Efficiency of Profile/Partial Likelihood in the Cox Model

Yuichi Hirose

School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, New Zealand

Summary. This paper shows equivalence and efficiency of the partial likelihood and profile likelihood estimator in the Cox regression model using the direct expansion of profile likelihood. This approach gives alternative proof to the one illustrated in Murphy and van der Vaart (2000).

Keywords: Cox model; Semi-parametric model; Profile likelihood; Partial likelihood; Efficiency; M-estimator; Maximum likelihood estimator; Efficient score; Efficient information bound.

1. Introduction

Cox (1972) introduced an proportional hazard model, known as the Cox model, where the cumulative hazard function of the survival time T for a subject with covariate $Z \in \mathbb{R}^k$ is given by

$$\Lambda(t|Z) = e^{\beta^T Z}\Lambda(t) \tag{1}$$

where $\Lambda(t)$ is an unspecified baseline cumulative hazard function. In the same paper, Cox also proposed an estimation of β using partial likelihood. Since then, several authors, Cox (1975), Tsiatis (1981), Andersen and Gill (1982), Bailey (1983, 1984), Johansen (1983) and Jacobsen (1984) have tried to justify the method of partial likelihood estimation, and establish the asymptotic equivalence of the partial likelihood estimator and the maximum likelihood estimator. We show that the profile likelihood is the most natural way to justify the partial likelihood in the Cox model and establish the asymptotic properties of its estimator. Murphy and van der Vaart (2000) discussed asymptotic normality of the profile likelihood estimator by applying an approximate least favorable submodel which was proposed in their paper. Our approach uses the direct asymptotic expansion of profile likelihood for the Cox regression model and show the estimator is efficient.

Suppose we observe (X, δ, Z) in time interval $[0, \tau]$, where $X = T \wedge C$, $\delta = 1_{\{T \le C\}}$, $Z \in \mathbb{R}^k$ is a regression covariate, T is a right-censored failure time with cumulative hazard is given by Equation (1), and C is a censoring time independent of T given Z and uninformative of (β, Λ). Let $N(t) = 1_{\{X \le t, \delta = 1\}}$, $Y(t) = 1_{\{X \ge t\}}$ and $M(t) = N(t) - \int_0^t Y(s) e^{\beta^T Z} d\Lambda(s)$. The log-likelihood for a single observation (X, δ, Z) is

$$\ell(X, \delta, Z; \beta, \Lambda) = (\beta^T Z + \log \Delta\Lambda(X))\delta - e^{\beta^T Z}\Lambda(X), \tag{2}$$

where $\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-)$.

For a cdf F, write $E_F f = \int f dF$ and define

$$\hat{\Lambda}(t; \beta, F) = \int_0^t \frac{E_F dN(s)}{E_F[Y(s)e^{\beta^T Z}]}. \tag{3}$$

The Breslow estimator is given by $\hat{\Lambda}(t;\beta,F_n)=\int_0^t \frac{E_{F_n}dN(s)}{E_{F_n}[Y(s)e^{\beta^T Z}]}$ where F_n is the empirical cdf. If F_0 is the cdf at the true value (β_0,Λ_0), then $\hat{\Lambda}(t;\beta_0,F_0)=\Lambda_0(t)$. We substitute the function $\hat{\Lambda}(\beta,F)=\hat{\Lambda}(t;\beta,F)$ in the log-likelihood (Equation (2)) and call it the induced model. The log-likelihood for an observation (X,δ,Z) in the induced model is

$$\ell(X,\delta,Z;\beta,\hat{\Lambda}(\beta,F)) = \left(\beta^T Z+\log\frac{E_F\Delta N(X)}{E_F[Y(X)e^{\beta^T Z}]}\right)\delta - e^{\beta^T Z}\int_0^X \frac{E_F dN(s)}{E_F[Y(s)e^{\beta^T Z}]} \quad (4)$$

The score function and its derivative at (X,δ,Z) in the induced model are

$$\begin{aligned}\dot{\ell}(X,\delta,Z;\beta,F) &= \frac{\partial}{\partial\beta}\ell(X,\delta,Z;\beta,\hat{\Lambda}(\beta,F)) \\ &= \int_0^\tau\left\{Z-\frac{E_F[ZY(t)e^{\beta^T Z}]}{E_F[Y(t)e^{\beta^T Z}]}\right\}\left\{dN(t)-Y(t)e^{\beta^T Z}d\hat{\Lambda}(t;\beta,F)\right\} \quad (5)\end{aligned}$$

and

$$\begin{aligned}&\ddot{\ell}(X,\delta,Z;\beta,F)=\frac{\partial}{\partial\beta}\dot{\ell}(X,\delta,Z;\beta,F) \\ &= -\int_0^\tau\left\{\frac{E_F[Z^{\otimes 2}Y(t)e^{\beta^T Z}]}{E_F[Y(t)e^{\beta^T Z}]}-\frac{(E_F[ZY(t)e^{\beta^T Z}])^{\otimes 2}}{(E_F[Y(t)e^{\beta^T Z}])^2}\right\}\left\{dN(t)-Y(t)e^{\beta^T Z}d\hat{\Lambda}(t;\beta,F)\right\} \\ &\quad -\int_0^\tau\left\{Z-\frac{E_F[ZY(t)e^{\beta^T Z}]}{E_F[Y(t)e^{\beta^T Z}]}\right\}^{\otimes 2}Y(t)e^{\beta^T Z}d\hat{\Lambda}(t;\beta,F). \quad (6)\end{aligned}$$

Since $\hat{\Lambda}(t;\beta_0,F_0)=\Lambda_0(t)$, the induced score function at (β_0,F_0),

$$\dot{\ell}(X,\delta,Z;\beta_0,F_0)=\int_0^\tau\left\{Z-\frac{E_0[ZY(t)e^{\beta_0^T Z}]}{E_0[Y(t)e^{\beta_0^T Z}]}\right\}dM(t)=:\dot{\ell}^*(X,\delta,Z) \quad (7)$$

is the efficient score function $\dot{\ell}^*(X,\delta,Z)$ and the efficient information matrix is given by

$$\begin{aligned}I_0^* &:= -E_0\ddot{\ell}(X,\delta,Z;\beta_0,F_0) \\ &= E_0\int_0^\tau\left\{Z-\frac{E_0[ZY(t)e^{\beta_0^T Z}]}{E_0[Y(t)e^{\beta_0^T Z}]}\right\}^{\otimes 2}Y(t)e^{\beta_0^T Z}d\Lambda_0(t) \quad (8)\end{aligned}$$

where $E_0=E_{F_0}$ is the expectation at the true value (cf. Murphy and van der Vaart (2000)).

We assume

(C1) $P(X\geq\tau)=E_0(Y(\tau))>0$, and

(C2) the range of Z is bounded;

(C3) the efficient information matrix I_0^* is invertible;

(C4) the empirical cdf F_n is $\sqrt{n}$-consistent, i.e., $\sqrt{n}(F_n-F_0)=O_P(1)$.

2. Efficiency in Profile and Partial likelihood

The partial likelihood and and the corresponding score equation are given by

$$L_n(\beta) = \prod_{i=1}^{n} \prod_{0\le t\le \tau} \left\{ \frac{Y_i(t)e^{\beta^T Z_i}}{\sum_{j=1}^n Y_j(t)e^{\beta^T Z_j}} \right\}^{\Delta N_i(t)}$$

and

$$\frac{\partial}{\partial\beta} \log L_n(\beta) \quad = \quad \mathbb{P}_n \int_0^\tau \left\{ Z - \frac{E_{F_n}[ZY(t)e^{\beta^T Z}]}{E_{F_n}[Y(t)e^{\beta^T Z}]} \right\} dN(t) = 0. \qquad (9)$$

On the other hand, for the empirical cdf F_n, $\mathbb{P}_n \ell(X,\delta,Z;\beta,\hat{\Lambda}(\beta,F_n))$ gives a version of profile (log-) likelihood, where $\ell(X,\delta,Z;\beta,\Lambda)$ is the log-likelihood for an observation given by Equation (2) and $\hat{\Lambda}(\beta,F_n)$ is the Breslow estimator given by Equation (3) . By Equation (5), the score equation for the profile likelihood is

$$\mathbb{P}_n \int_0^\tau \left\{ Z - \frac{E_{F_n}[ZY(t)e^{\beta^T Z}]}{E_{F_n}[Y(t)e^{\beta^T Z}]} \right\} \left\{ dN(t) - Y(t)e^{\beta^T Z} d\hat{\Lambda}(t;\beta,F_n) \right\} \quad = \quad 0. \qquad (10)$$

Since

$$\mathbb{P}_n \int_0^\tau \left\{ Z - \frac{E_{F_n}[ZY(t)e^{\beta^T Z}]}{E_{F_n}[Y(t)e^{\beta^T Z}]} \right\} Y(t)e^{\beta^T Z} d\hat{\Lambda}(t;\beta,F_n) = 0,$$

the score equations Equation (9) and Equation (10) are the same equation. This establishes the equivalence of the estimators based on the profile likelihood and the partial likelihood.

The following theorem shows that the estimator based on the profile likelihood and the partial likelihood are efficient.

THEOREM 1. *Suppose (C1)–(C4). The solution $\hat{\beta}_n$ to the score equation for the profile likelihood (Equation (10)) and the solution $\hat{\beta}_n$ to the score equation for the partial likelihood (Equation (9)) are both asymptotically linear estimators with the efficient influence function $(I_0^*)^{-1}\dot{\ell}^*$ so that*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \sqrt{n}\mathbb{P}_n (I_0^*)^{-1}\dot{\ell}^*(X,\delta,Z) + o_P(1) \xrightarrow{d} N(0,(I_0^*)^{-1}). \qquad (11)$$

where the efficient score $\dot{\ell}^*$ and the efficient information I_0^* are given by Equations (7) and (8), respectively.

Proof. Conditions (R0)–(R3) in Theorem A in Appendix A are verified in Appendix B. The claim follows from Theorem A. □

3. Discussion

The equivalence of the estimator in the partial likelihood and profile likelihood in the Cox regression model has been established by Bailey (1983, 1984), Johansen (1983), and Jacobsen (1984). Asymptotic behavior of the estimator has been studied by Tsiatis (1981) and Andersen and Gill (1982). Murphy and van der Vaart (2000)) discussed the profile likelihood estimator in the Cox regression model to illustrate the method of an approximate least favorable submodel that was used to establish the efficiency of the profile likelihood estimator for general semiparametric models. Our approach uses the direct expansion of profile likelihood (cf. Hirose (2009)) to show the efficiency of the profile likelihood estimator in the Cox regression model.

Appendix A: Theorem A

This section is a modification of the result in Hirose (2009).

Hadamard differentiability
We say that a map $\psi : B_1 \to B_2$ between two Banach spaces B_1 and B_2 is Hadamard differentiable at x if there is a continuous linear map $d\psi(x) : B_1 \to B_2$ such that

$$\frac{\psi(x+th') - \psi(x)}{t} \to d\psi(x)(h) \quad \text{as} \quad t \to 0 \quad \text{and} \quad h' \to h.$$

The map $d\psi(x)$ is called derivative of ψ at x, and is continuous in x. (For reference, see Gill (1989) and Shapiro (1990).)

Theorem and its assumptions
On the set of cdf functions $\mathcal{F}$, we use the sup-norm, i.e., for $F, F_0 \in \mathcal{F}$,

$$\|F - F_0\| = \sup_x |F(x) - F_0(x)|.$$

For $\rho > 0$, let

$$\mathcal{C}_\rho = \{F \in \mathcal{F} : \|F - F_0\| < \rho\}.$$

Suppose we consider a semi-parametric model of the form

$$\mathcal{P} = \{p(x;\beta,\eta) : \ \beta \in \Theta_\beta \subset \mathbb{R}^m, \ \eta \in \Theta_\eta\}$$

where β is the m-dimensional parameter of interest, and η is a nuisance parameter, which may be infinite-dimensional. Let (β_0, η_0) be the true value of (β, η). We assume Θ_β is a compact set containing an open neighborhood of β_0 in $\mathbb{R}^m$, and Θ_η is a convex set containing η_0 in a Banach space $\mathcal{B}$. The expectation at the true value (β_0, η_0) is denote by E_0.

For a map $\hat\eta : \Theta_\beta \times \mathcal{F} \to \Theta_\eta$, define a model (called the *induced model*) with log-likelihood for one observation

$$\ell(x;\beta,F) = \log p(x;\beta,\hat\eta(\beta,F)), \ \beta \in \Theta_\beta, \ F \in \mathcal{F}.$$

The score function in the induced model is denoted by

$$\dot\ell(x,\beta,F) = \frac{\partial}{\partial\beta}\ell(x;\beta,F). \tag{12}$$

We assume that:

(R0) $\hat{\eta}$ satisfies $\hat{\eta}(\beta_0, F_0) = \eta_0$ and the function

$$\dot{\ell}^*(x) = \dot{\ell}(x, \beta_0, F_0)$$

is the efficient score function.

(R1) The empirical process F_n is $\sqrt{n}$-consistent, i.e., $\sqrt{n}\|F_n - F_0\| = O_P(1)$, and for each $(\beta, F) \in \Theta_\beta \times \mathcal{F}$, the log-likelihood function $\ell(x; \beta, F)$ is twice continuously differentiable with respect to β and Hadamard differentiable with respect to F for all x.
(Derivatives are denoted by $\dot{\ell}(x, \beta, F) = \frac{\partial}{\partial\beta}\ell(x; \beta, F)$, $\ddot{\ell} = \frac{\partial}{\partial\beta}\dot{\ell}(x, \beta, F)$, $A(x, \beta, F) = d_F \ell(x; \beta, F)$ and $d_F \dot{\ell}(x, \beta, F)$.)

(R2) The efficient information matrix $I_0^* = E_0(\dot{\ell}^* \dot{\ell}^{*T})$ is invertible.

(R3) There exist a $\rho > 0$ and a neighborhood Θ_β of β_0 such that the class of functions $\{\dot{\ell}(x, \beta, F) : (\beta, F) \in \Theta_\beta \times \mathcal{C}_\rho\}$ is Donsker with square integrable envelope function, and such that the class of functions $\{\ddot{\ell}(x, \beta, F) : (\beta, F) \in \Theta_\beta \times \mathcal{C}_\rho\}$ is Glivenko-Cantelli with integrable envelope function.

Theorem A. *Suppose sets of assumptions* $\{(R0), (R1), (R2), (R3)\}$*, then a consistent solution* $\hat{\beta}_n$ *to the estimating equation*

$$\mathbb{P}_n \dot{\ell}(X, \hat{\beta}_n, F_n) = 0 \tag{13}$$

is an asymptotically linear estimator for β_0 *with the efficient influence function* $(I_0^*)^{-1}\dot{\ell}^*(x)$ *so that*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \sqrt{n}\mathbb{P}_n (I_0^*)^{-1}\dot{\ell}^*(X) + o_P(1) \xrightarrow{d} N\left(0, (I_0^*)^{-1}\right).$$

This demonstrates that the estimator $\hat{\beta}_n$ *is efficient.*

Proof.

Since (i) the range of the score operator $A(X, \beta_0, F_0) = d_F \ell(x; \beta_0, F_0) = d_F \log p(x; \beta_0, \hat{\eta}(\beta_0, F_0))$ for F is in the nuisance tangent space (the tangent space for η), and (ii) the function $\dot{\ell}(x, \beta_0, F_0)$ is the efficient score function, we have

$$E_0 d_F \dot{\ell}(X, \beta_0, F_0) = -E_0[\dot{\ell}(X, \beta_0, F_0) A(X, \beta_0, F_0)] = 0 \quad \text{(the zero operator)}. \tag{14}$$

For F_n and F_0 in $\mathcal{F}$, consider a path $F_n^*(t) = F_0 + t(F_n - F_0)$, $t \in [0, 1]$. Then $F_n^*(0) = F_0$ and $F_n^*(1) = F_n$. Under the assumption $\sqrt{n}\|F_n - F_0\| = O_P(1)$ (condition (R1)), we have that $\sup_{t\in[0,1]} |F_n^*(t) - F_0| = o_P(1)$.

By the mean value theorem for vector valued function (cf. Hall and Newell (1979)),

$$\begin{aligned}
& \|\sqrt{n} E_0 \dot{\ell}(X, \beta_0, F_n)\| \\
= \;& \|\sqrt{n} E_0 \dot{\ell}(X, \beta_0, F_n^*(1)) - \sqrt{n} E_0 \dot{\ell}(X, \beta_0, F_n^*(0))\| \\
\le \;& \sup_{t\in[0,1]} \|E_0 d_F \dot{\ell}(X, \beta_0, F_n^*(t))\| \sqrt{n}\|F_n - F_0\| \\
= \;& \|E_0 d_F \dot{\ell}(X, \beta_0, F_0) + o_p(1)\| \sqrt{n}\|F_n - F_0\| \quad (\text{since } \sup_{t\in[0,1]} |F_n^*(t) - F_0| = o_P(1)) \\
= \;& o_p(1)\sqrt{n}\|F_n - F_0\| \quad (\text{by Equation (14)}) \\
= \;& o_P(1) \quad (\text{since } \sqrt{n}\|F_n - F_0\| = O_P(1)).
\end{aligned} \tag{15}$$

Since the functions $\dot\ell(x,\beta,F)$ and $\ddot\ell(x,\beta,F)$ are continuous at (β_0,F_0), and they are dominated by the square integrable function and the integrable function, respectively, by dominated convergence theorem, for every $(\beta_n^*,F_n^*)\xrightarrow{P}(\beta_0,F_0)$, we have

$$E_0\|\dot\ell(X,\beta_0,F_n^*)-\dot\ell(X,\beta_0,F_0)\|^2\xrightarrow{P}0.$$

and

$$E_0\|\ddot\ell(X,\beta_n^*,F_n^*)-\ddot\ell(X,\beta_0,F_0)\|\xrightarrow{P}0.$$

Together with condition (R3), this implies that

$$\sqrt{n}\mathbb{P}_n\left\{\dot\ell(X,\beta_0,F_n)-\dot\ell(X_i,\beta_0,F_0)\right\}=\sqrt{n}E_0\left\{\dot\ell(X,\beta_0,F_n)-\dot\ell(X,\beta_0,F_0)\right\}+o_P(1)\quad(16)$$

by Lemma 13.3 in Kosorok (2008), and for every $(\beta_n^*,F_n^*)\xrightarrow{P}(\beta_0,F_0)$,

$$\mathbb{P}_n\ddot\ell(X,\beta_n^*,F_n^*)\xrightarrow{P}E_0\ddot\ell(X,\beta_0,F_0)=-I_0^*.\quad(17)$$

By combining Equation (15) and (16), we get

$$\sqrt{n}\mathbb{P}_n\dot\ell(X,\beta_0,F_n)=\sqrt{n}\mathbb{P}_n\dot\ell(X,\beta_0,F_0)+o_P(1).\quad(18)$$

Finally, by Taylor's expansion, for some β_n^* with $\|\beta_n^*-\beta_0\|\le\|\hat\beta_n-\beta_0\|\xrightarrow{P}0$,

$$\begin{aligned}0&=\sqrt{n}\mathbb{P}_n\dot\ell(X,\hat\beta_n,F_n)\\&=\sqrt{n}\mathbb{P}_n\dot\ell(X,\beta_0,F_n)+\mathbb{P}_n\ddot\ell(X,\beta_n^*,F_n)\sqrt{n}(\hat\beta_n-\beta_0)\\&=\sqrt{n}\mathbb{P}_n\dot\ell(X,\beta_0,F_0)+o_P(1)+\{-I_0^*+o_P(1)\}\sqrt{n}(\hat\beta_n-\beta_0)\end{aligned}$$

where the last equality is by Equations (17) and (18). Hence, by condition (R2),

$$\sqrt{n}(\hat\beta_n-\beta_0)=(I_0^*)^{-1}\sqrt{n}\mathbb{P}_n\dot\ell(X,\beta_0,F_0)+o_P(1)\{1+\sqrt{n}(\hat\beta_n-\beta_0)\}.$$

Since $(I_0^*)^{-1}\sqrt{n}\mathbb{P}_n\dot\ell(X,\beta_0,F_0)=O_P(1)$, this equality implies $\sqrt{n}(\hat\beta_n-\beta_0)=O_P(1)$ and

$$\sqrt{n}(\hat\beta_n-\beta_0)=(I_0^*)^{-1}\sqrt{n}\mathbb{P}_n\dot\ell(X,\beta_0,F_0)+o_P(1).$$

□

Appendix B: Proof of Theorem 1

To prove Theorem 1, Conditions (R0)–(R3) in Theorem A (in Appendix A) are verified here.

Condition (R0): This condition is verified by two lines below Equation (3) and Equation (7).

Condition (R1): Equation (4) is twice continuously differentiable with respect to β with the first and second derivatives (5) and (6). We show that Equation (4) is Hadmard differentiable with respect to F. Suppose Λ_t be a path such that $\Lambda_t\to\Lambda$ and $t^{-1}\{\Lambda_t-\Lambda\}\to g$ as $t\downarrow 0$. Then, as $t\downarrow 0$,

$$\begin{aligned}&t^{-1}\{\ell(x,\delta,z;\beta,\Lambda_t)-\ell(x,\delta,z;\beta,\Lambda)\}\\=\;&\delta t^{-1}\{\log\Delta\Lambda_t(x)-\log\Delta\Lambda(x)\}-e^{\beta'z}(t^{-1}\{\Lambda_t(x)-\Lambda(x)\})\\\to\;&\delta\frac{\Delta g(x)}{\Delta\Lambda(x)}-e^{\beta'z}g(x)\equiv[d_\Lambda\ell(\delta,z;\beta,\Lambda)](g)(x).\end{aligned}$$

This shows $\ell(x, \delta, z; \beta, \Lambda)$ is Hadmard differentiable with respect to Λ.

If we show Hadamard differentiability of the function $\hat{\Lambda}(t; \beta, F)$ (defined by Equation (3)) with respect to F, then, by the chain rule of Hadamard differentiable maps, Equation (4) is Hadamard differentiable with respect to F.

Suppose F_t be a path such that $F_t \to F$ and $t^{-1}\{F_t - F\} \to h$ as $t \downarrow 0$. Then, as $t \downarrow 0$,

$$\begin{aligned}
& t^{-1}\left\{\hat{\Lambda}(s; \beta, F_t) - \hat{\Lambda}(s; \beta, F)\right\} \\
= \quad & t^{-1}\left\{\int_0^s \frac{E_{F_t} dN(u)}{E_{F_t}[Y(u)e^{\beta^T Z}]} - \int_0^s \frac{E_F dN(u)}{E_F[Y(u)e^{\beta^T Z}]}\right\} \\
\to \quad & \int_0^s \frac{E_h dN(u)}{E_F[Y(u)e^{\beta^T Z}])} - \int_0^s \frac{E_F dN(u) E_h[Y(u)e^{\beta^T Z}]}{(E_F[Y(u)e^{\beta^T Z}])^2} \equiv [d_F\hat{\Lambda}(\beta, F)](h)(s),
\end{aligned}$$

Therefore, the function $\hat{\Lambda}(t; \beta, F)$ is Hadamard differentiable with respect to F and hence Condition (R1) is verified.

Condition (R2): We assumed that the efficient information matrix given by Equation (8) is invertible (C3).

Condition (R3): Let $\mathcal{F}$ be the set of cdf functions and for some $\rho > 0$ define $\mathcal{C}_\rho = \{F \in \mathcal{F} : \|F - F_0\|_\infty \le \rho\}$. We show that the class

$$\{\dot{\ell}(X, \delta, Z; \beta, F) : \beta \in \Theta, F \in \mathcal{C}_\rho\}$$

is Donsker with square integrable envelope function and the class

$$\left\{\ddot{\ell}(X, \delta, Z; \beta, F) : \beta \in \Theta, F \in \mathcal{C}_\rho\right\}$$

is Glivenko-Cantelli with integrable envelope function.

The set of cdf functions $\mathcal{F}$ is uniformly bounded Donsker. Hence the subset $\mathcal{C}_\rho \subset \mathcal{F}$ is uniformly bounded Donsker.

We assumed Z is bounded. The classes of functions $\{Z\}$, $\{N(t) : t \in [0, \tau]\}$ and $\{Y(t) : t \in [0, \tau]\}$ are uniformly bounded Donsker. The class $\{\beta^T Z : \beta \in \Theta\}$, with the compact set Θ, is uniformly bounded Donsker. It follows from $f(x) = e^x$ is a Lipschitz continuous function that $\{e^{\beta^T Z} : \beta \in \Theta\}$ is uniformly bounded Donsker.

By Example 2.10.8 in van der Vaart and Wellner (1996), the class of functions $\{Y(t)e^{\beta^T Z} : t \in [0, \tau], \beta \in \Theta\}$ is uniformly bounded Donsker. Since the map $(f, F) \to E_F f = \int f dF$ is Lipschitz, by Theorem 2.10.6 in van der Vaart and Wellner (1996), $\{E_F(Y(t)e^{\beta^T Z}) : t \in [0, \tau], \beta \in \Theta, F \in \mathcal{C}_\rho\}$ is Donsker since it is uniformly bounded. Similarly, the class $\{E_F(N(t)) : t \in [0, \tau], F \in \mathcal{C}_\rho\}$ is uniformly bounded Donsker.

We assumed $P(X \ge \tau) = E_0(Y(\tau)) > 0$. Since the map $F \to E_F f = \int f dF$ is continuous, there are $\rho > 0$ and $\rho_1 > 0$ such that for all $F \in \mathcal{C}_\rho$,

$$E_F(Y(\tau)) \ge \rho_1 > 0.$$

Since Z is bounded and β is in the compact set Θ, $0 < m < e^{\beta^T Z} < M$ for some $0 < m < M < \infty$. It follows that

$$0 < \rho_1 m \le m E_F(Y(s)) \le E_F(Y(s)e^{\beta^T Z}) \le M E_F(Y(s)) \le M < \infty.$$

By Example 2.10.9 in van der Vaart and Wellner (1996), the class

$$\left\{\frac{1}{E_F(Y(t)e^{\beta^T Z})} : t \in [0,\tau], \beta \in \Theta, F \in \mathcal{C}_\rho\right\}$$

is uniformly bounded Donsker.

Since the map $(f, F) \to E_F f = \int f dF$ is Lipschitz, by Theorem 2.10.6 in van der Vaart and Wellner (1996), the class of functions

$$\left\{\hat{\Lambda}(t;\beta,F) = \int_0^t \frac{dE_F[N(s)]}{E_F[Y(s)e^{\beta^T Z}]} : t \in [0,\tau], \beta \in \Theta, F \in \mathcal{C}_\rho\right\}$$

is uniformly bounded Donsker.

By Examples 2.10.7, 2.10.8 and 2.10.9 in van der Vaart and Wellner (1996), the class

$$\left\{N(t) - Y(t)e^{\beta^Z}\hat{\Lambda}(t;\beta,F) : t \in [0,\tau], \beta \in \Theta, F \in \mathcal{C}_\rho\right\}$$

is uniformly bounded Donsker.

Clearly, the class of functions

$$\left\{Z - \frac{E_F[ZY(t)e^{\beta^T Z}]}{E_F[Y(t)e^{\beta^T Z}]} : \beta \in \Theta, F \in \mathcal{C}_\rho\right\}$$

is uniformly bounded Donsker.

Again, since the map $(f, F) \to \int f dF$ is Lipschitz, by Theorem 2.10.6 in van der Vaart and Wellner (1996), the class of functions

$$\left\{\dot{\ell}(X,\delta,Z;\beta,F) = \int_0^\tau \left\{Z - \frac{E_F[ZY(t)e^{\beta^T Z}]}{E_F[Y(t)e^{\beta^T Z}]}\right\}\left\{dN(t) - Y(t)e^{\beta^T Z}d\hat{\Lambda}(t;\beta,F)\right\} : \beta \in \Theta, F \in \mathcal{C}_\rho\right\}$$

is uniformly bounded Donsker and hence it has square integrable envelope function.

Similarly, we can show that

$$\left\{\ddot{\ell}(X,\delta,Z;\beta,F) : \beta \in \Theta, F \in \mathcal{C}_\rho\right\}$$

is uniformly bounded Donsker, hence it is Glivenko-Cantelli with integrable envelope function.

References

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

Bailey, K. R. (1983). The asymptotic joint distribution of regression and survival parameter estimates in the Cox regression model. *Ann. Statist.* **11** 39–48.

Bailey, K. R. (1984). Asymptotic equivalence between the Cox estimator and the general ML estimators of regression and survival parameters in the Cox model. *Ann. Statist.* **12** 730–736.

Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

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

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89–99.

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

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

Gill, R.D. (1989). Non-and semi-parametric maximum likelihood estimators and the von Mises method (part 1). *Scand. J. Statist.* **16** 97–128.

Godambe, V.P. (1991). Orthogonality of estimating functions and nuisance parameters. *Biometrika* **78** 143–151.

Hall, W.S. and Newell, M.L. (1979). The mean value theorem for vector valued functions: a simple proof. *Math. Mag.* **52** 157–158.

Hirose, Y. (2009). Efficiency of profile likelihood in semi-parametric models, Submitted to Ann. Inst. Statist. Math.

Jacobsen, M. (1984). Maximum likelihood estimation in the multiplicative intensity model: a survey. *Int. Statist. Rev.* **52** 193–207.

Johansen, S. (1983). An extension of Cox's regression model *Int. Statist. Rev.* **51** 165–174.

Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Murphy, S.A. and van der Vaart, A.W. (2000). On profile likelihood (with discussion). *J. Amer. Statist. Assoc.* **95** 449–485.

Newey, W.K. (1990). Semi-parametric efficiency bounds. *J. Appl. Econ* **5** 99–135.

Newey, W.K. (1994). The asymptotic variance of semi-parametric estimators. *Econometrica* **62** 1349–1382.

Shapiro, A. (1990). On concepts of directional differentiability *J. Optim. Theory Appl.* **66** 477–487.

Tsiatis, A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93–108.

van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge Univ. Press, Cambridge.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.