ISSN 1066-5307, Mathematical Methods of Statistics, 2011, Vol. 20, No. 3, pp. 224–231. © Allerton Press, Inc., 2011.

Asymptotic Linear Expansion of Profile Likelihood in the Cox Model

Y. Hirose[1*]

[1] School of Math., Statist. and Operat. Research, Victoria Univ. of Wellington, New Zealand

Received February 25, 2010; in final form, May 27, 2011

Abstract—By introducing a new parameterization, Hirose [12] improved on the seminal work of Murphy and van der Vaart [16]: the improvement establishes the efficiency of the estimator through direct quadratic expansion of the profile likelihood, which requires fewer assumptions. This paper aims to demonstrate that the approach in [12] is fully applicable to the Cox proportional hazard model.

Keywords: Cox model, semiparametric model, profile likelihood, partial likelihood, efficiency, M-estimator, maximum likelihood estimator, efficient score, efficient information bound.

2000 Mathematics Subject Classification: 62F12, 62N02.

DOI: 10.3103/S1066530711030045

1. INTRODUCTION

We consider a semiparametric model of the form

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

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 is a compact set containing an open neighborhood of β_0 in $\mathbb{R}^m$, and $\mathcal{H}$ is a convex set containing η_0 in a Banach space $\mathcal{B}$.

Hirose [12] showed an asymptotic expansion of the profile likelihood by introducing a function $\hat\eta(\beta, F)$ of the parameter of interest β and a cdf F such that $\hat\eta(\beta_0, F_0) = \eta_0$ and the derivative

$$\left.\frac{\partial}{\partial\beta}\right|_{\beta=\beta_0} \log p(x;\beta,\hat\eta(\beta,F_0)) \tag{1}$$

is the efficient score function, where F_0 is the cdf for the density $p(x;\beta_0,\eta_0)$. Then for the empirical cdf F_n, the *profile log-likelihood* for β is defined by

$$\ell_n(\beta,\hat\eta(\beta,F_n)) = \sum_{i=1}^{n} \log p(X_i;\beta,\hat\eta(\beta,F_n)).$$

This representation of profile likelihood is an explicit function of sample size n, through F_n. In Theorem 1 at Section 2, we summarize the results in [12] with a modified proof that shows the asymptotic linear expansion of profile likelihood using this representation. This result gives an alternative approach to the one given by Murphy and van der Vaart [16]. The work in this paper is motivated by the referee's comments given for [12]: "Its assumption on the existence of $\hat\eta(\beta, F)$ in equation (1) still concerns me, as many semiparametric models used in survival context don't have this property. The paper should give more examples or acknowledge this limitation." In this paper, we aim to answer the comments by showing that, in the case of the Cox proportional hazard model, there exists a function $\hat\eta(\beta, F)$ and the results in [12] are fully applicable. This is demonstrated in Section 3.

*E-mail: Yuichi.Hirose@msor.vuw.ac.nz

2. SUMMARY OF RESULTS IN [12]: THEOREM AND ITS ASSUMPTIONS

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

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

For $\rho > 0$, let

$$\mathcal{C}_\rho = \{F \in \mathcal{F} \colon \|F - F_0\|_\infty < \rho\}. \tag{2}$$

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

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

The score function in the induced model is denoted by

$$\dot{\ell}(x, \beta, F) = \frac{\partial}{\partial \beta} \ell(x; \beta, F). \tag{3}$$

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 cdf F_n is $n^{1/2}$-consistent, i.e., $n^{1/2}\|F_n - F_0\|_\infty = O_P(1)$, and for each $(\beta, F) \in \Theta \times \mathcal{F}$, 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 $d_F \dot{\ell}(x, \beta, F)$. See [9] or [19] for the definition of Hadamard differentiability.)

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

(R3) There exists a $\rho > 0$ such that the class of functions $\{\dot{\ell}(x, \beta, F) \colon (\beta, F) \in \Theta \times \mathcal{C}_\rho\}$ is Donsker with square integrable envelope function, and the class of functions $\{\ddot{\ell}(x, \beta, F) \colon (\beta, F) \in \Theta \times \mathcal{C}_\rho\}$ is Glivenko–Cantelli with integrable envelope function. (Note: In the Introduction, we assumed the set Θ is a compact set containing an open neighborhood of β_0 in $\mathbb{R}^m$.)

Theorem 1. *Suppose that the assumptions* {(R0), (R1), (R2), (R3)} *are satisfied, then a consistent solution $\hat{\beta}_n$ to the estimating equation*

$$\sum_{i=1}^n \dot{\ell}(X_i, \hat{\beta}_n, F_n) = 0 \tag{4}$$

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

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = n^{-1/2} \sum_{i=1}^n (I^*)^{-1} \dot{\ell}^*(X_i) + o_P(1) \xrightarrow{d} N\left\{0, (I^*)^{-1}\right\},$$

where $N\{0, (I^*)^{-1}\}$ is a normal distribution with mean zero and variance $(I^*)^{-1}$. 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

$$E\{d_F\dot{\ell}(X,\beta_0,F_0)\} = -E\{\dot{\ell}(X,\beta_0,F_0)A(X,\beta_0,F_0)\} = 0 \quad \text{(the zero operator)}. \tag{5}$$

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 $n^{1/2}\|F_n - F_0\|_\infty = O_P(1)$ (condition (R1)), we have $\sup_{t\in[0,1]}\|F^*_{n,t} - F_0\|_\infty = o_P(1)$.

By the mean value theorem for vector-valued functions (see Hall and Newell [11]),

$$\begin{aligned}
&\|n^{1/2}E\{\dot{\ell}(X,\beta_0,F_n)\}\| \\
&\quad = \|n^{1/2}E\{\dot{\ell}(X,\beta_0,F_n)\} - n^{1/2}E\{\dot{\ell}(X,\beta_0,F_0)\}\| \\
&\quad \le \sup_{t\in[0,1]} \|E\{d_F\dot{\ell}(X,\beta_0,F^*_{n,t})\}\| n^{1/2}\|F_n - F_0\|_\infty \\
&\quad = \|E\{d_F\dot{\ell}(X,\beta_0,F_0)\} + o_p(1)\| n^{1/2}\|F_n - F_0\|_\infty \ \ \Big(\text{since} \sup_{t\in[0,1]}\|F^*_{n,t} - F_0\|_\infty = o_P(1)\Big) \\
&\quad = o_p(1)n^{1/2}\|F_n - F_0\|_\infty \ \ (\text{by Eq. (5)}) \\
&\quad = o_P(1) \ \ (\text{since } n^{1/2}\|F_n - F_0\|_\infty = O_P(1)).
\end{aligned} \tag{6}$$

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

$$E\|\dot{\ell}(X,\beta_0,F^*_n) - \dot{\ell}(X,\beta_0,F_0)\|^2 \xrightarrow{P} 0$$

and

$$E\|\ddot{\ell}(X,\beta^*_n,F^*_n) - \ddot{\ell}(X,\beta_0,F_0)\| \xrightarrow{P} 0.$$

Together with condition (R3), this implies that

$$n^{-1/2}\sum_{i=1}^n \big\{\dot{\ell}(X_i,\beta_0,F_n) - \dot{\ell}(X_i,\beta_0,F_0)\big\} = n^{1/2}E\big\{\dot{\ell}(X,\beta_0,F_n) - \dot{\ell}(X,\beta_0,F_0)\big\} + o_P(1). \tag{7}$$

By Lemma 13.3 in [15], and for every $(\beta^*_n,F^*_n) \xrightarrow{P} (\beta_0,F_0)$,

$$n^{-1}\sum_{i=1}^n \ddot{\ell}(X_i,\beta^*_n,F^*_n) \xrightarrow{P} E\ddot{\ell}(X_i,\beta_0,F_0) = -I^*. \tag{8}$$

By combining Eqs. (6) and (7), we get

$$n^{-1/2}\sum_{i=1}^n \dot{\ell}(X_i,\beta_0,F_n) = n^{-1/2}\sum_{i=1}^n \dot{\ell}(X_i,\beta_0,F_0) + o_P(1). \tag{9}$$

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

$$\begin{aligned}
0 &= n^{-1/2}\sum_{i=1}^n \dot{\ell}(X_i,\hat{\beta}_n,F_n) \\
&= n^{-1/2}\sum_{i=1}^n \dot{\ell}(X_i,\beta_0,F_n) + n^{-1}\sum_{i=1}^n \ddot{\ell}(X_i,\beta^*_n,F_n)n^{1/2}(\hat{\beta}_n - \beta_0)
\end{aligned}$$

$$= n^{-1/2}\sum_{i=1}^{n}\dot\ell(X_i,\beta_0,F_0) + o_P(1) + \{-I^* + o_P(1)\}n^{1/2}(\hat\beta_n-\beta_0),$$

where the last equality is by Eqs. (8) and (9). Hence, by condition (R2),

$$n^{1/2}(\hat\beta_n-\beta_0) = (I^*)^{-1}n^{-1/2}\sum_{i=1}^{n}\dot\ell(X_i,\beta_0,F_0) + o_P(1)\{1+n^{1/2}(\hat\beta_n-\beta_0)\}.$$

Since $(I^*)^{-1}n^{-1/2}\sum_{i=1}^{n}\dot\ell(X_i,\beta_0,F_0) = O_P(1)$, this equality implies $n^{1/2}(\hat\beta_n-\beta_0) = O_P(1)$ and

$$n^{1/2}(\hat\beta_n-\beta_0) = (I^*)^{-1}n^{-1/2}\sum_{i=1}^{n}\dot\ell(X_i,\beta_0,F_0) + o_P(1).$$

□

3. APPLICATION TO COX'S PROPORTIONAL HAZARD MODEL

Cox [7] introduced a proportional hazard model, known as the Cox model, where the cumulative hazard function of the survival time T for a subject with covariate $Z \in R^k$ is given by

$$\Lambda(t \mid Z) = e^{\beta^T Z}\Lambda(t). \tag{10}$$

Here $\Lambda(t)$ is an unspecified baseline cumulative hazard function. Murphy and van der Vaart [16] 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.

Suppose we observe (X,δ,Z) in time interval $[0,\tau]$, where $X = T\wedge C$, $\delta = 1_{\{T\le C\}}$, $Z\in R^k$ is a regression covariate, T is a right-censored failure time with cumulative hazard as given by Eq. (10), 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) for the usual discrete extension of the model is given by

$$\ell(X,\delta,Z;\beta,\Lambda) = \{\beta^T Z + \log\Delta\Lambda(X)\}\delta - e^{\beta^T Z}\Lambda(X), \tag{11}$$

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

We denote the empirical cdf of the observations $\{(X_i,\delta_i,Z_i)\colon i=1,\ldots,n\}$ by F_n, and the cdf that generates an observation (X,δ,Z) by F_0. For any cdf F of (X,δ,Z), write $E_F(f) = \int f\,dF$ and define a function

$$\hat\Lambda_{\beta,F}(t) = \int_0^t \frac{E_F\{dN(s)\}}{E_F\{Y(s)e^{\beta^T Z}\}}. \tag{12}$$

Note that (i) for an estimator $\hat\beta$ of β, the Breslow estimator is given by $\hat\Lambda_{\hat\beta,F_n}(t)$; (ii) if (β_0,Λ_0) is the true value of the parameter (β,Λ), then we have $E_{F_0}N(t) = \int_0^t E_{F_0}\{Y(s)e^{\beta_0^T Z}\}\,d\Lambda_0(s)$ and $\hat\Lambda_{\beta_0,F_0}(t) = \Lambda_0(t)$.

We substitute the function $\hat\Lambda_{\beta,F}(t)$ into the log-likelihood (Eq. (11)) and call it the induced model. The log-likelihood for an observation (X,δ,Z) in the induced model is

$$\ell(X,\delta,Z;\beta,\hat\Lambda_{\beta,F}) = \left[\beta^T Z + \log\frac{E_F\{\Delta N(X)\}}{E_F\{Y(X)e^{\beta^T Z}\}}\right]\delta - e^{\beta^T Z}\int_0^X \frac{E_F\{dN(s)\}}{E_F\{Y(s)e^{\beta^T Z}\}}. \tag{13}$$

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

$$\dot\ell(X,\delta,Z;\beta,F)=\frac{\partial}{\partial\beta}\ell(X,\delta,Z;\beta,\hat\Lambda_{\beta,F})=\int_0^\tau\left[Z-\frac{E_F\{ZY(t)e^{\beta^TZ}\}}{E_F\{Y(t)e^{\beta^TZ}\}}\right]d\hat M_{\beta,F}(t) \tag{14}$$

and

$$\begin{aligned}\ddot\ell(X,\delta,Z;\beta,F)&=\frac{\partial}{\partial\beta}\dot\ell(X,\delta,Z;\beta,F)\\&=-\int_0^\tau\left(\frac{E_F\{Z^{\otimes2}Y(t)e^{\beta^TZ}\}}{E_F\{Y(t)e^{\beta^TZ}\}}-\frac{[E_F\{ZY(t)e^{\beta^TZ}\}]^{\otimes2}}{[E_F\{Y(t)e^{\beta^TZ}\}]^2}\right)\hat M_{\beta,F}(t)\\&\quad-\int_0^\tau\left[Z-\frac{E_F\{ZY(t)e^{\beta^TZ}\}}{E_F\{Y(t)e^{\beta^TZ}\}}\right]^{\otimes2}Y(t)e^{\beta^TZ}\,d\hat\Lambda_{\beta,F}(t).\end{aligned} \tag{15}$$

Here $\hat M_{\beta,F}(t)=N(t)-Y(t)e^{\beta^TZ}\hat\Lambda_{\beta,F}(t)$.

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

$$\dot\ell(X,\delta,Z;\beta_0,F_0)=\int_0^\tau\left[Z-\frac{E_{F_0}\{ZY(t)e^{\beta_0^TZ}\}}{E_{F_0}\{Y(t)e^{\beta_0^TZ}\}}\right]dM(t)=:\dot\ell^*(X,\delta,Z) \tag{16}$$

is the efficient score function $\dot\ell^*(X,\delta,Z)$ (cf. [16]) and the efficient information matrix is given by

$$I^*:=-E\{\ddot\ell(X,\delta,Z;\beta_0,F_0)\}=E\left(\int_0^\tau\left[Z-\frac{E_{F_0}\{ZY(t)e^{\beta_0^TZ}\}}{E_{F_0}\{Y(t)e^{\beta_0^TZ}\}}\right]^{\otimes2}Y(t)e^{\beta_0^TZ}\,d\Lambda_0(t)\right). \tag{17}$$

We assume that

(C1) $\mathrm{pr}(X\ge\tau)=E\{Y(\tau)\}>0$, and

(C2) the range of Z is bounded and the parameter β is in a compact set Θ that contains an open neighborhood of β_0;

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

(C4) the empirical cdf F_n is $n^{1/2}$-consistent, i.e., $n^{1/2}\|F_n-F_0\|_\infty=O_P(1)$.

Now, the partial likelihood and the corresponding score equation are given by

$$L_n(\beta)=\prod_{i=1}^n\prod_{0\le t\le\tau}\left\{\frac{Y_i(t)e^{\beta^TZ_i}}{\sum_{j=1}^nY_j(t)e^{\beta^TZ_j}}\right\}^{\Delta N_i(t)}$$

and

$$\frac{\partial}{\partial\beta}\log L_n(\beta)=\sum_{i=1}^n\int_0^\tau\left[Z_i-\frac{E_{F_n}\{ZY(t)e^{\beta^TZ}\}}{E_{F_n}\{Y(t)e^{\beta^TZ}\}}\right]dN_i(t)=0. \tag{18}$$

At the same time, for the empirical cdf F_n, $\sum_{i=1}^n\ell(X_i,\delta_i,Z_i;\beta,\hat\Lambda_{\beta,F_n})$ gives a version of profile log-likelihood, where $\ell(X,\delta,Z;\beta,\Lambda)$ is the log-likelihood for an observation given by Eq. (11) and $\hat\Lambda_{\beta,F_n}$ is given by Eq. (12). Then the score equation for the profile likelihood is

$$\sum_{i=1}^n\int_0^\tau\left[Z_i-\frac{E_{F_n}\{ZY(t)e^{\beta^TZ}\}}{E_{F_n}\{Y(t)e^{\beta^TZ}\}}\right]\hat M_{\beta,F_n}(t)=0. \tag{19}$$

Since

$$\sum_{i=1}^{n}\int_{0}^{\tau}\left[Z_i-\frac{E_{F_n}\{ZY(t)e^{\beta^T Z}\}}{E_{F_n}\{Y(t)e^{\beta^T Z}\}}\right]Y_i(t)e^{\beta^T Z_i}\,d\hat{\Lambda}_{\beta,F_n}(t)=0,$$

the score equations (18) and (19) are the same equation.

The following theorem shows asymptotic linearity of the estimators based on the profile likelihood and the partial likelihood.

Theorem 2. *Suppose* (C1)–(C4). *The solution* $\hat{\beta}_n$ *to the score equation for the profile likelihood* (*Eq.* (19)) *and the solution* $\hat{\beta}_n$ *to the score equation for the partial likelihood* (*Eq.* (18)) *are both asymptotically linear estimators with the efficient influence function* $(I^*)^{-1}\dot{\ell}^*$, *so that*

$$n^{1/2}(\hat{\beta}_n-\beta_0)=n^{-1/2}\sum_{i=1}^{n}(I^*)^{-1}\dot{\ell}^*(X_i,\delta_i,Z_i)+o_P(1)\xrightarrow{d}N\{0,(I^*)^{-1}\},\tag{20}$$

where the efficient score $\dot{\ell}^*$ *and the efficient information* I^* *are given by Eqs.* (16) *and* (17)*, respectively, and* $N\{0,(I^*)^{-1}\}$ *is a normal distribution with mean zero and variance* $(I^*)^{-1}$.

Proof. In the following, the set of assumptions (R0)–(R3) in Theorem 1 are verified. Then the claim follows from Theorem 1.

Condition (R0): This condition is verified by three lines below Eqs. (12) and (16).

Condition (R1): Equation (13) is twice continuously differentiable with respect to β with the first and second derivatives (14) and (15). We show that Eq. (13) is Hadamard 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 $t\downarrow 0$,

$$\begin{aligned}t^{-1}\{\ell(x,\delta,z;\beta,\Lambda_t)-\ell(x,\delta,z;\beta,\Lambda)\}&=\delta t^{-1}\{\log\Delta\Lambda_t(x)-\log\Delta\Lambda(x)\}-e^{\beta^T z}t^{-1}\{\Lambda_t(x)-\Lambda(x)\}\\&\to\delta\frac{\Delta g(x)}{\Delta\Lambda(x)}-e^{\beta^T z}g(x)\equiv d_\Lambda\ell(\delta,z;\beta,\Lambda)(g)(x).\end{aligned}$$

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

If we show Hadamard differentiability of the function $\hat{\Lambda}_{\beta,F}(t)$ (defined by Eq. (12)) with respect to F, then, by the chain rule of Hadamard differentiable maps, Eq. (13) 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 $t\downarrow 0$,

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

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

Condition (R2): We assumed that the efficient information matrix given by Eq. (17) is invertible (C3).

Condition (R3): Let $\mathcal{F}$ be the set of cdf functions and $\mathcal{C}_\rho$ be defined by (2). For some $\rho>0$, we show that the class

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

is Donsker with square integrable envelope function and the class

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

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 $\{N(t)\colon t \in [0,\tau]\}$ and $\{Y(t)\colon t \in [0,\tau]\}$ are uniformly bounded Donsker. The class $\{\beta^T Z\colon \beta \in \Theta\}$ with compact set Θ is uniformly bounded Donsker. Since $f(x) = e^x$ is a Lipschitz continuous function, we have that $\{e^{\beta^T Z}\colon \beta \in \Theta\}$ is uniformly bounded Donsker.

By Example 2.10.8 in van der Vaart and Wellner [22], the class of functions $\{Y(t)e^{\beta^T Z}\colon t \in [0,\tau], \beta \in \Theta\}$ 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 [22], $\{E_F\{Y(t)e^{\beta^T Z}\}\colon t \in [0,\tau], \beta \in \Theta, F \in \mathcal{C}_\rho\}$ is Donsker since it is uniformly bounded. Similarly, the class $\{E_F\{N(t)\}\colon t \in [0,\tau], F \in \mathcal{C}_\rho\}$ is uniformly bounded Donsker.

We assumed $P(X \geq \tau) = E\{Y(\tau)\} > 0$. 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 $F \in \mathcal{C}_\rho$,

$$E_F\{Y(\tau)\} \geq \rho_1 > 0.$$

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

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

By Example 2.10.9 in van der Vaart and Wellner [22], the class

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

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 [22], the class of functions

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

is uniformly bounded Donsker.

By Examples 2.10.7, 2.10.8, and 2.10.9 in van der Vaart and Wellner [22], the class

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

is uniformly bounded Donsker.

Clearly, the class of functions

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

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 [22], the class of functions

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

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

Similarly, we can show that

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

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

REFERENCES

1. P. K. Andersen and R. D. Gill, “Cox’s Regression Model for Counting Processes: a Large Sample Study”, Ann. Statist. **10**, 1100–1120 (1982).
2. K. R. Bailey, “The Asymptotic Joint Distribution of Regression and Survival Parameter Estimates in the Cox Regression Model”, Ann. Statist. **11**, 39–48 (1983).
3. K. R. Bailey, “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 (1984).
4. J. M. Begun, W. J. Hall, W. M. Huang, and J. A. Wellner, “Information and Asymptotic Efficiency in Parametric–Nonparametric Models”, Ann. Statist. **11**, 432–452 (1983).
5. P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models* (Johns Hopkins Univ. Press, Baltimore, 1993).
6. N. Breslow, “Covariance Analysis of Censored Survival Data”, Biometrics **30**, 89–99 (1974).
7. D. R. Cox, “Regression Models and Life Tables (with Discussion)”, J. Roy. Statist. Soc. B **34**, 187–220 (1972).
8. D. R. Cox, “Partial Likelihood”, Biometrika **62**, 269–276 (1975).
9. R. D. Gill, “Non- and Semiparametric Maximum Likelihood Estimators and the von Mises Method (Part 1)”, Scand. J. Statist. **16**, 97–128 (1989).
10. V. P. Godambe, “Orthogonality of Estimating Functions and Nuisance Parameters”, Biometrika **78**, 143–151 (1991).
11. W. S. Hall and M. L. Newell, “The Mean Value Theorem for Vector Valued Functions: a Simple Proof”, Math. Mag. **52**, 157–158 (1979).
12. Y. Hirose, “Efficiency of Profile Likelihood in Semiparametric Models”, Ann. Inst. Statist. Math. (2010). DOI 10.1007/s10463-010-0280-y.
13. M. Jacobsen, “Maximum Likelihood Estimation in the Multiplicative Intensity Model: a Survey”, Int. Statist. Rev. **52**, 193–207 (1984).
14. S. Johansen, “An Extension of Cox’s Regression Model”, Int. Statist. Rev. **51**, 165–174 (1983).
15. M. R. Kosorok, *Introduction to Empirical Processes and Semiparametric Inference* (Springer, New York, 2008).
16. S. A. Murphy and A. W. van der Vaart, “On Profile Likelihood (with Discussion)”, J. Amer. Statist. Assoc. **95**, 449–485 (2000).
17. W. K. Newey, “Semiparametric Efficiency Bounds”, J. Appl. Econ. **5**, 99–135 (1990).
18. W. K. Newey, “The Asymptotic Variance of Semiparametric Estimators”, Econometrica **62**, 1349–1382 (1994).
19. A. Shapiro, “On Concepts of Directional Differentiability”, J. Optim. Theory Appl. **66**, 477–487 (1990).
20. A. Tsiatis, “A Large Sample Study of Cox’s Regression Model”, Ann. Statist. **9**, 93–108 (1981).
21. A. W. van der Vaart, *Asymptotic Statistics* (Cambridge Univ. Press, Cambridge, 1998).
22. A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes* (Springer, New York, 1996).