_0(\cdot) \quad \text{for } \omega \in B : P_0^\infty(B) = 1.$$

We now claim that $\exists$ constant $M_2 > 0$ such that for n large enough and $\omega \in B$,

(2.10) $$\frac{1}{F_n(u-\theta z;\omega) - F_n(t-\theta z;\omega)} \le M_2 \quad \forall (t,u,z) \in S_J : \{t-\theta z, u-\theta z\} \in [a,b].$$

If this is not true, then for $(t_*, u_*, z_*) \in S_J$, $(t_* - \theta z_*, u_* - \theta z_*) \in [a,b]^{\otimes 2}$ and with positive probability

(2.11) $$\frac{1}{F_n(u_* - \theta z_*;\omega) - F_n(t_* - \theta z_*;\omega)} > M \quad \forall M > 0.$$

Let $A = \{(x,y) : a \le t_* - \theta z \le x < y \le u_* - \theta z \le b\}$. Note that by the monotonicity of $F_n(\cdot;\omega)$, (2.11) holds for every point in A. Moreover A has positive Lebesgue measure. Therefore

$$\eta \int_{A_\eta} \frac{\gamma}{F_n(u-\theta z;\omega) - F_n(t-\theta z;\omega)} d\boldsymbol{P}_n(\cdot;\omega)$$

$$> \eta M \int_{A_\eta \cap A} \gamma d\boldsymbol{P}_n(\cdot;\omega)$$

$$= \eta M \boldsymbol{P}_n(A_\eta \cap A \cap \{t < x \le u\};\omega)$$

$$\geq \frac{1}{10}\eta\, M\, P_0\left(A_\eta \cap A \cap \{t < x \leq u\}\right)$$

$$> 1 \quad \text{for large } n, \text{M}.$$

Since this violates (2.8), we conclude that (2.10) holds.

Similarly we can prove that

$$(2.12) \qquad \frac{1}{F_n(t-\theta z;\omega)} \leq M_1 \quad \forall (t,u,z) \in S_J \ : t-\theta z \in [a,b]$$

and

$$(2.13) \qquad \frac{1}{\overline{F}_n(u-\theta z;\omega)} \leq M_3 \quad \forall (t,u,z) \in S_J \ : u-\theta z \in [a,b] \quad \text{with } \omega \in B : P_0^\infty(B) = 1.$$

The result now follows from (2.10), (2.12) and (2.13) for $M = M_1 + M_2 + M_3$.

We conclude this section with the theorem that establishes the uniform strong law of large numbers for our empirical process $\boldsymbol{W}_n(\theta) \equiv \boldsymbol{P}_n \tilde{l}_n(Y,\theta)$.

THEOREM 2. *In addition to the assumptions in Lemma 1, assume that*

(i) $\tilde{k}$ *is a bounded function on* $[a_0,b_0]^{\otimes 2}$.

(ii) $\tilde{f}$ *is Lipschitz, i.e.*

$$\left|\tilde{f}(x) - \tilde{f}(y)\right| \leq m|x-y| \quad \forall x, y.$$

Then $$\lim \boldsymbol{W}_n(\theta_0) = W(\theta_0) \quad \text{as } P_0^\infty.$$

Proof of Theorem 2. Recall the definitions of

$$(2.14) \qquad \boldsymbol{W}_n(\theta) \equiv \boldsymbol{P}_n \tilde{l}_n(Y,\theta) = (\boldsymbol{P}_n - P)\tilde{l}_n(\theta;Y) + P\tilde{l}_n(\theta;Y) \qquad \theta \in \Theta.$$

and

$$\tilde{l}_n(y;\theta) \equiv -\left[z - \tilde{k}(t_\#,u_\#)\right] \left\{ \left[\frac{\delta}{F_n(t_\#;\theta)} - \frac{\gamma}{F_n(u_\#;\theta) - F_n(t_\#;\theta)} \right] \tilde{f}(t_\#) \right.$$

$$+\left[\frac{\gamma}{F_n(u_\#;\theta)-F_n(t_\#;\theta)}-\frac{1-\delta-\gamma}{\bar{F}_n(u_\#;\theta)}\right]\tilde{f}(u_\#)\ \}.$$

Let r_n be a sequence of positive numbers such that $r_n \downarrow 0$. Write

$$\sup_{|\theta-\theta_0|\le r_n}\left|(\boldsymbol{P}_n-P)\tilde{l}_n(\theta,Y)\right| \le (\boldsymbol{P}_n-P)\sup_{|\theta-\theta_0|\le r_n}\left|\tilde{l}_n(\theta,Y)\right| \qquad \forall n.$$

From (i) and (ii) in our hypotheses and Lemma 1 we obtain

$$\sup_{|\theta-\theta_0|\le r_n}\left|\tilde{l}_n(\theta,y)\right| \le M_1\sup_{|\theta-\theta_0|\le r_n}\tilde{f}(t-\theta z)+M_2\sup_{|\theta-\theta_0|\le r_n}\tilde{f}(u-\theta z)$$

$$\le M_1\sup_{|\theta-\theta_0|\le r_n}\left|\tilde{f}(t-\theta z)-\tilde{f}(t-\theta_0 z)\right|+M_2\sup_{|\theta-\theta_0|\le r_n}\left|\tilde{f}(u-\theta z)-\tilde{f}(u-\theta_0 z)\right|$$

$$+M_1\tilde{f}(t-\theta_0 z)+M_2\tilde{f}(u-\theta_0 z)$$

$$\le\left(\tilde{M}_1+\tilde{M}_2\right)r_n+M_1\tilde{f}(t-\theta_0 z)+M_2\tilde{f}(u-\theta_0 z), \quad \text{as } P_0^\infty .$$

Taking the limit as n goes to infinity and from the strong law of large numbers applied to

$M_1\tilde{f}(t-\theta_0 z)+M_2\tilde{f}(u-\theta_0 z)$, we obtain

(2.15) $$\lim_{n\to\infty}\sup_{|\theta-\theta_0|\le r_n}\left|(\boldsymbol{P}_n-P_0)\tilde{l}_n(\theta,Y)\right|=0 \quad \text{as } P_0^\infty .$$

To prove that $P_0\tilde{l}_n(\theta_0,Y)\to P_0\tilde{l}(\theta_0,Y)$ we note that $\tilde{f}(t-\theta_0 z)$, $\tilde{f}(u-\theta_0 z)$ are bounded in a closed subset of $[a_0,b_0]$. Thus Lemma 1, (2.1) and the Lebesgue dominated convergence theorem establish

(2.16) $$P_0\tilde{l}_n(\theta_0,Y)\to P_0\tilde{l}(\theta_0,Y).$$

Now from (2.15) and (2.16) we obtain

$$\boldsymbol{W}_n(\theta_0)\to W(\theta_0,P_0)=0 \quad \text{as } P_0^\infty .$$

The theorem is now proved.

Chapter 6

SIMULATIONS

We have seen in chapter 5 that the profile likelihood approach doesn't always work in the AFT model as nicely as it does in the PH model. The source of most problems with this model is the frequent lack of smoothness in the function $\theta \mapsto F_n(\cdot,\theta)$. In fact, very little one can say about this function and its properties. Nevertheless, smoothness is essential in obtaining the maximizer in step 2 of the profile likelihood algorithm. To shed some light in the behavior of $\theta \mapsto F_n(\cdot,\theta)$, a limited simulation study is conducted to examine the behavior of the profile likelihood as a function of θ.

We work with the model

$$X = \theta Z + \varepsilon$$

where Z is a single covariate having Bernoulli(.5) distribution and ε a normally distributed random variable. We present the results of the simulation for sample sizes of n=100 and n=1000 observations. Non negative, independent random variables T_1, T_2 are generated from preselected distributions and the censoring pair (inspection times) is constructed according to

$$T \equiv T_1, \;\; U \equiv T_1 + T_2 .$$

In an interval centered around the true θ, we take a grid of points and for every such point we compute the MLE $F_n(\cdot,\theta_i): i = 1,2,\ldots,k$, with k indicating the cardinality of the grid. We then record the value of the profile likelihood function $\pi(\theta_i) = l(\theta_i, F_n(\cdot,\theta_i))$ and plot the pairs $(\theta_j, \pi(\theta_j))$. For $\theta_* \equiv \arg\max_{1\le j\le k} \pi(\theta_j)$, we set $\hat{\theta}_n \equiv \theta_*$ and $\hat{F}_n(\cdot) \equiv F_n(\cdot,\theta_*)$.

Table 6.1: Profile Likelihood

X=2Z+N(0,1) $T_1 \sim Exp(.5), T_2 \sim Exp(.5)$			
n=100		n=1000	
θ	π(θ)	θ	π(θ)
-1.0	-58.54	1.0	-475.9
-0.5	-51.66	1.2	-460.5
0	-44.15	1.4	-445.1
0.5	-33.19	1.6	-433.7
1	-26.7	1.8	-426.8
1.5	-25.16	2.0	-424.4
2	-25.6	2.2	-428.2
2.5	-28.7	2.4	-434.7
3	-29.4	2.6	-441.4
3.5	-29.5	2.8	-448.7
4.0	-29.5	3.0	-457.6
4.5	-29.6		
5	-29.8		

The proportion of censored observations in the two samples by type of censoring is (.60,.30,.10) and (.60,.25,.15) respectively, for left, interval, right censoring. The π function is maximized for $\theta_*=1.5$ when n=100 and for $\theta_*=2$ when n=1000.

Table 6.2: Profile Likelihood

X=.5Z+N(4,1)			
$T_1 \sim Unif[3,5]$, $T_2 \sim Unif[0,1]$			
n=100		n=1000	
θ	π(θ)	θ	π(θ)
-1.0	-93.18	0	-860.0
-0.8	-90.88	0.1	-854.0
-0.6	-88.31	0.2	-850.2
-0.4	-86.60	0.3	-845.0
-0	-87.0	0.4	-840.2
0	-83.29	0.5	-839.8
0.2	-79.45	0.6	-837.6
0.4	-80.61	0.7	-841.5
0.6	-78.34	0.8	-840.6
0.8	-77.59	0.9	-845.5
1.0	-79.71	1.0	-850.4

The proportion of (left,interval,right) censoring is (.40,.18,.42) and (.38,.18,.44) respectively, in the two groups. With 50% of the data essentially censored (to the right), convergence of the profile estimator is much slower. With n=100, we get $\theta_{*}=.8$, while increasing the sample size to n=1000, we only obtain $\theta_{*}=.6$. The π function is obviously "less" smooth here. In the next figures we plot the π function and the maximum (profile) likelihood estimate of the error distribution.

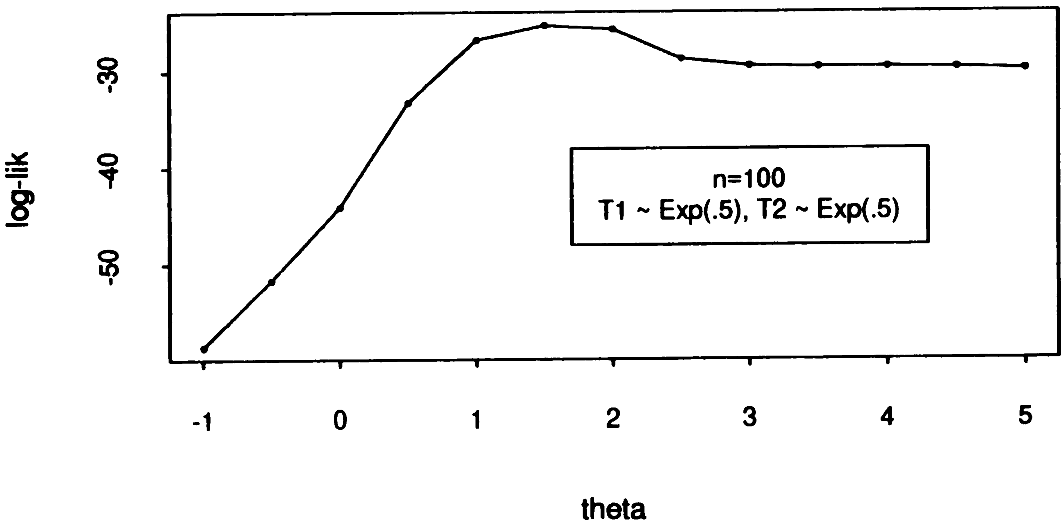

Figure 6.1a: Profile Likelihood: X=2Z+N(0,1)

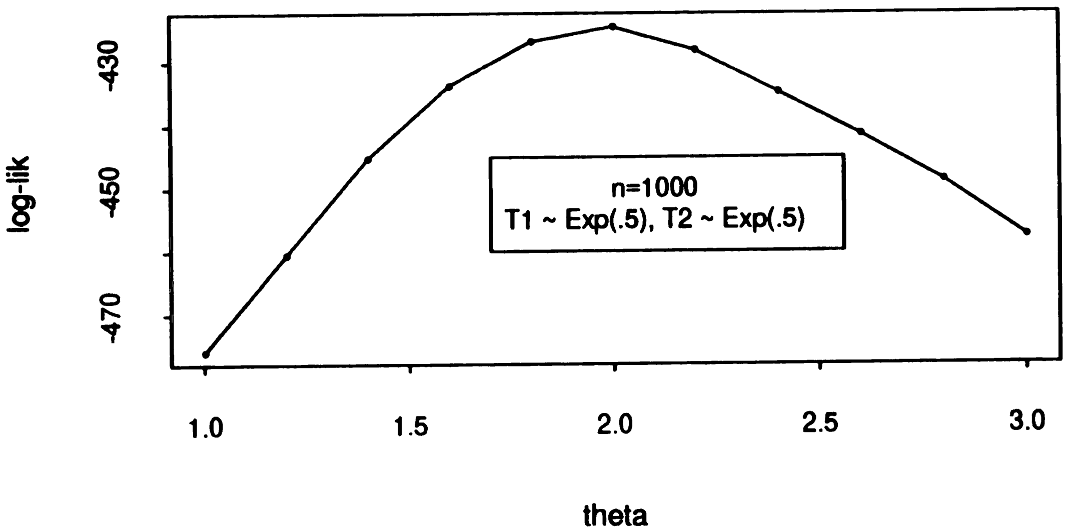

Figure 6.1b: Profile Likelihood: X=2Z+N(0,1)

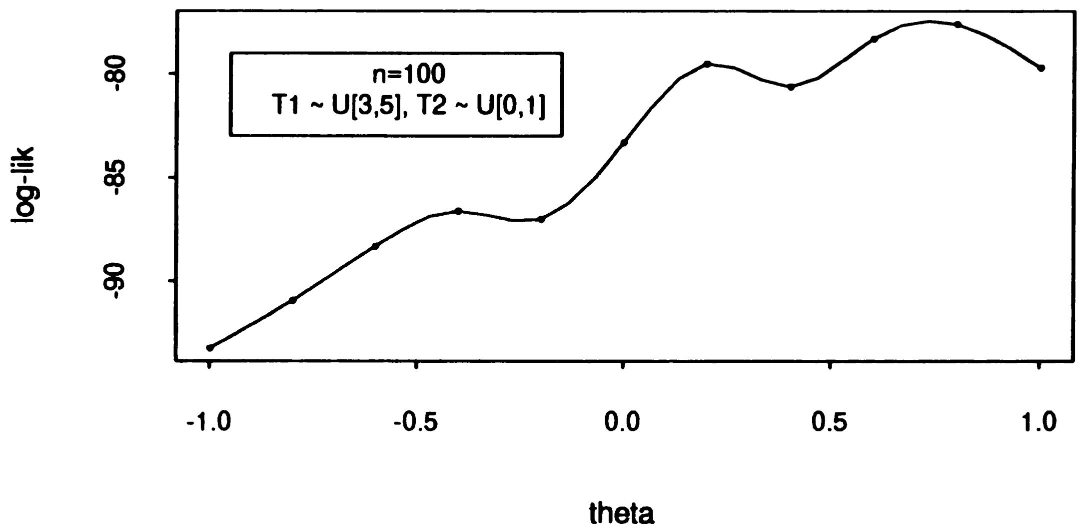

Figure 6.2a: Profile Likelihood: X=.5Z+N(4,1)

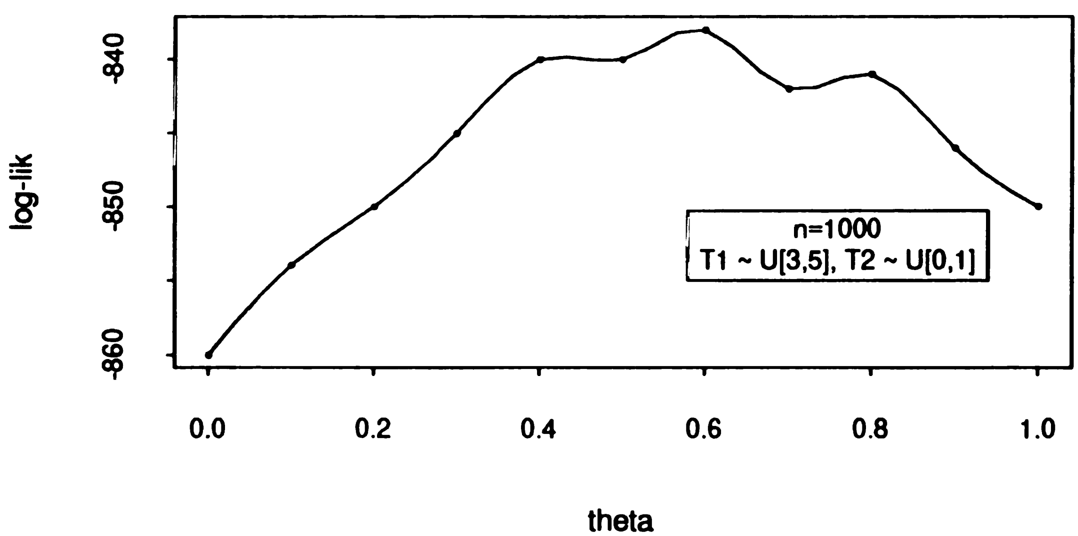

Figure 6.2b: Profile Likelihood: X=.5Z+N(4,1)

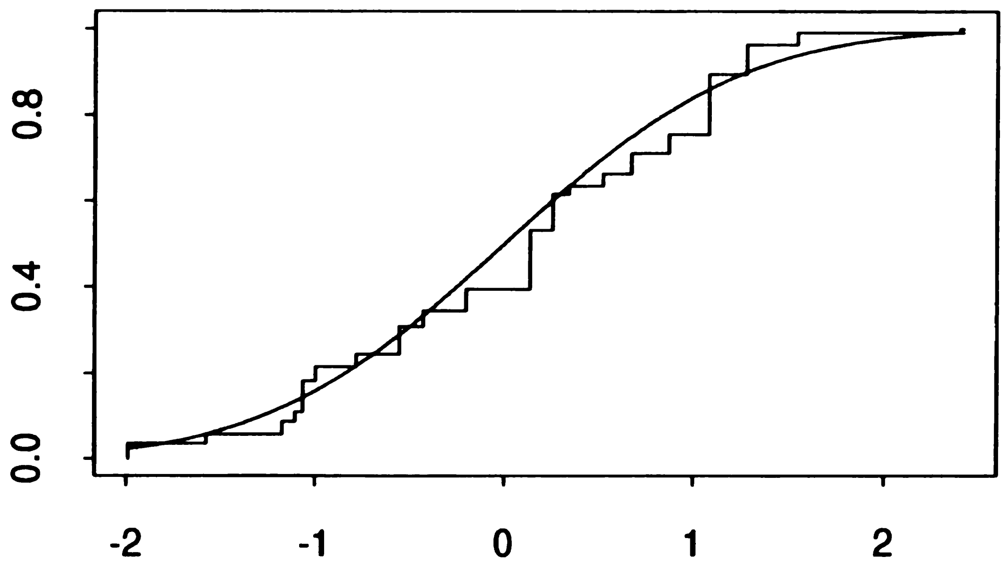

Figure 6.3a: M.L.E. of error distribution: X=2Z+N(0,1)

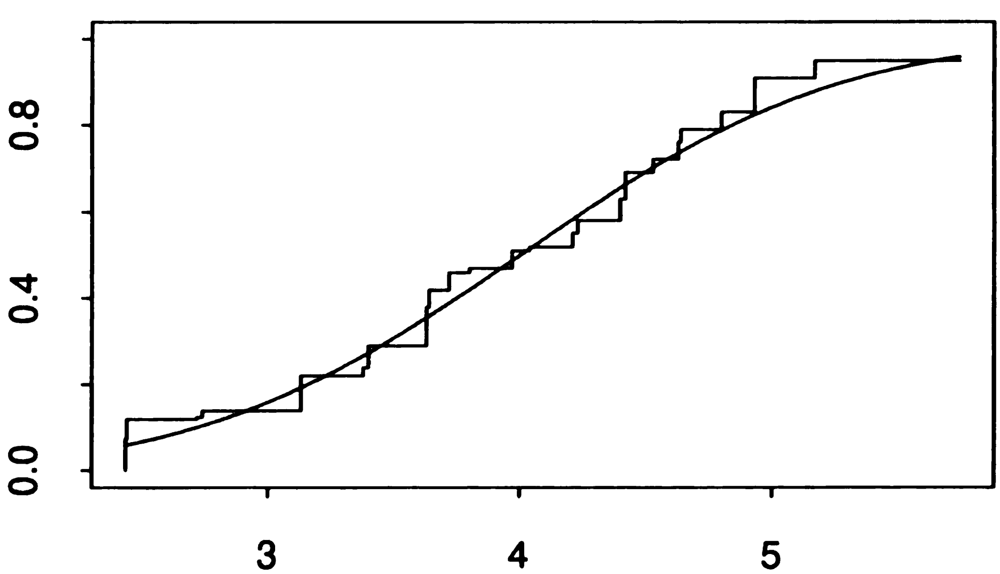

Figure 6.3b: M.L.E. of error distribution: X=.5Z+N(4,1)

APPENDIX

Appendix

1. Concavity of the function $\Lambda \mapsto l_n(\theta;\Lambda)$**:** In chapter 2 we have used without proof the concavity of the log-likelihood function with respect to Λ for every fixed $\theta \in \Theta$. In this section we give a short proof of this claim.

Let $a>0$ and consider the function

$$\psi_a(t,u) = \delta \log\left(1-e^{-ta}\right) + \gamma \log\left(e^{-ta} - e^{-ua}\right) - (1-\gamma-\delta)au .$$

Then $\frac{\partial}{\partial t}\psi_a(t,u) = \frac{\delta a e^{-ta}}{1-e^{-ta}} - \frac{\gamma a e^{-ta}}{e^{-ta}-e^{-ua}}$, $\frac{\partial}{\partial u}\psi_a(t,u) = \frac{\gamma a e^{-ta}}{e^{-ta}-e^{-ua}} - (1-\gamma-\delta)a$.

Set

$$\psi_{tt}(\delta,\gamma) \equiv \frac{\delta e^{-ta}\left(1-e^{-ta}\right)+\delta e^{-2ta}}{\left(1-e^{-ta}\right)^2} + \frac{\gamma e^{-(t+u)a}}{\left(e^{-ta}-e^{-ua}\right)^2}, \quad \psi_{uu}(\delta,\gamma) \equiv \frac{\gamma e^{-a(t+u)}}{\left(e^{-ta}-e^{-ua}\right)^2}$$

$$\psi_{tu}(\delta,\gamma) \equiv -\frac{\gamma e^{-a(t+u)}}{\left(e^{-ta}-e^{-ua}\right)^2}.$$

Then we can write

$$\frac{\partial^2}{\partial t^2}\psi_a(t,u) = -a^2\psi_{tt}(\delta,\gamma), \quad \frac{\partial^2}{\partial u^2}\psi_a(t,u) = -a^2\psi_{uu}(\delta,\gamma),$$

$$\frac{\partial^2}{\partial u \partial t}\psi_a(t,u) = -a^2\psi_{tu}(\delta,\gamma).$$

Now the matrix of second derivatives is

$$\mathcal{B}(\delta,\gamma) \equiv -a^2\begin{bmatrix} \psi_{tt}(\delta,\gamma) & \psi_{tu}(\delta,\gamma) \\ \psi_{tu}(\delta,\gamma) & \psi_{uu}(\delta,\gamma) \end{bmatrix} = -a^2\widetilde{\mathcal{B}}(\delta,\gamma).$$

It is easy to verify that $\widetilde{\mathscr{D}}(\delta,\gamma)$ is non-negative definite. Thus $\psi_a(t,u)$ is concave as claimed.

2. Alternative characterization of MLE: We give here an alternative characterization of the maximum likelihood estimator in both models that we considered. It is based on a geometric interpretation of the NPMLE, as the left derivative of the gratest convex minorant of a cumulative sum diagram. The idea is given in Robertson, Wright and Dykstra (1988) and was first implemented in the interval censored problems by Groeneboom, see Groeneboom and Wellner (1992).

The process W associated with the first derivative of the log-likelihood function with respect to Λ in the Cox model is

$$W_{\Lambda,q}(t) = \sum_{i:\, T_i \le t} \left(\frac{\delta_i}{1 - e^{-\Lambda(T_i)e^{q_i}}} - \frac{\gamma_i}{e^{-\Lambda(T_i)e^{q_i}} - e^{-\Lambda(U_i)e^{q_i}}} \right) e^{q_i - \Lambda(T_i)e^{q_i}} +$$
$$\sum_{i:\, U_i \le t} \left(\frac{\gamma_i}{e^{-\Lambda(T_i)e^{q_i}} - e^{-\Lambda(U_i)e^{q_i}}} - \frac{1 - \gamma_i - \delta_i}{e^{-\Lambda(U_i)e^{q_i}}} \right) e^{q_i - \Lambda(U_i)e^{q_i}} .$$

Let

$$G_{\Lambda,q}(t) = -\frac{d}{dt} W_{\Lambda,q}(t) \quad \text{and} \quad dV_{\Lambda,q}(t) = dW_{\Lambda,q}(t) + \Lambda(t)\, dG_{\Lambda,q}(t).$$

The following proposition is adapted from Groeneboom and Wellner (1992). It provides an alternative characterization of the MLE of Λ, thus giving an equivalent statement to the one presented in our Theorem 2.6.1.

PROPOSITION 1. *For fixed* $\theta \in \Theta$, *let* $q_i = \theta' z_i$, $i \in \{1,2,\ldots,n\}$. *Suppose that* $\delta_{(1)} = 1$ *and* $\delta_{(m)} = \gamma_{(m)} = 0$. *Then* $\widetilde{\Lambda}_n(\cdot,\theta)$ *is the NPMLE of* Λ_0 *if and only if* $\widetilde{\Lambda}_n(\cdot,\theta)$

is the left derivative of the convex minorant of the "cumulative sum diagram", consisting of the points

$$P_{j,q} = \left(G_{\tilde{\Lambda}_n(\cdot,\theta),q}\left(\eta_{(j)}\right), V_{\tilde{\Lambda}_n(\cdot,\theta),q}\left(\eta_{(j)}\right)\right)$$

where $P_0 = (0,0)$ and $\eta_{(j)} \in J_n$, $j = 1,2,\ldots,m$.

A similar proposition can be formulated to characterize the MLE in the linear regression model. It can be used as an equivalent statement to Theorem 2.6.2. The associated W process is given by

$$W_{F,\theta}(t) = \sum_{i=1}^{n} 1_{\{T_i^\theta \le t\}} \left(\frac{\delta_i^\theta}{F(T_i^\theta)} - \frac{\gamma_i^\theta}{F(U_i^\theta) - F(T_i^\theta)} \right)$$
$$+ \sum_{i=1}^{n} 1_{\{U_i^\theta \le t\}} \left(\frac{\gamma_i^\theta}{F(U_i^\theta) - F(T_i^\theta)} - \frac{1 - \delta_i^\theta - \gamma_i^\theta}{1 - F(U_i^\theta)} \right).$$

3. Some results from empirical process theory: In this section we summarize results that can be used to prove uniform central limit theorems and laws for large numbers. We see the need for such machinery in our chapter 5, in our effort to verify the hypotheses of the master theorem of BKRW (1994). Most of them can be found in Dudley (1984), (1987) and Pollard (1984), (1989). Let (S,d) be a metric space, $B \subset S$, $\varepsilon > 0$ and $\mathcal{J} \subset \mathcal{L}^r(P)$, a family of functions for some $r>0$. Denote by $\ln N(\varepsilon, B, d)$, $\ln D(\varepsilon, B, d)$, $\ln N_B(\varepsilon, \mathcal{J}, d)$, the ε–entropy, ε–capacity and ε– bracket entropy of B respectively. The following proposition is a simple consequence of the definition of N,D.

PROPOSITION 1: *For every* $\varepsilon > 0$ *and* $\forall$ *set* B *in a metric space* (S,d),

$$D(2\varepsilon, B, d) \le N(\varepsilon, B, d) \le D(\varepsilon, B, d).$$

We now define the concept of a manageable class of functions . Pollard (1989) introduced these classes and obtained results that go beyond the Vapnik-Cervonenkis (1971) theory of VC classes of sets, thus extending the availability of central limit theorems to estimators that depend on larger classes of functions.

DEFINITION 1: Let $\mathcal{J}$ be a class of functions with an envelope F, that is $|f| \le F \;\; \forall f \in \mathcal{J}$ and let $\|\cdot\|_{2,Q}$ indicate the $L_2(Q)$ norm.. We say that $\mathcal{J}$ *is manageable for the envelope* F if there exists a decreasing function $\Gamma(\cdot)$ for which

(i) $\int_0^1 (\log \Gamma(x))^{1/2} dx < \infty$ and (ii) $\sup_Q D\left(\varepsilon \|F\|_{2,Q}, \mathcal{J}, \|\cdot\|_{2,Q}\right) \le \Gamma(\varepsilon) \quad \forall\; 0 < \varepsilon \le 1$,

where the supremum is taken over all measures with finite support.

From the above definition it follows, see Dudley (1987), that

1. Every subclass $\mathcal{J}$ of a VC-subgraph class is manageable for $F = \sup_{\mathcal{J}} |f|$.
2. Every subclass $\mathcal{J}$ of a VC-hull class is manageable for $F = \sup_{\mathcal{J}} |f|$.
3. Every subclass $\mathcal{J}$ of a VC-major class is manageable for constant $F = \sup_{\mathcal{J}} |f|$.

Moreover one can construct more complex manageable classes of functions by starting from simple classes and then use their stability properties to build new ones. For further details see Pollard (1989). We are now ready to give the most appealing result associated with these classes.

THEOREM 1 (Pollard 1989): *Let* $\mathcal{J}$ *be a manageable class for an envelope* F *with* $PF^2 < \infty$. *Let* $\gamma_n \equiv \sqrt{n}(\boldsymbol{P}_n - P)$ *and for subclasses* $\mathcal{J}(n)$, *n=1,2,...*

$$\textit{(i)}\ \ 0 \in \mathcal{F}(n)\ \ \forall n \quad \textit{and} \quad \textit{(ii)}\ \ \sup_{\mathcal{F}(n)} P|f| \to 0 \quad \text{as } n \to \infty.$$

Then

$$E \sup_{\mathcal{F}(n)} |\gamma_n f|^2 \to 0 \quad \text{as} \quad n \to \infty.$$

The concept of a manageable class comes close to Dudley's definition of *functional Donsker classes*, see Dudley (1987). Although a manageable class for a constant envelope is a functional Donsker class, not all Donsker classes are manageable. Let $\mathcal{F}$ be a class of uniformly bounded functions on a probability space $(\mathcal{X}, \mathcal{A}, P)$. Set

$$\nu_n(f) \equiv \sqrt{n}(\boldsymbol{P}_n - P)f \quad f \in \mathcal{F}.$$

DEFINITION 2. The class of functions $\mathcal{F}$ is said to be *a functional Donsker class* if and only if

(i) $\mathcal{F}$ is totally bounded for the sup-norm.

(ii) $\exists\ \delta > 0$ such that $\sup_{\|f-g\|<\delta} |\nu_n(f) - \nu_n(g)| = o_p(1)$, modulo measurability constraints.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aalen, O.O. (1978b). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-26.

Andersen, P. K., Borgan, O., Gill, R.D., Keiding, N. (1993*). Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

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

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.

Buckley, J., and James, I. R. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Cox, D. R. (1972). Regression models and life-tables. *J. R. Statist. Soc. B*, **34**, 187-220.

Leblanc, M., and Crowley, J. (1995). Semiparametric regression functionals. *JASA*, **429**, 95-105.

Dudley, R. M. (1984). A course on empirical processes. *École d' Été de Probabilités de Saint-Flour XII-1982*. Lecture Notes in Math. **1097,** 2-142. Springer, New York.

Dudley, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15,** 1306-1326.

Finkelstein, D. M. and Wolfe R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41**, 933-945.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845-854.

Fleming, T. R., Harrington, D. P. (1991). *Counting processes and survival analysis.* Wiley, New York.

Green, P. and Yandell, B., (1985). Semiparametric generalized linear models. *Proceedings 2nd International GLIM conference,* Lecture notes in Statistics, **32**, Springer - Verlag, Berlin.

Heckman, N. E. (1986). Spline smoothing in a partially linear model, *J. Roy. Statist. Soc.,B,* **48**, 244-248.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Prob. Theory and Related Fields*, **81**, 79-109.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* DMV Seminar Band **19**, Birkhäuser, Basel.

Hoel, D. G. and Walburg, H. E. (1972). Statistical analysis of survival experiments. *J. Nat. Canc. Inst.*, **49**, 361-372.

Huang, J. (1993). Maximum scored likelihood estimation of a linear regression model with interval censored data. *Tech. Report*, No. **253**, Dept. of Statist., Univ. of Washington.

Huang, J. and Wellner, J. A. (1993). Regression models with interval censoring. *Proceedings of the Kolmogorov Seminar, Euler Mathematics Institute*, St. Petersburg, Russia.

Koul, H. L., Sousarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly righr censored data. *Ann. Statist.* **9**, 1276-1288.

Leiderman, P.H., Babu, D., Kagia, J., Kramer, H.C., Leiderman, G.F. (1973). African infant precocity and some social influence during the first year. *Nature*, **242**, 247-249.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics*, **3** , 205-228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *J Econometrics*, **27**, 313-333.

Paneth, N., Pinto-Martin, J., Gardiner, J., Wallenstein, S., Katsikiotis, V., Hegyi, T., Hiatt, M., Susser, M. (1993). Incidence and timing of Germinal Matrix/ Intraventricular hemorrhage in low birth weight infants. *Am. J. Epidemiol.* **137**, 1167-1176.

Peto, R. and Peto, J.(1972). Asymptotically efficient rank invariant test procedures. *J. R. Statist. Soc. A,* **135**, 185-206.

Pfanzagl, J. (1988). Consistency of maximum likelihood estimators for certain nonparametric families, in particular: Mixtures. *J. Statist. Plann. Inference*, **19**, 137-158.

Pinto-Martin, J., Paneth, N., Witomski, T., Stein, I., Schonfeld, S., Rosenfeld, D., Rose, W., Kazam, E., Kairam, R., Katsikiotis, V., Susser, M. (1992). The central New Jersey neonatal brain haemorrhage study: design of the study and reliability of ultrasound diagnosis. *Paediatr. Perinat. Epidemiol.* **6**, 273-84.

Pollard, D., (1984). *Convergence of Stochastic Processes*. Springer, New York.

Pollard, D., (1989). Asymptotics via Empirical Processes. *Statist. Sci.* **4**, 341-366

Ritov, Y., (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303-328.

Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

Rücker, G., Messerer, D.,(1988). Remission duration: An example of interval censored observations. *Statist. in Medicine*, **7**, 1139-1145.

Schick, A.,(1993). On efficient estimation in regression models. *Ann. Statist.* **21**, 1486-1521.

Silverman, B. M. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69**, 169-173.

Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist. Soc. B*, **38**, 290-295.

Van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14-44.

Wang, J.-L. (1985). Strong consistency of approximate maximum likelihood estimators with applications to nonparametrics. *Ann. Statist.* **13**, 932-946.

Wang, Z., Gardiner, C. J., Ramamoorthi, V. R. (1994). Identifiability in interval censhorship models. *Statist. Prob. Letters*, **21**, 215-221.

Whittemore, S. A. and Keller, B. J. (1986). Survival estimation using splines. *Biometrics*, **42**, 495-506.