東海大學統計研究所

碩士論文

指導教授:沈葆聖博士

雙設限資料下之Aalen模型

Aalen's Linear Model for Doubly Censored Data

研究生:林育璞

中華民國一百零四年六月

論文謝詞

此篇論文的完成，首先要感恩我的指導老師——沈葆聖教授，在這完成學業的過程中用心的指導與鼓勵，在論文完成的背後有著更多看不到的是老師的愛與關懷。

此外也要感謝遠從台北前來東海大學的戴政教授，在口試的過程中給予了許多寶貴的意見，不只是學術的知識，包括了社會上的經驗，讓學生受益良多。以及感謝系上的林正祥教授，在口試過程中不辭辛勞的審查與給予建議。

感謝同學與朋友對我的幫忙與祝福，最後感謝家人無盡的關懷與付出，和無限的支持與鼓勵。

Aalen's Linear Model for Doubly Censored Data

Director: Pao-sheng Shen

Student: Yui-pu Lin

Department of Statistics

Tunghai University

Taichung, Taiwan 40704

Abstract

In this article, we consider estimation of Aalen's nonparametric regression coefficients when data is subject to double censoring. We propose two estimation techniques. The first type of estimators, including ordinary least squared (OLS) estimator and weighted least squared (WLS) estimators, are obtained using martingale arguments. The second type of estimator, called the maximum likelihood (ML) estimator, is obtained via EM algorithm that treat the survival times of left censored observations as missing. Simulation study indicated the ML estimate is more efficient than both OLS and WLS estimators.

Keywords: Additive model; Martingale; EM algorithm.;

Contents

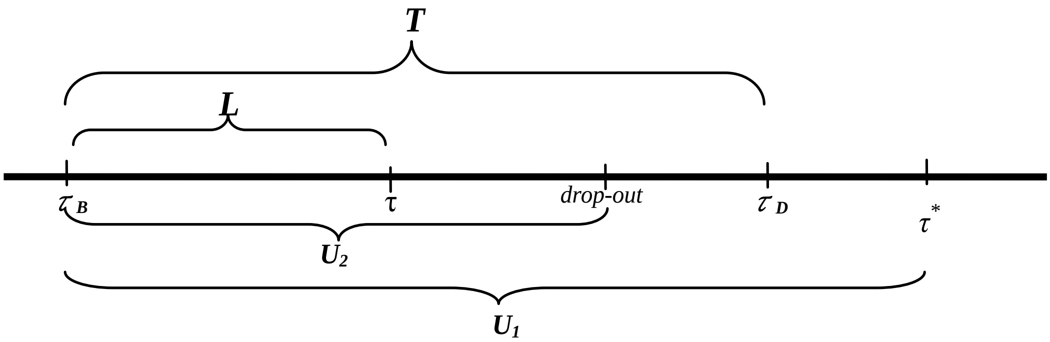

Figure 1. Schematic depiction of doubly censored data described in Example 1

Chapter 1

Introduction

In survival or reliability studies, the observed data is typically censored. Left-censoring and right censoring together naturally occur in doubly censored data. Double censoring arises when T represents an outcome variable that can only be accurately measured within a certain range, $[L, U]$, where L and U are the left- and right-censoring variables, respectively. In some case, L is always observed. Consider the following applications.

Example 1: Follow-up Studies

In early childhood learning centers, interest often focuses upon determining when a child learns to accomplish certain tasks. Consider a follow-up study (from calendar time τ to τ^*) for determining the ages T at which a child first develops the skill to accomplish certain task. Let τ_B and τ_D denote the birth date and the date of developing the skill. Let $L = \tau - \tau_B$ denote the child's age at entry and $U = \min(U_1, U_2)$, where $U_1 = \tau^* - \tau_B$ denotes the child's age at the termination of the program and U_2 denotes the child's age when he or she is lost to follow-up. One can observe $T = \tau_D - \tau_B$ if the child develops the skill to accomplish the task after entering the program, i.e. $L \leq T \leq U$. However, for some children in the program, the development may have been completed before entry, i.e. $T < L$ (left-censoring). Furthermore, a child may drop out or has not developed the skill by the time of the termination of the program, i.e. $U < T$ (right-censoring). Since the age at entry (i.e. L) is always observed, we observe a vector $(X,\ \delta,\ L)$, where $X = \min(U, \max(L, T))$, and $\delta = 1$ if $X = T$, $\delta = 2$, if $X = U$, and $\delta = 3$, if $X = L$. Figure 1 highlights all the different times for doubly censored data as described in Example 1.

Assume for each individual, data is available on time-dependent covariates $Z(t)$. Suppose one is interested in investigating the relationship between T and $Z(t)$. Assume that T, L and U are continuous. Further, assume that given $Z(t)$, T and (L, U) are independent of each other but L and U are dependent with $P(L \leq U) = 1$. Suppose that the left and right endpoints of T, U, and L are independent of $Z(t)$. Let a_F and b_F denote the left and right support of T, and similarly, define (a_G, b_G) and (a_Q, b_Q) as the left and right support of L and U, respectively. Throughout this article, for identifiability of $S(t|Z(t)) = P(T > |Z(t))$, we assume that $a_G \leq a_F \leq a_Q$.

Cox's proportional hazards model (1972) has so far been the most popular model for the regression analysis of censored survival data. Kim et al. (2010) derived the asymptotic properties of the maximum likelihood estimator for the Cox's proportional hazards model with doubly censored data. Furthermore, Kim et al. (2013) proposed an EM algorithm for estimating parameters in the Cox's proportional hazard model. Using martingale arguments of Chen et al. (2002), Shen (2011) propose an estimator (denoted by $\tilde{\beta}$) for estimating regression coefficients of transformation model when L is always observed. Under Cox proportional hazards model, the proposed estimator is equivalent to the partial likelihood estimator for left-truncated and right-censored data if the left-censoring variables L were regarded as left-truncated variables (see Pan and Chappell (2002)). When L and U are always observed, Cai and Cheng (2004) proposed an alternative estimator under transformation with doubly censored data. Notice that the transformation model can be written as $S(t|Z(t)) = g\{h(t) + \beta^T Z(t)\}$, where the continuous, strictly decreasing link function $g(\cdot)$ is given, $h(\cdot)$ is a completely unspecified strictly increasing function satisfying $h(a_F) = -\infty$, and β is a $(p+1) \times 1$ vector of unknown regression coefficients. One disadvantage of transformation model is that it does not allow time-varying coefficients. A useful and flexible alternative to the transformation model is the Aalen's additive risk model (Aalen (1980, 1989, 1993), McKeague (1988); and Huffer and McKeague (1991)). The model is useful to deal with time-varying covariate effects in a simple manner, and it is important to know the temporal effects of the covariates on the time of interest.

Let $(L_i, X_i, \delta_i, Z_i(t))$ $(i = 1, \ldots, n)$ denote the observed sample, where $Z_i(t) = [1, z_{1i}(t), \ldots, z_{pi}(t)]^T$ is a $(p+1) \times 1$ vector of covariate for individual i. The additive risk model assumes that for individual i, the conditional hazard function at time t, given $Z_i(t)$, is

$$\lambda(t|Z_i(t)) = Z_i(t)^T \beta(t), \tag{1.1}$$

where $t \in [a_F, b_F]$, $\beta(t) = [\beta_0(t), \beta_1(t), \ldots, \beta_p(t)]^T$ is a $(p+1) \times 1$ vector of regression function, which are assumed to satisfy

$$\int_{a_G}^{b_Q} \beta_i(s) ds < \infty, \quad i = 0, \ldots, p.$$

In Section 2, we propose two estimation techniques for estimating

$$B(t) = [B_0(t), B_1(t), \ldots, B_p(t)]^T = \int_0^t \beta(s) ds.$$

The first type of estimators, including ordinary least squared (OLS) estimator and weighted least squared (WLS) estimators, are obtained using martingale arguments. The martingale approach are the same as those used in Shen (2014), where left-truncated and right-censored (LTRC) data are analyzed using Aalen's linear model. Furthermore, using the approach of Kim et al. (2013), we propose the second type of estimator, called the maximum likelihood (ML) estimator. The ML estimator is obtained via EM algorithm that treat the survival times of left censored observations as missing. In Section 3, a simulation study is conducted to investigate the performance of the two types of estimators.

Chapter 2

The Proposed Estimators

2.1 The OLS and WLS estimators

Let $Y_i(t) = I_{[L_i<t\leq X_i]}$ and $N_i(t) = I_{[X_i\leq t,\delta_i=1]}$. Let $\mathcal{F}_t$ denote the complete σ-field generated by $\{L_i, Z_i(x), Y_i(x), N_i(x); x \leq t\}$. Let $M_i(t) = N_i(t) - \int_0^t Y_i(s)Z_i(s)^T\beta(s)ds$. Under model (1.1), we have

$$E[dN_i(t)|Z_i(t), \mathcal{F}_{t-}] = P(t \leq T_i < t + dt, \delta_i = 1|Z_i(t), \mathcal{F}_{t-})$$

$$= P(t \leq T_i < t + dt, L_i < t \leq U_i|Z_i(t), \mathcal{F}_{t-}) = Y_i(t)Z_i^T(t)\beta(t),$$

it follows that $M_i(t)$ is a martingale process with respect to $\mathcal{F}_t$. Let $R(t) = [R_1(t), \ldots, R_n(t)]^T$ be a $n \times (p + 1)$ matrix, where $R_i(t) = [Y_i(t), Y_i(t)z_{1i}(t), \ldots, Y_i(t)z_{pi}(t)]^T$ is a $(p + 1) \times 1$ vector. Let $N(t) = [N_1(t), \ldots, N_n(t)]^T$ be a $n \times 1$ vector, where $N_i(t) = I_{[X_i\leq t,\delta_i=1]}$. Then $M(t) = N(t) - \int_0^t R(u)\beta(u)du$ is a $n \times 1$ vector of martingales. Thus, for doubly censored data, the following ordinary least squared (OLS) estimating function can be used for the estimation of $B(t)$ (see Aalen 1980):

$$\hat{B}(t) = \sum_{X_i\leq t} R^-(X_i)N(dX_i) = (\hat{B}_0(t), \hat{B}_1(t), \ldots, \hat{B}_p(t))^T,$$

where $N(dX_i) = N(X_i) - N(X_i-)$ and $R^-(X_i) = [R(X_i)^TR(X_i)]^{-1}R(X_i)^T$ is a generalized inverse of $R(X_i)$. Under the assumptions that (a) $a_G \leq a_F \leq b_Q$, (b) $Z_i(t)$ is bounded for $t \in [a_F, b_F]$ and the limit of $n^{-1}R(t)^TR(t)$ is nonsingular, it follows by Aalen (1980) that $\hat{B}(t) - B(t)$ coincides in probability, as $n \to \infty$, with the $(p+1)$-variate martingale $\tilde{M}(t) = \int_{a_F}^t R^-(u)dM(u)$. The predictable covariation process $\langle\tilde{M}\rangle(t) = \int_{a_F}^t R^-(u)[\text{Diag } H(u)][R^-(u)]^Tdu$, where Diag $H(u)$ is the $n \times n$ diagonal matrix with (i, i)th element $Z_i(u)^T\beta(u)Y_i(u)$. The variance $\langle\tilde{M}\rangle(t)$ can be estimated by $\hat{\Sigma}(t) = \int_{a_F}^t R^-(u)[\text{Diag } dN(u)][R^-(u)]^Tdu$, where Diag $dN(u)$ is the $n\times n$ diagonal matrix with (i, i)th element $dN_i(u)$. Hence, $\sqrt{n}(\hat{B}(t)-B(t))$ converges in distribution to a zero-mean Gaussian process and a pointwise $100(1-\alpha)\%$ confidence intervals for $B_j(t)$ $(j = 0, \ldots, p)$ can be calculated using the formula $\hat{B}_j(t) \pm z_{\alpha/2}\hat{\sigma}_j(t)$, $j = 1, \ldots, p$, where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal, $\hat{B}_j(t)$ $(j = 0, \ldots, p)$ is the j^{th} cumulative regression function of the OLS estimator $\hat{B}(t)$, and $\hat{\sigma}_j^2(t)$ is the j^{th} diagonal entry of the covariance matrix $\hat{\Sigma}$.

For right censored data, to obtain a more efficient estimator, Huffer and McKeague (1991) and McKeague (1988) considered a weighted least-squared generalized inverse $R_w^-(u) = [R(u)^TW(u)R(u)]^{-1}R(u)^TW(u)$, where $W(u)$ is an $n \times n$ diagonal matrix taken to have the (j, j)th element, $W_{jj}(u)$, proportional to the inverse of the variance of $dM_i(t)$. A kernel-smoothed estimator of $\beta_j(t)$ is needed for estimating the variance of $dM_i(t)$, which is given by $\lambda(t|Z_i(t)) = Z_i(t)^T\beta(t)$. We consider a two-step estimation procedure as follows.

The first step:

Using $\hat{B}(u)$, a kernel-smoothed estimator of $\beta_j(t)$ is given by

$$\hat{\beta}_j(t) = \int_{a_F}^{b_F} \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) d\hat{B}_j(u),$$

where $K(\cdot)$ is a left-continuous function on $(0,1]$ with $\int_{(0,1]} K(u)du = 1$ and h_n is a positive bandwidth parameter that tends to 0 as $n \to \infty$.

The second step:

The estimators $\hat{\beta}_j(u)$ $(j = 1, \ldots, p)$ are used to estimate the weight function $W_{jj}(u) = [\lambda_j(u|Z_j(u))]^{-1}$ by $\hat{W}_{jj}(u) = [\hat{\beta}_0(u) + \sum_{j=1}^{p} z_{ji}(u)\hat{\beta}_j(u)]^{-1}$. These weights are then used to compute $\hat{R}_w^-(u) = [R(u)^T \hat{W}(u) R(u)]^{-1} R(u)^T \hat{W}(u)$, where $\hat{W}(u)$ is the diagonal matrix with (j,j)th element, $\hat{W}_{jj}(u)$.

Thus, for doubly censored data, we obtain the following weighted least squared (WLS) estimator:

$$\hat{B}_w(t) = \sum_{X_i \le t} \hat{R}_w^-(X_i) N(dX_i) = (\hat{B}_{w0}(t), \hat{B}_{w1}(t), \ldots, \hat{B}_{wp}(t))^T.$$

By Theorem 4.1.2 of Ramlau-Hansen (1983), if $h_n \to 0$ and $nh_n \to \infty$, then the smoothed least squared $\hat{\beta}_j(u)$ is uniformly consistent over $[h_n, b_F]$. It follows that the estimated matrix $\hat{W}(u)$ is a uniformly consistent estimator of the true weight $W(u)$ with (j,j)th element $W_{jj}(u) = [\beta_0(u) + \sum_{j=1}^{p} z_{ji}\beta_j(u)]^{-1}$ over $[h_n, b_F]$. Under the assumption that the limit of $n^{-1}R(t)^T \hat{W}(t) R(t)$ is nonsingular, it follows by Aalen (1980) that $\hat{B}_w(t) - B(t)$ coincides in probability, as $n \to \infty$, with the $(p+1)$-variate martingale $\tilde{M}_w(t) = \int_{a_F}^{t} R_w^-(X_i) dM_w(u)$. The predictable covariation process $\langle \tilde{M}_w \rangle(t) = \int_{a_F}^{t} R_w^-(u)[\text{Diag } H(u)][R_w^-(u)]^T du$, which can be estimated by $\hat{\Sigma}_w(t) = \int_{a_F}^{t} \hat{R}_w^-(u)[\text{Diag } dN(u)][\hat{R}_w^-(u)]^T du$. Thus, pointwise $100(1-\alpha)\%$ confidence intervals for $B_j(t)$ can be calculated using the formula $\hat{B}_{wj}(t) \pm z_{\alpha/2}\hat{\sigma}_{wj}(t)$, where $\hat{B}_{wj}(t)$ $(j = 0, \ldots, p)$ is the j^{th} cumulative regression function of the OLS estimator $\hat{B}_w(t)$, and $\hat{\sigma}_{wj}^2(t)$ is the j^{th} diagonal entry of the covariance matrix $\hat{\Sigma}_w$.

Similar to the approach of Shen (2014), we choose bandwidth using the integrated square error (ISE) for an estimator $\hat{\beta}_j(t)$, which is defined as

$$\begin{aligned} \text{ISE}(\hat{\beta}_j) &= \int_{t_1}^{t_2} (\hat{\beta}_j(s) - \beta_j(s))^2 ds \\ &= \int_{t_1}^{t_2} (\hat{\beta}_j(s))^2 dt - 2\int_{t_1}^{t_2} \hat{\beta}_j(s)\beta_j(s) ds + \int_{t_1}^{t_2} (\beta_j(s))^2 dt, \qquad (2.1) \end{aligned}$$

where $[t_1, t_2] \subset [a_F, b_F]$. Since the last term of (2.1) does not depend on bandwidth or kernel function selected for $\hat{\beta}_j(s)$, the ISE is proportional to

$$U_j(h_n) = \int_{t_1}^{t_2} (\hat{\beta}_j(s) - \beta_j(s))^2 ds$$

$$= \int_{t_1}^{t_2} (\hat{\beta}_j(s))^2 ds - 2\int_{t_1}^{t_2} \int_{a_F}^{b_F} \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) I_{[u \neq t]} d\hat{B}_j(u) dB_j(s)$$

$$\simeq \hat{U}_j(h_n) = \int_{t_1}^{t_2} (\hat{\beta}_j(s))^2 ds - 2\int_{t_1}^{t_2} \int_{a_F}^{b_F} \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) I_{[u \neq t]} d\hat{B}_j(u) d\hat{B}_j(s).$$

To minimize $\hat{U}_j(h_n)$, we need to specify lower and upper bound for h_n, between which we can identify the value that minimize $\hat{U}_j(h_n)$.

2.2 The ML Estimator

In Section 2.1, the proposed based estimators OLS and WLS are based on martingale process, which requires that L is always observed. Furthermore, they may have the disadvantage of less efficiency compared with ML approach. This argument is explained as follows.

Under model (1.1), the likelihood function is given by

$$L(\beta) = \prod_{i=1}^{n} [\lambda(X_i|Z_i(t))]^{I_{[\delta_i=1]}} \left[e^{-\int_0^{X_i} \lambda(u|Z_i(u))du} \right]^{I_{[\delta_i \neq 3]}} \left[1 - e^{-\int_0^{X_i} \lambda(u|Z_i(u))du} \right]^{I_{[\delta_i=3]}}.$$

Given L_i's, the conditional likelihood for the observations with $\delta_i \neq 3$ is given by

$$L_C(\beta) = \prod_{\delta_i \neq 3} [\lambda(X_i|Z_i(t))]^{I_{[\delta_i=1]}} \left[e^{-\int_{L_i}^{X_i} \lambda(u|Z_i(u))du} \right]^{I_{[\delta_i \neq 3]}}.$$

Then, by Greenwood and Wefelmeyer (1990, 1991) and Sasieni (1992), consider a one-dimensional parametric submodel with $\beta(t) = \alpha_\eta(t)$ and $d\alpha_\eta(t)/d\eta = a(t) = (a_1(t), \ldots, a_p(t))^T$ so that $\partial \log \lambda(X_i|Z_i(t))/\partial\eta = Z_i(t)^T a(t)/\lambda_i(t|Z_i(t)) = w_i(t|Z_i(t))$. The score function by differentiating the conditional log likelihood of L_C with respect to η is given by

$$\dot{l}_c(\eta) = \sum_{\delta_i \neq 3} \left\{ \int_{a_F}^{b_F} \frac{a(t)^T Z_i(t)}{\lambda_i(t|Z_i(t))} dN_i(t) - \int_{a_F}^{b_F} I_{[L_i < t \leq X_i]} \frac{a(t)^T Z_i^T(t)}{\lambda_i(t|Z_i(t))} Z_i^T(t) dB(t) \right\}$$

$$= \int_{a_F}^{b_F} \sum_{\delta_i \neq 3} w_i(t|Z_i(t)) dM_i(t).$$

Setting $\dot{l}_c(\eta) = 0$ and solving for $B(t)$, say $\hat{B}_c(t)$, we obtain

$$\int_{a_F}^{b_F} a^T(t) \sum_{\delta_i \neq 3} \frac{Z_i(t)}{\lambda_i(t)|Z_i(t))} dN_i(t) = \int_{a_F}^{b_F} a^T(t) \sum_{\delta_i \neq 3} I_{[L_i < t \leq X_i]} \frac{Z_i(t) Z_i^T(t)}{\lambda_i(t)|Z_i(t))} d\hat{B}_c(t)$$

for all such $a(t)$. Let $R_w^-(u) = [R(u)^T W(u) R(u)]^{-1} R(u)^T W(u)$. Substituting $\hat{B}_c(t) = \sum_{X_i \leq t} R_w^-(X_i) N(dX_i)$ into above equation gives a solution for any vector of function $a(t)$. However, $\hat{B}_c(t)$ is not an estimator since $W(u)$ is unknown. Replacing $W(t)$ by $\hat{W}(t)$, one is led to consider the WLS estimator $\hat{B}_w(t)$. Thus, the WLS estimator is not efficient since it is an approximate conditional maximum likelihood type estimator.

In this section, using empirical likelihood, we consider an alternative approach for estimating $B(t)$. This approach has the advantage that it does not requires that the left censored variable L is always observed. To construct an empirical likelihood for the Aalen's model, we first discretize the function $B(t)$ by assuming that $B(t)$ has jump only at times in the set $\mathcal{S}$, where $\mathcal{S}$ consists of all X_i with $\delta_i = 1$ and $X_{(1)}$, the smallest order statistic among $X_1, \ldots, X_n$, if $X_{(1)}$ is left censored. The reason for the choice of $\mathcal{S}$ is that maximum empirical likelihood (ML) estimator without a covariate becomes the self-consistent estimator (Mykland and Ren (1996)). Let $t_1 < t_2 \cdots < t_d$ denote the distinct point in the set $\mathcal{S}$. Let $\Lambda(t|b; Z_i(t)) = \sum_{k=1}^{d} I_{[t_k \le t]} Z_i(t_k)^T dB(t_k)$ and $\Delta\Lambda(t|b; Z_i(t)) = \Lambda(t|b; Z_i(t)) - \Lambda(t-|b; Z_i(t))$, where $b = (b(t_1), \ldots, b(t_d))^T$, $b(t_k) = dB(t_k) = (dB_0(t_k), dB_1(t_k), \ldots, dB_p(t_k))^T$. Then an empirical likelihood is given by

$$L(b) = \prod_{i=1}^{n} [\Delta\Lambda(X_i|b; Z_i(X_i)) \exp\{-\Lambda(X_i|b; Z_i(X_i))\}]^{I_{[\delta_i=1]}}$$

$$[\exp\{-\Lambda(X_i|b; Z_i(X_i))\}]^{I_{[\delta_i=2]}} [1 - \exp\{(-\Lambda(X_i|b; Z_i(X_i))\}]^{I_{[\delta_i=3]}},$$

and the log empirical likelihood becomes

$$l(b) = \sum_{i=1}^{n} I_{[\delta_i=1]} [\log(\Delta\Lambda(X_i|b; Z_i(X_i)) - \Lambda(X_i|b; Z_i(X_i))]$$

$$- \sum_{i=1}^{n} I_{[\delta_i=2]} \Lambda(X_i|b; Z_i(X_i)) + \sum_{i=1}^{n} I_{[\delta_i=3]} [\log(1 - \exp\{(-\Lambda(X_i|b; Z_i(X_i))\}].$$

The ML estimator of b can be obtained by maximizing the empirical likelihood $l(b)$ with respect to b_j under the constraint $\Delta\Lambda(t|dB; Z_i(t)) \ge 0$ for $t \in \mathcal{S}$. However, one may encounter computational difficulties since the number of parameters maximized is proportional to the number of uncensored observations which can be huge when the sample size n is large. Hence, using the approach of Kim et al. (2013), we consider an alternative approach, namely EM algorithm, for obtaining ML of b. The main idea of the EM algorithm is to treat the survival times of left censored observations as missing. Notice that the empirical likelihood implies that given X_i, $\delta_i = 3$ and $Z_i(t)$ $(t \le X_i)$ the conditional distribution of the unobserved lifetime, say T_i, is discrete having mass on $\{t_k : t_k \le X_i\}$ and

$$P(T_i = t_k | X_i, \delta_i = 3, b) = \frac{p_i(t_k; b)}{\sum_{t_l \le X_i} p_i(t_l; b)}, \tag{2.2}$$

where $p_i(t_j; b) = \Delta\Lambda(t_k|b; Z_i(t_k)) \exp\{-\Lambda(t_k|b; Z_i(t_k))\}$. Define $\tilde{X}_i = X_i$ if $\delta_i = 1$ or 2 and $\tilde{X}_i = T_i$ if $\delta_i = 3$. Similarly, let $\tilde{\delta}_i = \delta_i$ if $\delta_i = 1$ or 2 and $\tilde{\delta}_i = 1$ if $\delta_i = 3$. For subject i, define $\Delta N_{ik} = I_{[\tilde{X}_i = t_k, \tilde{\delta}_i = 1]}$ and $Y_{ik} = I_{[\tilde{X}_i \geq t_k]}$. If T_i were observed then the complete empirical likelihood based on $(\tilde{X}_i, \tilde{\delta}_i)$ $(i = 1, \ldots, n)$ is given by

$$L_c(b) = \prod_{i=1}^{n} \prod_{k=1}^{d} [\Delta\Lambda(t_k|b; Z_i(t_k))]^{\Delta N_{ik}} \exp\{-Y_{ik}\Lambda(t_j|b; Z_i(t_k))\},$$

and the corresponding log empirical likelihood becomes

$$l_c(b) = \sum_{i=1}^{n} \sum_{k=1}^{d} \{\Delta N_{ik}[\log(\Delta\Lambda(t_k|b; Z_i(t_k))] - Y_{ik}\Lambda(t_k|b; Z_i(t_k))\}.$$

E-step:

Define $D = \{(X_1, \delta_1), \cdots, (X_n, \delta_n)\}$. The E-step is to calculate

$$l_E(b) = E_{N,Y}[l_c(b)|D, b^{(m)})] = \sum_{i=1}^{n} \sum_{k=1}^{d} \{E_{N,Y}[\Delta N_{ik}|D, b^{(m)}][\log(\Delta\Lambda(t_k|b^{(m)}; Z_i(t_k))]$$

$$-E_{N,Y}[Y_{ik}|D, b^{(m)}]\Lambda(t_k|b^{(m)}, Z_i(t_k))\},$$

where $E_{N,Y}[\cdot|D, b^{(m)}]$ is the conditional expectation of N_{ik}'s and Y_{ik}'s given the data and the m^{th} iteration parameter values $b^{(m)} = (b^{(m)}(t_1), \ldots, b^{(m)}(t_d))^T$, $b^{(m)}(t_k) = (dB_0^{(m)}(t_k), dB_1^{(m)}(t_k), \ldots, dB_p$ For $\delta_i = 1$, both N_{ik} and Y_{ik} are known and hence no expectation is needed. For $\delta_i = 3$, the conditional expectation (2.2) implies that

$$e_{ik} = E_{N,Y}[\Delta N_{ik}|D, b^{(m)}] = \frac{p_{ik}(b^{(m)})}{\sum_{t_l \leq X_i} p_{il}(b^{(m)})} \quad \text{for } t_k \leq X_i,$$

and $e_{ik} = 0$ for $t_k > X_i$, where

$$p_{ik}(b^{(m)}) = \Delta\Lambda(t_k|b^{(m)}; Z_i(t_k)) \exp\{-\Lambda(t_k|b^{(m)}; Z_i(t_k))\}.$$

Similarly, we have

$$r_{ik} = E_{N,Y}[Y_{ik}|D, b^{(m)}] = 1 - \sum_{l=1}^{k-1} e_{ik}$$

for $\delta_i = 3$ and $t_k \leq X_i$, and $r_{ik} = 0$ for $t_k > X_i$.

M-step:

The M-step updates the parameter values $b^{(m)}$ by $b^{(m+1)}$ which maximizes $l_E(b)$. Next, we explain how to obtain $b^{(m+1)}$.

Let $e_{ik} = \Delta N_{ik}$ and $r_{ik} = Y_{ik}$ for $\delta_i = 1$ or $\delta_i = 2$. Let $\tilde{Y}_i(t) = \sum_{t_k \geq t} r_{ik}$, $\tilde{N}_i(t) = \sum_{t_k \leq t} e_{ik}$, $\tilde{R}_i(t) = [\tilde{Y}_i(t), \tilde{Y}_i(t) z_{1i}(t), \ldots, \tilde{Y}_i(t) z_{pi}(t)]^T$, and a $\tilde{N}(t) = [\tilde{N}_1(t), \ldots, \tilde{N}_n(t)]^T$. Using $B^{(m)}(u) = \sum_{t_k \leq u} b^{(m)}(t_k)$, a kernel-smoothed estimator of $\beta_j^{(m)}(t)$ is given by

$$\beta_j^{(m)}(t) = \int_{a_F}^{b_F} \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) dB_j^{(m)}(u).$$

Let $(R_w^{(m)}(u))^- = [R(u)^T W^{(m)}(u) R(u)]^{-1} R(u)^T W^{(m)}(u)$, where $W^{(m)}(u)$ is an $n \times n$ diagonal matrix taken to have the (j,j)th element, $W_{jj}^{(m)}(u) = [\beta_0^{(m)}(u) + \sum_{j=1}^{p} z_{ji} \beta_j^{(m)}(u)]^{-1}$. Then, by Greenwood and Wefelmeyer (1990, 1991) and Sasieni (1992), an approximate maximum likelihood estimate $B^{(m+1)}$ is given by

$$B^{(m+1)}(t) = \sum_{X_i \leq t} (\tilde{R}_w^{(m)}(t_k))^- \tilde{N}(dt_k).$$

Based on $B^{(m+1)}$, a kernel-smoothed estimator of $b^{(m+1)}(t)$ is given by

$$b^{(m+1)}(t) = \int_{a_F}^{b_F} \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) dB_j^{(m+1)}(u).$$

Let $\hat{b}_M = (\hat{b}_M(t_1), \ldots, \hat{b}_M(t_d))^T$ denote the converged estimator using EM algorithm, where $\hat{b}_M(t_k) = (\hat{b}_{M0}(t_k), \hat{b}_{M1}(t_k), \ldots, \hat{b}_{Mp}(t_k))^T$.

Chapter 3

Simulation Studies

A simulation study is conducted to compare the performance of $\hat{B}_j(t)$, $\hat{B}_{wj}(t)$ and $\hat{B}_{Mj}(t)$. We consider the simulation model $\lambda(t|z_{1i}) = 1 + z_{1i}^T\beta_1(t)$, where $\beta_1(t) = t$ and z_{1i}'s are generated from discrete distribution with $P(z_{1i} = j) = 0.25$ for $j = 1, 2, 3, 4$. The left censoring time L is independent of the failure time and exponentially distributed with parameter $\theta_L = 2, 4, 8$ (mean=$1/\theta_L$). The right censoring variable U was generated from $L+1$. Sample size is set at 100 and 200. The optimum bandwidth h_n is selected from the grid $\{0.15, 0.2, 0.25, 0.3, 0.35, 0.4\}$ by minimizing $\hat{U}_j(h_n)$, where we set $[t_1, t_2] = [T_{(1)}, T_{(n_d)}]$, $T_{(1)}$ and $T_{(n_d)}$ is the smallest and largest uncensored observation of X_i's, respectively. Sample size is set at 100 and 300. The replication time is 1000. Tables 1 and 2 shows the mean average biases (bias) over all simulation runs and empirical root mean squared errors (rmse) of $\hat{B}_i(t_p)$, $\hat{B}_{wi}(t_p)$ and $\hat{B}_{Mi}(t_p)$ at $p = 0.25, 0.75$. Table 1 also shows the ratio (denoted by ratio) of the simulated root mean squared error (rmse) of $\hat{B}_{Mi}$ to that of $\hat{B}_{wi}$.

Table 1. Simulated biases and rmse of $\hat{B}_0(t_p)$, $\hat{B}_{w0}(t_p)$ and $\hat{B}_{M0}(t_p)$

						$\hat{B}_0(t_p)$		$\hat{B}_{w0}(t_p)$		$\hat{B}_{M0}(t_p)$		
θ_L	p	n	h_n	p_L	p_R	bias	rmse	bias	rmse	bias	rmse	ratio
2	0.25	100	0.30	0.43	0.13	0.018	0.404	0.019	0.395	0.012	0.387	0.98
2	0.25	300	0.25	0.43	0.13	0.009	0.205	0.007	0.203	0.005	0.196	0.97
2	0.75	100	0.30	0.43	0.13	-0.094	0.653	-0.085	0.643	-0.045	0.631	0.98
2	0.75	300	0.25	0.43	0.13	-0.037	0.293	-0.013	0.288	-0.010	0.280	0.97
4	0.25	100	0.25	0.25	0.17	-0.019	0.320	-0.016	0.316	-0.012	0.307	0.97
4	0.25	300	0.20	0.25	0.17	-0.016	0.214	-0.012	0.209	-0.005	0.198	0.95
4	0.75	100	0.30	0.25	0.17	-0.067	0.441	-0.056	0.430	-0.031	0.413	0.96
4	0.75	300	0.25	0.25	0.17	-0.014	0.270	0.008	0.261	-0.007	0.249	0.95
8	0.25	100	0.25	0.13	0.22	-0.007	0.244	-0.008	0.238	0.013	0.226	0.95
8	0.25	300	0.20	0.13	0.22	-0.005	0.164	-0.006	0.162	0.007	0.151	0.93
8	0.75	100	0.30	0.13	0.22	-0.022	0.461	-0.016	0.453	0.018	0.437	0.96
8	0.75	300	0.25	0.13	0.22	-0.015	0.215	0.006	0.206	0.010	0.193	0.94

Table 2. Simulated biases and rmse of $\hat{B}_1(t_p)$, $\hat{B}_{w1}(t_p)$ and $\hat{B}_{M1}(t_p)$

						$\hat{B}_1(t_p)$		$\hat{B}_{w1}(t_p)$		$\hat{B}_{M1}(t_p)$		
θ_L	p	n	h_n	p_L	p_R	bias	rmse	bias	rmse	bias	rmse	ratio
2	0.25	100	0.30	0.43	0.13	-0.014	0.168	-0.015	0.162	-0.019	0.155	0.97
2	0.25	300	0.25	0.43	0.13	-0.006	0.079	-0.006	0.077	-0.009	0.074	0.96
2	0.75	100	0.30	0.43	0.13	0.024	0.278	0.021	0.270	-0.017	0.265	0.98
2	0.75	300	0.25	0.43	0.13	-0.005	0.119	0.001	0.117	-0.006	0.114	0.97
4	0.25	100	0.25	0.25	0.17	-0.015	0.131	0.010	0.127	0.014	0.120	0.94
4	0.25	300	0.20	0.25	0.17	-0.011	0.075	0.007	0.075	0.006	0.071	0.95
4	0.75	100	0.30	0.25	0.17	0.016	0.206	0.012	0.200	0.010	0.193	0.97
4	0.75	300	0.25	0.25	0.17	-0.002	0.125	-0.005	0.122	0.004	0.116	0.95
8	0.25	100	0.25	0.13	0.22	-0.007	0.122	0.008	0.118	0.007	0.112	0.95
8	0.25	300	0.20	0.13	0.22	-0.005	0.066	0.002	0.062	0.002	0.057	0.92
8	0.75	100	0.30	0.13	0.22	0.008	0.207	0.007	0.205	0.011	0.194	0.95
8	0.75	300	0.25	0.13	0.22	0.005	0.095	0.002	0.091	0.006	0.085	0.93

Based on the results of Tables 1 and 2, we have the following conclusions:

(1) For the estimation of $B_i(t)$, the standard deviations of all the three estimators increase as the proportion of censoring (i.e. $p_L + p_R$) increases. The optimum bandwidth decreases as sample size increases.

(2) In term of rmse, the ML estimator $\hat{B}_{Mi}(t)$ outperforms both $\hat{B}_{wi}(t)$ and $\hat{B}_i(t)$. The ratio of the simulated rmse of $\hat{B}_{Mi}$ to that of $\hat{B}_{wi}$ ranges from 0.92 to 0.98.

(3) Although the WLS estimator is superior to the OLS estimator, we encountered moderate gain in efficiency when $n = 100$. Bandwidth selection using the ISE scheme appears to have some room for improvement.

Chapter 4

Conclusions

In this article, for doubly censored data, we have considered three estimators, namely, the OLS, WLS and ML estimators. For doubly censored data, we pointed out that the WLS estimator is an approximate conditional maximum likelihood type estimator. Simulation results indicate that the ML estimator performs better than OLS and WLS estimators. Further research is needed in establishing the asymptotic distribution of the ML estimator.

References

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations, *Scandinavian Journal of Statistics*, **5**, 141-150.

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting process. In:Klonecki, W., Kozek, A., Rosiski, J. (Eds). Lecture Notes on Mathematical Statistics and Probability, Vol. 2. Springer, New York, pp. 1-25.

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

Aalen, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine*, **17**, 1569-1588.

Cai, T. and Cheng, S. (2002). Semiparametric regression analysis for doubly censored data. *Biometrika*, **91**, 277-290.

Chen, K., Jin, Z and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, **89**, 659-668.

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

Greenwood, P. E. and Wefelmeyer, W. (1990). Efficiency of estimators for partially specified filtered models. *Stochastic Processes and their Applications*, **36**, 353-370.

Greenwood, P. E. and Wefelmeyer, W. (1991). Efficient estimating equations for nonparametric filtered models. In: *Statistical Inferences in Stochastic Processes 1*, (Prabhu and Basawa, eds.), 107-141.

Huffer, F. W. and Mckeague, F. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, **86**, 114-129.

Kim, Y., Kim, B. and Jang, W. (2010). Asymptotic properties of the maximum likelihood estimator for the proportional hazard model with doubly censored data. *Journal of Multivariate Analysis*, **101**, 1339-1351.

Kim, Y., Kim, J. and Jang, W. (2013). An EM algorithm for the proportional hazards model with doubly censored data. *Computational Statistics and Data Analysis*, **57**, 41-51.

McKeagues, I. W. (1988). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *in Statistical Inference from Stochastic Processes*, page 139-152.

Mykland, P, A. and Ren, J. (1996), Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *Ann. Statist.* **24**, 1740-1764.

Pan, W. and Chappell, R. (2002). Estimation in the Cox proportional hazards model with left-truncated and interval-censored data. *Biometrics*, **58**, 64-70.

Ramula-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics*, **11**, 453-466.

Ranga, R. R. (1963). The law of large numbers for $D[0,1]$-valued random variables, *Theory of Probability and its applications*, **8**, 70-74.

Rebolledo, R (1980). Central limit theorems for local martingals *Zeitschrift für Wahrsheinlichkeitstheorie und Verwandte Gebiete*, **51**, 269-286.

Sasieni, P. D. (1992). Information bounds for the additive and multiplicative intensity models. In: *Survival Analysis of the art*, (J.P. Klein and P.K. Goel, eds.), 249-265, Kluwer.

Shen, P.-S. (2011). Semiparametric analysis of transformation models with doubly censored data. *Journal of Applied Statistics*, **38**, 675-682.

Shen, P.-S. (2014). Aalen's additive risk model for left-truncated and right-censored data. *Communications in Statistics-Simulation and Computation*, **43**, 1006-1019.