Semiparametric Analysis of Transformation Models with Left-truncated and Right-Censored Data

Director :Pao-sheng Shen

Graduate student: Yu-ching Wu

Department of Statistics
Tunghai University, Taichung, Taiwan, 40704
psshen@thu.edu.tw

Abstract

We analyze left-truncated and right-censored (LTRC) data using semiparametric transformation models. It is demonstrated that the approach of Chen et al. (2002) can be extended to LTRC data. A simulation study is conducted to investigate the performance of the proposed estimators.

Key Words: Left truncation; Semiparametric transformation model.

Chapter 1. Introduction

Left-truncated and right-censored (LTRC) data often arise in epidemiology and individual follow-up studies (see Wang, 1991). Their importance stems from the common use of prevalent cohort study designs to estimate survival from onset of a specified disease. Consider the following applications.

Example 1.1: prevalent cohort data

In epidemiology, a prevalent cohort is defined as a group of diseased individuals who are recruited for a prospective study. Suppose that the disease population in a certain area is a representative sample from a large disease population. The target interest of a research project is to study the natural history of the disease for individuals who developed the disease during the calendar time period (τ_0, τ), $\tau_0 < \tau$. Consider the sampling under which all of the individuals in the area who have experienced a first event (such as being diagnosed as having Alzheimer's disease or AIDS) between τ_0 and τ and have not experienced a second event (such as death) are recruited at the time τ for a prospective follow-up study. The follow-up study is terminated at τ^* $(\tau^* > \tau)$. Suppose that the initial time of the first event (denoted by T_s) can be quite accurately determined. For example, the dates of Alzheimer's disease or vascular dementia onset can be provided by the caregivers of those patients (e.g. Canadian study of Health and Aging (see The CSHA working group, 1994). Let T be the time from T_s to death. Let V denote the time from T_s to τ, and C denote the time from T_s to censoring. Note that the censoring time can be written as $C = \min(C_1, C_2)$ and $P(C \geq V) = 1$, where $C_1 = V + \tau^* - \tau$ denotes the time from onset of disease to the end of study, and C_2 denotes the time from onset of disease to drop-out or death due to other causes. Left truncation arises because those individuals who have been diagnosed as having the disease and die prior to time of recruitment (τ) are excluded from the cohort. Assume for each individual, data is available on $Z_1, \ldots, Z_p$ covariates (e.g. sex, education, smoking status, myocardial infarction). It is important to investigate the association between these covariates Z_i's and survival rate. Figure 1 highlights all the different times for LTRC data as described in Example 1.1

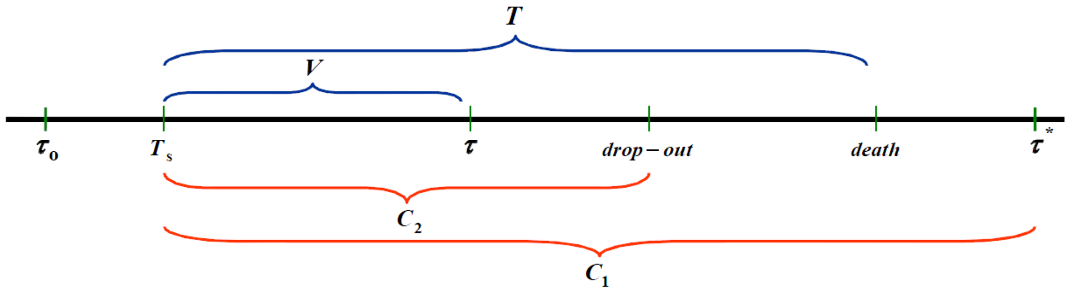

Figure 1. Schematic depiction of LTRC data described in Example 1.1

Example 1.2 (Channing House data)

Channing House is a retirement centre in Palo Alto, California. The data were collected between the opening of the house in calendar time $\tau = 1964$ (in years) and $\tau^* = 1975.5$ (Hyde, 1980; Klein and Moeschberger, 1997). In that time 97 men and 365 women passed through the center. A distinctive feature of these individuals was that all were covered by a health care program provided by the center which allowed for easy access to medical care without additional financial burden to the resident. Suppose that an individual must survival to an age of 60 (in years) to enter the retirement community. Let τ_0 denote the calendar time 1915.5 (in years). Define a population as the resident who were born before τ_0 and covered by a health cure program provided the center. The target interest of a research project is to study the survival time for the population defined above. Let T_b denote the calendar time (in years) of birth. Let T_e denote the calendar time (in years) of entry. Let $V = T_e - T_b$ denote the entry age into the Channing House. Let T be the time from T_b to death. It is clear that only subjects with entry age (V) smaller than or equal to the age on death (T), i.e. $T \geq V$, can become part of the sample. Moreover, a large number of the observations were right-censored due to the residents being alive on 1975.5 (termination of the follow-up). Let C denote the age at the end of study. Hence $C = V + \tau^* - T_e$ and $P(C \geq V) = 1$. Figure 2 highlights all the different times for LTRC data as described in Example 1.2

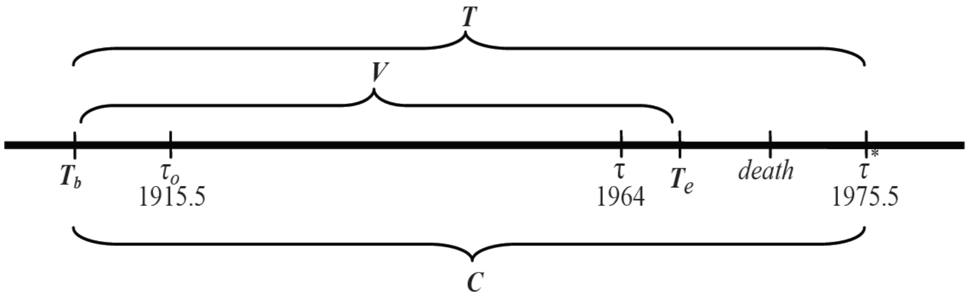

Figure 2. Schematic depiction of LTRC data described in Example 1.2

Following the notations in Examples 1.1 and 1.2, let (T, C, V) denote the lifetime, censoring time and truncation time, respectively. Let $Z = [Z_1, \ldots, Z_p]^T$ represent a $p \times 1$ vector of covariates. Assume that T, V and C are continuous. Further, assume that given Z, T and (V, C) are independent of each other but V and C are dependent with $P(C \geq V) = 1$. For LTRC data, one can observe nothing if $T < V$ and observe (X, V, δ, Z), with $\delta = I_{[T<C]}$ and $X = \min(T, C)$, if $T \geq V$. In Example 1, the calendar time of the potential censoring point must be greater than τ, since only those individuals in the follow-up study might be observed subject to right censoring. Therefore, the relationship $C \geq V$ is always satisfied, i.e. $P(C \geq V) = 1$.

We consider the following transformation model:

$$S(t|Z) = g\{h(t) + \beta^T Z\}, \tag{1.1}$$

where $S(t|Z) = P(T > t|Z)$ is the survival function of T given Z, the continuous, strictly decreasing link function $g(\cdot)$ is given or specified up to a finite-dimensional parameter, $h(\cdot)$ is a completely unspecified strictly increasing function, and β is a $p \times 1$ vector of unknown regression coefficients. Note that when $g(\cdot) = \exp\{-\exp(\cdot)\}$, (1.1) gives the Cox proportional hazard model (Cox, 1972). The other family of linear transformation models whose link functions are indexed by a single parameter η is given by $g(\cdot) = [1/\{1 + \eta\exp(\cdot)\}]^{-1/\eta}$ $(\eta > 0)$, which corresponds to the proportional odds model (Bennett, 1983; Murphy et al., 1997; Ying and Prentice,

1991) when $\eta = 1$. Further, note that model (1.1) has an equivalent form

$$h(T) = -\beta^T Z + \epsilon,$$

where the distribution of the error ϵ is $P(\epsilon \leq x) = F_\epsilon(x) = 1 - g(x)$.

When $g(\cdot)$ is completely specified, Chen et al. (2002) proposed an estimation procedure for the analysis of right-censored data. The procedure proposed by Chen et al. (2002) is easily implemented numerically and the estimator is the same as the Cox partial likelihood estimator in the case of the proportional hazards model. In Section 2, it is demonstrated that the approach of Chen et al. (2002) can be easily extended to LTRC data. In Section 3, simulation studies are conducted to investigate the performance of the proposed estimator.

Chapter 2. The Proposed Estimators

2.1 The Estimator based on the approach of Chen et al.

Let $F(t|Z) = P(T \leq t|Z)$ denote the cumulative distribution function of T given Z. Let $Q(t|Z) = P(C \leq t|Z)$ and $G(t) = P(V \leq t|Z)$ denote the cumulative distribution functions of C and V given Z, respectively. Suppose that the left and right endpoints of T are independent of Z. Let a_F and b_F denote the left and right endpoints of F, and similarly, define (a_G, b_G) and (a_Q, b_Q) as the left and right endpoint of V, and C, respectively. Throughout this article, for identifiabilities of $F(t|Z)$, we assume that

$$a_G = a_F = a_Q = 0, b_G \leq \min(b_F, b_Q) \text{ and } b_F \leq b_Q. \tag{2.1}$$

Let $(X_i, V_i, \delta_i, Z_i)$ $(i = 1, \ldots, n)$ be the observed truncated sample. Let $Y_i(t) = I_{[V_i \leq t \leq X_i]}$ and $N_i(t) = I_{[X_i \leq t, \delta_i = 1]}$. Let $\mathcal{F}(t)$ denote the complete σ-field generated by

$$\{V_i, Z_i, Y_i(x), I_{[V_i \leq X_i]}, \delta_i I_{[V_i < X_i \leq t]}, I_{[V_i < X_i \leq x]}, x \leq t; i = 1, \ldots, n\}.$$

Let $p(Z_i) = P(V \leq T|Z_i)$. Note that $E[Y_i(t)|Z_i] = P(V_i \leq t \leq X_i|Z_i) = p(Z_i)^{-1}P(V \leq t \leq C|Z_i)P(T \geq t|Z_i)$, and $E[N_i(t)|Z_i] = p(Z_i)^{-1}P(V \leq T \leq C, T \leq t|Z_i)$. Let $\lambda_\epsilon(\cdot)$ and $\Lambda_\epsilon(\cdot)$ denote the hazard and cumulative hazard functions of ϵ, respectively. Let $h_0(\cdot)$ and β_0 denote the true values of $h(\cdot)$ and β, respectively. Let $M_i(t) = N_i(t) - \int_0^t Y_i(s) d\Lambda_\epsilon(\beta_0^T Z_i + h_0(s))$. It follows that $M_i(t)$

is a martingale process with respect to $\mathcal{F}(t)$. Similar to the approach of Chen et al. (2002), i.e. by mimicking generalized estimating equation, we consider the following two estimating equations:

$$U(\beta, h) = \sum_{i=1}^{n} \int_{0}^{\tau_c} Z_i[dN_i(t) - Y_i(t)d\Lambda_\epsilon(\beta^T Z_i + h(t))] = 0, \quad (2.2)$$

and

$$\sum_{i=1}^{n}[dN_i(t) - Y_i(t)d\Lambda_\epsilon(\beta^T Z_i + h(t))] = 0, \quad (2.3)$$

where h is a nondecreasing function satisfying $h(0) = -\infty$ and $\tau_c < b_F$ is a pre-specified constant . This requirement ensures that $\Lambda_\epsilon(a + h(0)) = 0$ for any finite a. Let $\tilde{\beta}$ and $\tilde{h}(t; \tilde{\beta})$ denote the solution of (2.2) and (2.3). Note that $\tilde{h}(t; \tilde{\beta})$ is a step function in t that rises at the distinct jump points of $\{I_{[X_i \le t, \delta_i = 1]}; i = 1, \dots, n\}$.

Consider the special case of the Cox model, in which $\lambda_\epsilon = \exp(t)$. It then follows from (2.2) and (2.3) that the estimator $\tilde{\beta}$ satisfies the following equation (see Appendix):

$$\sum_{i=1}^{n} \int_{0}^{\tau_c} \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j Y_j(t) \exp(\beta^T Z_j)}{\sum_{j=1}^{n} Y_j(t) \exp(\beta^T Z_j)} \right\} dN_i(t) = 0,$$

which is precisely the Cox partial likelihood score equation for left-truncated and right-censored data (Pan and Chappell, 2002). Equations (2.2) and (2.3) suggest the following iterative algorithms for computing $\tilde{\beta}$ and $\tilde{h}(t; \tilde{\beta})$:

Step 0: Choose an initial value of β, denoted by $\tilde{\beta}^{(0)}$.

Step 1: Let $t_1 < t_2 < \cdots < t_{n_d} < \tau_c$ denote the distinct uncensored points. Obtain $\tilde{h}^{(0)}(t_1; \tilde{\beta}^{(0)})$ by solving

$$\sum_{i=1}^{n} Y_i(t_1)\Lambda_\epsilon(\beta^T Z_i + h(t_1)) = 1,$$

with $\beta = \tilde{\beta}^{(0)}$. Then, obtain $\tilde{h}(t_k)$ for $k = 2, \dots, n_d$, one-by-one by solving the equation

$$\sum_{i=1}^{n} Y_i(t_k)\Lambda_\epsilon(\beta^T Z_i + h(t_k)) = 1 + \sum_{i=1}^{n} Y_i(t_k)\Lambda_\epsilon(\beta^t Z_i + h(t_k-)),$$

with $\beta = \tilde{\beta}^{(0)}$.

Step 2: Obtain a new estimate of β by solving (2.2) with $h(t_k) = \tilde{h}^{(0)}(t_k; \tilde{\beta}^{(0)})$.

Step 3: Set $\tilde{\beta}^{(0)}$ to be the estimate obtained in Step 2 and repeat Steps 1 and 2 until prescribed convergence criteria are met.

2.2 The asymptotic Properties of the Proposed Estimator

For any vector x, let $x^{\otimes 2} = xx^T$. Similar to Proposition of Chen et al. (2002), under suitable regularity conditions, we have the following proposition.

Theorem 1. Under assumption (2.1) and regularity conditions (Fleming and Harrington, 1991), we have that $n^{\frac{1}{2}}(\tilde{\beta} - \beta) \to N(0, \Sigma_{\tilde{\beta}})$ in distribution, as $n \to \infty$, where $\Sigma_{\tilde{\beta}} = \Sigma_2^{-1}\Sigma_1(\Sigma_2^{-1})^T$

$$\Sigma_1 = E\left[\int_0^{\tau_c} [Z_1 - \mu_z(t;\beta_0)]^{\otimes 2} \lambda_\epsilon(h_0(t) + \beta_0^T) Y_1(t)] dh_0(t)\right],$$

$$\Sigma_2 = E\left[\int_0^{\tau_c} [Z_1 - \mu_z(t;\beta_0)] Z_1^T \dot{\lambda}_\epsilon(h_0(t) + \beta_0^T) Y_1(t)] dh_0(t)\right],$$

where

$$\mu_z(t) = \frac{E[Z_1 \lambda_\epsilon(h_0(X_1) + \beta_0^T Z_1) Y_1(t) B(t; X_1)]}{E[\lambda_\epsilon(h_0(t) + \beta_0^T Z_1) Y_1(t)]},$$

where

$$B(t,s) = \exp\left(\int_s^t \frac{E[\dot{\lambda}_\epsilon(h_0(x) + \beta_0^T Z_1) Y_1(x)]}{E[\lambda_\epsilon(h_0(x) + \beta_0^T Z_1) Y_1(x)} dh_0(x)\right).$$

Proof:

Let $\mathcal{H}$ be the collection of all nondecreasing step functions on $[0, \tau_c]$ with $h(0) = -\infty$ and with jumps only at the observed failure times. For any two nondecreasing functions h_1 and h_2 on $[0, \tau_c]$ such that $h_1(0) = h_2(0) = -\infty$, define

$$d(h_1, h_2) = \sup(|\exp\{h_1(t)\} - \exp\{h_2(t)\}| : t \in [0, \tau_c]).$$

Let M be a mapping on $\mathcal{H}$ defined by

$$M(h)(t) = n^{-1} \sum_{i=1}^n \int_0^t [dN_i(s) - Y_i(s) d\Lambda_\epsilon(\beta^T Z_i + h(s))].$$

For an arbitrary but fixed ϵ_*, consider $h_1 \in \mathcal{H}$ and $h_2 \in \mathcal{H}$ such that $d(h_1, h_2) \geq \epsilon_*$. There exists a $t_* \in (0, \tau_c]$ such that $|\exp\{h_1(t_*)\} - \exp\{h_2(t_*)\}| \geq \epsilon_*/2$. Then

$$\sup_{t\in[0,\tau_c]} |M(h_1)(t) - M(h_2)(t)| = n^{-1} \sup_{t\in[0,\tau_c]} \left|\sum_{i=1}^n \Lambda_\epsilon(\beta_0^T Z_i + h_1(t \wedge X_i)) - \Lambda_\epsilon(\beta_0^T Z_i + h_2(t \wedge X_i))\right|$$

$$\geq n^{-1}\left|\sum_{i=1}^{n}\int_{h_2(t_*\wedge X_i)}^{h_1(t_*\wedge X_i)}\lambda_\epsilon(s+\beta_0^TZ_i)ds\right|\geq n^{-1}\left|\sum_{i=1}^{n}I_{[X_i\geq\tau_c]}\int_{h_2(t_*)}^{h_1(t_*)}\lambda_\epsilon(s+\beta_0^TZ_i)ds\right|$$

$$\geq n^{-1}\left|\sum_{i=1}^{n}I_{[X_i\geq\tau_c]}\inf\left\{\int_{\log a}^{\log b}\lambda_\epsilon(s+c)ds : 0<a<b<m, b-a<\epsilon_*/2, c<m\right\}\right.,$$

where m is a large but fixed constant. Similar to the arguments of Step A1 of Chen et al. (2002), it follows that $d(\tilde{h}(\cdot;\beta_0)-h(\cdot,\beta_0))\to 0$ almost surely, where $\tilde{h}(\cdot;\beta_0)\in\mathcal{H}$ is the function implicitly defined as the unique solution of (2.2) given β. By Steps A2 through A6 of Chen et al. (2002), the proof is completed.

Note that Σ_1 and Σ_2 can be consistently estimated by

$$\hat{\Sigma}_1=n^{-1}\sum_{i=1}^{n}\int_0^{\tau_c}[Z_i-\bar{Z}(t;\tilde{\beta})]^{\otimes 2}\lambda_\epsilon(\tilde{\beta}^TZ_i+\tilde{h}(t;\tilde{\beta}))Y_i(t)d\tilde{h}(t;\tilde{\beta}),$$

and

$$\hat{\Sigma}_2=n^{-1}\sum_{i=1}^{n}\int_0^{\tau_c}[Z_i-\bar{Z}(t;\tilde{\beta})]Z_i^T\dot{\lambda}_\epsilon(\tilde{\beta}^TZ_i+\tilde{h}(t;\tilde{\beta}))Y_i(t)d\tilde{h}(t;\tilde{\beta}),$$

respectively, where $\dot{\lambda}_\epsilon(x)=d\lambda_\epsilon(x)/dx$,

$$\bar{Z}(t;\tilde{\beta})=\sum_{i=1}^{n}\frac{Z_i\lambda_\epsilon(\tilde{\beta}^TZ_i+\tilde{h}(t;\tilde{\beta}))Y_i(t)\hat{B}(t,X_i)}{\sum_{i=1}^{n}\lambda_\epsilon(\tilde{\beta}^TZ_i+\tilde{h}(t;\tilde{\beta}))Y_i(t)},$$

$$\hat{B}(t,s)=\exp\left(\int_s^t\frac{\sum_{i=1}^{n}\dot{\lambda}_\epsilon(\tilde{\beta}^TZ_i+\tilde{h}(x;\tilde{\beta}))Y_i(x)}{\sum_{i=1}^{n}\lambda_\epsilon(\tilde{\beta}^TZ_i+\tilde{h}(x;\tilde{\beta}))Y_i(x)}d\tilde{h}(x;\tilde{\beta})\right).$$

Hence, a consistent estimator of $\Sigma_{\tilde{\beta}}$ is given by $\tilde{\Sigma}_{\tilde{\beta}}=\hat{\Sigma}_2^{-1}\hat{\Sigma}_1(\hat{\Sigma}_2^{-1})^T$.

Chapter 3. Simulation Studies

We generated T following the proportional odds model with $h(t)=\log(t/10)$ and $\beta=(\beta_1=1,\beta_2=1)^T$. The resulting T has the survivorship function

$$P(T>t|Z_1,Z_2)=\frac{1}{1+\exp\{\log(t/10)+Z_1+Z_2\}},$$

where Z_1 is an ordinal variable with $P(Z_1=i)=0.25$ for $i=1,2,3,4$ and Z_2 is a Bernouli random variable with probability 0.5. Note that under this set-up, the p^{th} percentile of T at (Z_1,Z_2) is $t_p=10\exp\{\log((1-p)/p)-(Z_1+Z_2)\}$, which decreases as Z_1 or Z_2 increases. We generated the left-truncation variable V from uniform $U(0,\theta)$, and right censoring variable C was generated from $D+V$, where D is exponentially distributed with mean θ_d. The values of θ_d are set at 0.1 and 0.5, and the values of θ are set at 0.25, 2 and 5. Sample size is set at $n = 100$ and 300. The replication time is 1000. The values of τ_a and τ_b are set at the smallest and largest values of X_i's, respectively. For each simulated dataset, we obtained $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$. Tables 1 and 2 show the simulated biases, standard deviations (std), and root mean squared error (rmse) of $\hat{\beta}_i$ $(i = 1, 2)$ for estimating β_1 and β_2, respectively. Tables 1 and 2 also shows the proportion of left-truncation $\sum_{k=1}^{K} \hat{f}_k P(T < V|z_k)$ (denoted by q) and right-censoring (denoted by $p_c = P(\delta_i = 0)$), where $\hat{f}_k$ is the simulated proportion of the observations of the subgroup k.

Table 1. Simulated biases, std. and rmse of $\hat{\beta}_1$

θ_d	θ	p_c	q	n	bias	std	rmse
0.1	0.25	0.21	0.24	100	-0.034	0.361	0.363
0.1	0.25	0.21	0.24	300	-0.020	0.217	0.218
0.1	2	0.31	0.56	100	-0.040	0.381	0.383
0.1	2	0.31	0.56	300	-0.024	0.252	0.253
0.1	10	0.41	0.81	100	-0.046	0.413	0.416
0.1	10	0.41	0.81	300	-0.031	0.272	0.274
0.5	0.25	0.45	0.24	100	-0.038	0.396	0.398
0.5	0.25	0.45	0.24	300	-0.026	0.257	0.258
0.5	2	0.59	0.56	100	-0.043	0.456	0.458
0.5	2	0.59	0.56	300	-0.026	0.324	0.325
0.5	10	0.65	0.81	100	-0.060	0.491	0.495
0.5	10	0.65	0.81	300	-0.032	0.384	0.385

Table 2. Simulated biases, std. and rmse of $\hat{\beta}_2$

θ_d	θ	p_c	q	n	bias	std	rmse
0.1	0.25	0.21	0.24	100	-0.023	0.372	0.373
0.1	0.25	0.21	0.24	300	-0.010	0.239	0.239
0.1	2	0.31	0.56	100	-0.041	0.388	0.390
0.1	2	0.31	0.56	300	-0.018	0.262	0.263
0.1	10	0.41	0.81	100	-0.069	0.434	0.439
0.1	10	0.41	0.81	300	-0.047	0.295	0.299
0.5	0.25	0.45	0.24	100	-0.028	0.397	0.398
0.5	0.25	0.45	0.24	300	-0.017	0.225	0.226
0.5	2	0.59	0.56	100	-0.034	0.439	0.440
0.5	2	0.59	0.56	300	-0.015	0.260	0.260
0.5	10	0.65	0.81	100	-0.053	0.467	0.470
0.5	10	0.65	0.81	300	-0.042	0.329	0.332

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

The standard deviations of both estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ increase as the proportion of left-truncation q and right censoring (p_c) increase.

Chapter 4. Conclusions

In this note, we have demonstrated that the approach of Chen et al. (2002) can be easily extended to LTRC data. For right-censored data, Zeng and Lin (2007) considered maximum likelihood estimation in transformation model with random effects for dependent failure times. It is not easy to extend their approach to LTRC data due to complexity of the likelihood function. However, our approach can be extended to the following random-effects linear transformation model:

$$h(T_{li}) = -\beta^T Z_{li} - b_l^T \tilde{Z}_{li} + \epsilon_{li}, \quad (l = 1, \dots, c; i = 1, \dots, n_l)$$

where T_{li} is the failure time for the i^{th} individual in the l^{th} cluster, $\tilde{Z}_{li}$ is a set of covariates and b_l ($l = 1, \dots, c$) are independent zero-mean random vectors with multivariate density function $f(b_l; \gamma)$ indexed by a set of parameters γ. We may obtain a consistent estimator of β by modifying equations (2.2) and (2.3) as

$$U(\beta, h) = \sum_{l=1}^{c}\sum_{i=1}^{n_l}\int_b\int_0^{\tau_c} Z_{li}[dN_{li}(t) - Y_{li}(t)d\Lambda_\epsilon(\beta^T Z_{li} + b^T\tilde{Z}_{li} + h(t))]f(b;\gamma)db = 0,$$

and

$$\sum_{l=1}^{c}\sum_{i=1}^{n_l}\int_b [dN_{li}(t) - Y_{li}(t)d\Lambda_\epsilon(\beta^T Z_{li} + b^T\tilde{Z}_{li} + h(t))]f(b;\gamma)db = 0,$$

where $N_{li}(\cdot)$ and $Y_{li}(\cdot)$ are defined analogously to $N_i(\cdot)$ and $Y_i(\cdot)$ of Section 2.1. Given γ, let $\tilde{\beta}_\gamma$ and $\tilde{h}_\gamma$ be the solutions of the above two equations. Then, optimal estimators of $\tilde{\beta}$ and $\tilde{h}$ can be obtained by minimizing the following statistics with respect to γ:

$$\sum_{l=1}^{c}\sum_{i=1}^{n_l}\int_b [dN_{li}(t) - Y_{li}(t)d\Lambda_\epsilon(\tilde{\beta}_\gamma^T Z_{li} + b^T\tilde{Z}_{li} + \tilde{h}_\gamma(t))]^2 f(b;\gamma)db.$$

Appendix:

Proof that (2.2) and (2.3) are reduced to partial likelihood when $g(\cdot) = \exp\{-\exp(\cdot)\}$

proof:

$$d\Lambda_\epsilon(\beta^T Z_i + h(t)) = \exp(\beta^T Z_i)d(\exp(h(t)))$$

$$\Rightarrow \sum_{i=1}^{n}[dN_i(t) - Y_i(t)\exp(\beta^T Z_i)d(\exp(h(t)))] = 0$$

$$\Rightarrow d(\exp(h(t))) = \frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{i=1}^{n} Y_i(t)\exp(\beta^T Z_i)}$$

$$\sum_{i=1}^{n} \int_0^{\tau_c} Z_i[dN_i(t) - Y_i(t)d\Lambda_\epsilon(\beta^T Z_i + h(t))] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \int_0^{\tau_c} Z_i[dN_i(t) - Y_i(t)\ \exp(\beta^T Z_i)d(\exp(h(t)))] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \int_0^{\tau_c} Z_i[dN_i(t) - Y_i(t)\frac{\sum_{i=1}^{n} dN_i(t)\exp(\beta^T Z_i)}{\sum_{i=1}^{n} Y_i(t)\exp(\beta^T Z_i)}] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \int_0^{\tau_c} Z_i \left\{1 - \frac{\sum_{i=1}^{n} Y_i(t)\exp(\beta^T Z_i)}{\sum_{i=1}^{n} Y_i(t)\exp(\beta^T Z_i)}\right\} dN_i(t) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \int_0^{\tau_c} \left\{Z_i - \frac{\sum_{j=1}^{n} Z_j Y_j(t)\exp(\beta^T Z_j)}{\sum_{j=1}^{n} Y_j(t)\exp(\beta^T Z_j)}\right\} dN_i(t) = 0$$

References

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

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

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.

The CSHA working group (1994). Canadian study of health and aging: study methods and prevalence of dementia. *J. Can. Med. Asso.*, **150**, 899-913.

Hyde, J., (1980). Survival analysis with incomplete observations. In Biostatistics Casebook, R. G. Miller, B. Efron, B. W. Brown, and L. E. Moses, eds, New York: John Wiley and Sons, pp. 31-46.

Fleming, T. R. and Harrington, D. P. (1991). Counting Processes and Survival Analysis. New York: Wiley.

Klein, J. P. and Moeschberger, M. L. (1997). Survival analysis: Tenchniques for censored andtruncated data. Springer.

Lai, T. L. and Ying, Z. (1991), Estimating a distribution function with truncated and censored data. The Annal of Statistics, 19, 417-422.

Murphy, S. A., Rossini, A. J. and van der Vaart, A. W. (1996). Maximum likelihood estimation in the proportional odds model. *J. Am. Statist. Assoc.*, **92**, 968-976.

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

Tsai, W.-Y., Jewell, N. P., Wang, M.-C. (1987). A note on the product-limit estimate under right censoring and left truncation. *Biometrika*, **74**, 883-886.

Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Amer. Statist. Ass.*, **86**, 130-143.

Ying, S and Prentice, R. (1999). Semiparametric inference in the proportional odds regression model. *J. Am. Statist. Assoc.*, **92**, 968-976.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression model with censored data. *J. R. Statist. Soc. B*, **69**, 507-564.