An asymptotic linear representation for the Breslow estimator

Hendrik P. Lopuhaä and Gabriela F. Nane
Delft University of Technology
November 25, 2011

Abstract

We provide an asymptotic linear representation for the Breslow estimator for the baseline cumulative hazard function in the Cox model. The representation consists of an average of independent random variables and a term involving the difference between the maximum partial likelihood estimator and the underlying regression parameter. The order of the remainder term is arbitrarily close to $n^{-1}$.

1 Introduction

Within survival analysis, the Cox proportional hazards model is one of the most acknowledged approaches to model right-censored time to event data in the presence of covariates. Cox (1972) introduced this semiparametric model and focused on estimating the underlying regression coefficients of the covariates. His proposed estimator was later shown (Cox, 1975) to be a partial maximum likelihood estimator and its asymptotic properties were broadly enquired, see for example Tsiatis (1981), Andersen et al. (1993), Oakes (1977), Slud (1982). Furthermore, in the Cox model, different functionals of the lifetime distribution are commonly investigated. The (cumulative) hazard function is of particular interest, as it represents an important feature of a process under study, e.g., death or a certain disease. In the discussion following the Cox’s (1972) paper, Breslow proposed a nonparametric maximum likelihood estimator for the baseline cumulative hazard function. Asymptotic properties of the Breslow estimator, such as consistency and the asymptotic distribution, were derived by Tsiatis (1981) and Andersen et al. (1993). For a general overview of the Breslow estimator, see Lin (2007).

Estimators in censorship models in the case of no covariates, i.e., the Kaplan-Meier estimator or the Nelson-Aalen estimator, perceived tremendous interest, especially in the 1980s. Established large sample properties include consistency and asymptotic normality (Breslow and Crowley, 1974), rate of strong uniform consistency (Csörgő and Horváth, 1983), strong approximation or Hungarian embedding (Burke, Csörgő and Horváth, 1981) and linearization results (Lo and Singh, 1985). Lo and Singh (1985) expressed the difference between the Kaplan-Meier estimator and the underlying distribution function in terms of a sum of independent identically distributed random variables, almost surely, with a remainder term of the order $n^{-3/4}(\log n)^{3/4}$. The rate was later improved to $n^{-1} \log n$ by Lo, Mack and Wang (1989). To the authors’ best knowledge, a strong approximation result for the Breslow estimator is not available in the literature. Kosorok (2008) establishes a representation of the Breslow estimator in terms of counting processes. Although this can be turned into an asymptotic
linear representation of the type as in Lo and Singh (1985), the covariates are assumed to be in a bounded set and the remainder term is only shown to be of the order $o_p(n^{-1/2})$.

In this paper, we derive a similar linearization result for the Breslow estimator, i.e., we prove that the difference between the estimator $\Lambda_n(t)$ and the cumulative baseline hazard function $\Lambda_0(t)$ can be represented as a sum of independent random variables and a term involving the difference between the regression parameter and its partial maximum likelihood estimator. However, we allow unbounded covariates and we show that the remainder term is of the order $n^{-1}a_n^{-1}$, where $a_n$ may be any sequence tending to zero. As $a_n$ can be chosen to converge to zero arbitrarily slow, this means that the order of the remainder term is arbitrarily close to $n^{-1}$. Our main motivation is isotonic estimation of the baseline distribution in the Cox model. An example is the Grenander type estimator $\tilde{\lambda}_n$ for an increasing baseline hazard $\lambda_0$, considered in Lopuhaä and Nane (2011), which is defined as the left-hand slope of the greatest convex minorant of the Breslow estimator. The limit behavior of $\tilde{\lambda}_n$ at a fixed point $t_0$ essentially follows from the limit behavior of the process

$$t \mapsto n^{2/3} \left\{ (\Lambda_n - \Lambda_0) \left( t_0 + n^{-1/3}t \right) - (\Lambda_n - \Lambda_0)(t_0) \right\}.$$  

In the absence of a strong approximation result for the process $\Lambda_n - \Lambda_0$, an alternative to obtain the limit process is to apply the results in Kim and Pollard (1990) to the linear representation for $\Lambda_n - \Lambda_0$, provided that the remaining terms in the representation are of order smaller than $n^{-2/3}$. This cannot be ensured by the representation in Kosorok (2008), whereas the order $n^{-1}a_n^{-1}$ can be chosen sufficiently small, for suitable choices of $a_n$.

The paper is organized as follows. The Cox model and the Breslow estimator are introduced in Section 2. Furthermore, the assumed conditions are stated and notations are presented. Section 3 is devoted to the main result of the paper and its proof as well as to preparatory lemmas.

## 2 Notation and assumptions

Survival analysis focuses on the study of time to event data and the random variable of interest, the survival time, is denoted by $X$. However, the survival data is usually subject to right-censoring, indicating that there are subjects still alive or lost to follow-up at the end of the study. The random variable $C$ denotes the censoring time. Now, define $T = \min(X, C)$ as the generic follow-up time and $\Delta = \{X \leq C\}$ as its corresponding indicator, where $\{\cdot\}$ denotes the indicator function. Hence, $\Delta = 1$ indicates an observed survival time, while $\Delta = 0$ indicates a censored observation. Moreover, suppose that at the beginning of the study, extra information such as sex, age, status of a disease, etc. is recorded for each subject as covariates. Let $Z \in \mathbb{R}^p$ denote the covariate vector. Therefore, suppose we observe the following independent, identically distributed triplets $(T_i, \Delta_i, Z_i)$, with $i = 1, 2, \ldots, n$. The censoring mechanism is assumed to be non-informative. Hence, given the covariate $Z$, the survival time $X$ is assumed to be independent of the censoring time $C$. Moreover, the covariate vector $Z \in \mathbb{R}^p$ is assumed to be time invariant and non-degenerate.

The hazard function is of particular interest in survival analysis, as it represents an important feature of the survival time. In respect to this, in the Cox model, the distribution of the survival time is related to the corresponding covariate by

$$\lambda(x|z) = \lambda_0(x) e^{\beta'z},$$
where \( \lambda(x|z) \) is the hazard function for a subject with covariate vector \( z \in \mathbb{R}^p \), \( \lambda_0 \) represents the underlying baseline hazard function, corresponding to a subject with \( z = 0 \) and \( \beta \in \mathbb{R}^p \) is the vector of the underlying regression coefficients. Conditionally on \( Z = z \), the survival time \( X \) is assumed to be a nonnegative random variable, with an absolutely continuous distribution function \( F(x|z) \) with density \( f(x|z) \). The same assumptions hold for the censoring variable \( C \) and its distribution function \( G \). Let \( H \) be the distribution function of the follow-up time \( T \) and let \( \tau_H = \inf \{ t : H(t) = 1 \} \) be the end point of the support of \( H \). Moreover, let \( \tau_F \) and \( \tau_G \) be the end points of the support of \( F \) and \( G \), respectively. We employ the usual assumptions for deriving large sample properties of Cox proportional hazards estimators, see for example Tsiatis (1981).

(A1) \( \tau_H = \tau_G < \tau_F \).

(A2) There exists \( \varepsilon > 0 \) such that

\[
\sup_{|\beta - \beta_0| \leq \varepsilon} \mathbb{E} \left[ |Z| e^{2\beta'Z} \right] < \infty,
\]

where \( | \cdot | \) denotes the Euclidean norm.

Let \( X_{(1)} < X_{(2)} < \cdots < X_{(m)} \) denote the ordered, observed survival times. Cox (1972) introduced the proportional hazards model and proposed the partial likelihood estimator \( \hat{\beta} \) (Cox, 1975) as an estimator for the underlying regression coefficients \( \beta_0 \). Breslow (1972) on the other hand, focused on estimating the baseline cumulative hazard function, \( \Lambda_0(x) = \int_0^x \lambda_0(u) \, du \) and proposed

\[
\Lambda_n(x) = \sum_{i : X_{(i)} \leq x} \frac{d_i}{\sum_{j=1}^n \{ T_j \geq X_{(i)} \} e^{\hat{\beta}'Z_j}},
\]

(2.1)
as an estimator for \( \Lambda_0 \), where \( d_i \) is the number of events at \( X_{(i)} \) and \( \hat{\beta} \) is the partial maximum likelihood estimator of the regression coefficients. The estimator \( \Lambda_n \) is most commonly referred to as the Breslow estimator. Under the assumption of a piecewise constant baseline hazard function and assuming that all the censoring times are shifted at the preceding observed survival time, Breslow showed that the partial maximum likelihood estimator \( \hat{\beta} \) along with the baseline cumulative hazard estimator \( \Lambda_n \) can be obtained by jointly maximizing the full loglikelihood function.

Let

\[
\Phi(\beta, x) = \int \{ u \geq x \} e^{\beta'z} \, dP(u, \delta, z),
\]

\[
\Phi_n(\beta, x) = \int \{ u \geq x \} e^{\beta'z} \, dP_n(u, \delta, z),
\]

(2.2)

where \( P \) is the underlying probability measure corresponding to the distribution of \( (T, \Delta, Z) \) and \( P_n \) is the empirical measure of the triplets \( (T_i, \Delta_i, Z_i) \)’s. Furthermore, let \( H^{uc}(x) = \mathbb{P}(T \leq x, \Delta = 1) \) be the sub-distribution function of the uncensored observations. Then, using the derivations in Tsiatis (1981), it can be deduced that

\[
\lambda_0(u) = \frac{dH^{uc}(u)/du}{\Phi(\beta_0, u)}. \quad (2.3)
\]
Consequently, it can be derived that

\[ \Lambda_0(x) = \int \frac{\delta\{u \leq x\}}{\Phi(\beta_0, u)} \, dP(u, \delta, z). \tag{2.4} \]

Note that from (A1), it follows that \( \Lambda_0(\tau_H) < \infty \). In view of the above expression, an intuitive baseline cumulative hazard function estimator is obtained by replacing \( \Phi \) in (2.4) by \( \Phi_n \) and by plugging in \( \hat{\beta} \), which yields exactly the Breslow estimator in (2.1),

\[ \Lambda_n(x) = \int \frac{\delta\{u \leq x\}}{\Phi_n(\beta, u)} \, dP_n(u, \delta, z). \tag{2.5} \]

The asymptotic properties of the Breslow estimator have been investigated extensively, see for example Tsiatis (1981), Andersen et al.(1993) and Lin (2007) for a general overview. Kosorok (2008) provided the strong uniform consistency of the Breslow estimator and the process convergence of \( \sqrt{n}(\Lambda_n - \Lambda_0) \), yet under the strong assumption of bounded covariates. Using standard empirical processes methods, Lopuhaä and Nane (2011) establish strong uniform consistency at rate \( n^{-1/2} \) for the Breslow estimator under the relatively mild conditions (A1) and (A2).

### 3 Asymptotic representation

The following two lemmas will be used in proving the main result of the paper.

**LEMMA 3.1.** Suppose that condition (A2) holds and let \( \Phi_n \) and \( \Phi \) be defined in (2.2). With \( \epsilon > 0 \) taken from (A2), for \( |\beta - \beta_0| < \epsilon \), let

\[ D_n^{(1)}(\beta, x) = \frac{\partial \Phi_n(\beta, x)}{\partial \beta} = \int \{u \geq x\} z e^{\beta'z} \, dP_n(u, \delta, z) \in \mathbb{R}^p. \]

Then,

\[ \sqrt{n} \sup_{x \in \mathbb{R}} |\Phi_n(\beta_0, x) - \Phi(\beta_0, x)| = O_p(1) \]

\[ \sqrt{n} \sup_{x \in \mathbb{R}} \left| D_n^{(1)}(\beta_0, x) - D^{(1)}(\beta_0, x) \right| = O_p(1). \tag{3.2} \]

**Proof.** Consider the class of functions \( \mathcal{G} = \{g(u, z; x) : x \in \mathbb{R}\} \), where, for each \( x \in \mathbb{R} \) and \( \beta_0 \in \mathbb{R}^p \) fixed,

\[ g(u, z; x) = \{u \geq x\} \exp(\beta_0'z) \]

is a product of an indicator and a fixed function. It follows that \( \mathcal{G} \) is a VC-subgraph class (e.g., see Lemma 2.6.18 in van der Vaart and Wellner, 1996) and its envelope \( G = \exp(\beta_0'z) \) is square integrable under condition (A2). Standard results from empirical process theory (van der Vaart and Wellner, 1996) yield that the class of functions \( \mathcal{G} \) is a Donsker class, i.e.,

\[ \sqrt{n} \int g(u, z; x) \, d(\mathbb{P}_n - P)(u, \delta, z) = O_p(1), \]
so that the first statement in (3.2) follows by continuous mapping theorem. To prove the second statement, it suffices to consider each $j$th coordinate, for $j = 1, 2, \ldots, p$ fixed. In this case, we deal with the class $G_j = \{g_j(u, z; x) : x \in \mathbb{R}\}$, where 

$$g_j(u, z; x) = \{u \geq x\}z_j \exp(\beta'_0z).$$

From here the argument is exactly the same, which proves the lemma.

**Lemma 3.2.** Assume (A1) and (A2). Then, for all $0 < M < \tau_H$,

$$a_n n \sup_{x \in [0, M]} \left| \int \delta\{u \leq x\} \left( \frac{1}{\Phi_n(\beta_0, u)} - \frac{1}{\Phi(\beta_0, u)} \right) d(P - P)(u, \delta, z) \right| = O_p(1),$$

for any sequence $a_n \to 0$.

**Proof.** Consider the class of functions $F_n = \{f_n(u, \delta, z; x) : 0 \leq x \leq M\}$, where 

$$f_n(u, \delta, z; x) = \delta\{u \leq x\} \left( \frac{1}{\Phi_n(\beta_0, u)} - \frac{1}{\Phi(\beta_0, u)} \right).$$

Correspondingly, consider the class $G_{n,M,\alpha}$ consisting of functions 

$$g(u, \delta, z; y, \Psi) = \delta\{u \leq y\} \left( \frac{1}{\Psi(u)} - \frac{1}{\Phi(\beta_0, u)} \right),$$

where $0 \leq y \leq M$ and $\Psi$ is nonincreasing left continuous, such that 

$$\Psi(M) \geq K$$

and 

$$\sup_{u \in [0, M]} |\Psi(u) - \Phi(\beta_0, u)| \leq \alpha,$$

where $K = \Phi(\beta_0, M)/2$. Then, for any $\alpha > 0$, we have $P(F_n \subset G_{n,M,\alpha}) \to 1$, by Lemma 3.1. Furthermore, the class $G_{n,M,\alpha}$ has envelope $G(u, \delta, z) = \alpha/K^2$. Since the functions in $G_{n,M,\alpha}$ are products of indicators and a difference of bounded monotone functions, its entropy with bracketing satisfies 

$$\log N_{[]}([\varepsilon, G_{n,M,\alpha}, L_2(P)]) \lesssim \frac{1}{\varepsilon},$$

see e.g., Theorem 2.7.5 in van der Vaart and Wellner (1996) and Lemma 9.25 in Kosorok (2008). Hence, for any $\delta > 0$, the bracketing integral 

$$J_{[]}(\delta, G_{n,M,\alpha}, L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[]}([\varepsilon\|G\|_2, G_{n,M,\alpha}, L_2(P)])} d\varepsilon < \infty.$$

By Theorem 2.14.2 in van der Vaart and Wellner (1996), we have 

$$\mathbb{E} \left\| \sqrt{n} \int g(u, \delta, z; y, \Psi) d(P - P)(u, \delta, z) \right\|_{G_{n,M,\alpha}} \leq J_{[]}(1, G_{n,M,\alpha}, L_2(P))\|G\|_{P,2} = O(\alpha),$$

where $\|\cdot\|_\mathcal{F}$ denotes the supremum over the class of functions $\mathcal{F}$. Now, let $(a_n)$ be a sequence that tends to zero. Then, according to (3.2), 

$$a_n \sqrt{n} \sup_{x \in \mathbb{R}} |\Phi_n(\beta_0, x) - \Phi(\beta_0, x)| \to 0,$$
in probability. Therefore, if we choose $\alpha = n^{-1/2}a_n^{-1}$, this gives
\[
E \left\| \int g(u, \delta, z; y, \Psi) d(\mathbb{P}_n - P)(u, \delta, z) \right\|_{\mathcal{G}_{n,m,\alpha}} = \mathcal{O}(na_n^{-1})
\]
and hence, by the Markov inequality, this proves the lemma. \qed

The asymptotic linear representation of the Breslow estimator is provided by the next theorem.

**Theorem 3.1.** Assume (A1) and (A2). Let $\Phi$ and $D^{(1)}$ be defined in (2.2) and (3.1). Then, for all $0 < M < \tau_H$ and $x \in [0, M]$,
\[
\Lambda_n(x) - \Lambda_0(x) = \frac{1}{n} \sum_{i=1}^{n} \xi(T_i, \Delta_i, Z_i; x) + (\hat{\beta} - \beta_0)'A_0(x) + R_n(x),
\]
where $\hat{\beta}$ is the maximum partial likelihood estimator,
\[
A_0(x) = \int_{0}^{x} \frac{D^{(1)}(\beta_0, u)}{\Phi(\beta_0, u)} \lambda_0(u) \, du
\]
and
\[
\xi(t, \delta, z; x) = -e^{\beta_0 z} \int_{0}^{x \wedge t} \frac{\lambda_0(u)}{\Phi(\beta_0, u)} \, du + \delta\{t \leq x\} \frac{\Phi(\delta, t)}{\Phi(\beta_0, t)}
\]
and $R_n$ is such that
\[
\sup_{x \in [0, M]} |R_n(x)| = \mathcal{O}_p(n^{-1}a_n^{-1}),
\]
for any sequence $a_n \to 0$.

**Proof.** For $\beta \in \mathbb{R}^p$, define
\[
\Lambda_n(\beta, x) = \int \delta\{u \leq x\} \frac{1}{\Phi_n(\beta, u)} \, d\mathbb{P}_n(u, \delta, z).
\]
Hence, the Breslow estimator in (2.5) can also be written as $\Lambda_n(\hat{\beta}, x)$. For $x \in [0, M]$, consider the following decomposition
\[
\Lambda_n(x) - \Lambda_0(x) = T_{n1}(x) + T_{n2}(x),
\]
where $T_{n1}(x) = \Lambda_n(\hat{\beta}, x) - \Lambda_n(\beta_0, x)$ and $T_{n2}(x) = \Lambda_n(\beta_0, x) - \Lambda_0(x)$.

For the term $T_{n1}$, first notice that a Taylor expansion of $\Lambda_n(\cdot, x)$ around $\beta_0$ yields that
\[
\Lambda_n(\hat{\beta}, x) - \Lambda_n(\beta_0, x) = (\hat{\beta} - \beta_0)'A_n(x) + \frac{1}{2}(\hat{\beta} - \beta_0)'R_{n1}(x)(\hat{\beta} - \beta_0),
\]
where the vector $A_n$ and matrix $R_{n1}$ are given by
\[
A_n(x) = \int \delta\{u \leq x\} \frac{D^{(1)}(\beta_0, u)}{\Phi_n(\beta_0, u)} \, d\mathbb{P}_n(u, \delta, z),
\]
\[
R_{n1}(x) = \int \delta\{u \leq x\} \frac{2D^{(1)}(\beta^*, u)D^{(1)}(\beta^*, u)' - D^{(2)}(\beta^*, u)}{\Phi_n(\beta^*, u)} \, d\mathbb{P}_n(u, \delta, z),
\]
for some $|\beta^* - \beta_0| \leq |\hat{\beta} - \beta_0|$, with $D_n^{(1)}$ as defined in (3.1) and

$$ D_n^{(2)}(\beta, x) = \frac{\partial^2 \Phi_n(\beta, x)}{\partial \beta^2} = \int \{ u \geq x \} \, z z' e^{z' \beta} \, dP_n(u, \delta, z) \in \mathbb{R}^p \times \mathbb{R}^p. $$

Note that, according to (A2), we have $|D^{(1)}(\beta_0, x)| \leq \mathbb{E} ||Z| \exp(\beta_0 Z)| < \infty$, for all $x \in \mathbb{R}$, and similarly

$$ |D_n^{(1)}(\beta_0, x)| \leq \frac{1}{n} \sum_{i=1}^n |Z_i| e^{\beta_0 Z_i} \rightarrow \mathbb{E} \left[ |Z| e^{\beta_0 Z} \right] < \infty, $$

with probability one. Likewise,

$$ |D_n^{(2)}(\beta_0, x)| \leq \frac{1}{n} \sum_{i=1}^n |Z_i|^2 e^{\beta_0 Z_i} \rightarrow \mathbb{E} \left[ |Z|^2 e^{\beta_0 Z} \right] < \infty, $$

with probability one. Furthermore, for all $x \in [0, M]$,

$$ 0 < \Phi(\beta_0, M) \leq \Phi(\beta_0, x) \leq \Phi(\beta_0, 0) = \mathbb{E} \left[ e^{\beta_0 Z} \right] < \infty $$

and $\Phi_n(\beta_0, M) \leq \Phi_n(\beta_0, x) \leq \Phi_n(\beta_0, 0)$, where $\Phi_n(\beta_0, M) \rightarrow \Phi(\beta_0, M)$ and $\Phi_n(\beta_0, 0) \rightarrow \Phi(\beta_0, 0)$, with probability one. It follows, that there exist constants $K_1, K_2 > 0$, such that for all $x \in [0, M]$,

$$ |D^{(1)}(\beta_0, x)| \leq K_2, \quad |D^{(2)}(\beta_0, x)| \leq K_2 \quad \text{and} \quad K_1 \leq \Phi(\beta_0, x) \leq K_2 \quad (3.5) $$

and for $n$ sufficiently large,

$$ |D_n^{(1)}(\beta_0, x)| \leq K_2, \quad |D_n^{(2)}(\beta_0, x)| \leq K_2 \quad \text{and} \quad K_1 \leq \Phi_n(\beta_0, x) \leq K_2, \quad (3.6) $$

with probability one. Note that, according to (2.3),

$$ \frac{\delta}{\Phi(\beta_0, u)} \, dP(u, \delta, y) = \frac{dH^{ue}(u)}{\Phi(\beta_0, u)} \lambda_0(u) \, du, \quad (3.7) $$

so that $A_0$, as defined in (3.3), is equal to

$$ A_0(x) = \int \delta\{ u \leq x \} \frac{D^{(1)}(\beta_0, u)}{\Phi(\beta_0, u)} \, dP(u, \delta, z) \in \mathbb{R}^p, $$

Then, for the $A_n$ term in (3.4), it can be deduced that

$$ \sup_{0 \leq x \leq M} |A_n(x) - A_0(x)| \leq \sup_{0 \leq u \leq M} \left| \frac{D_n^{(1)}(\beta_0, u)}{\Phi_n(\beta_0, u)} - \frac{D^{(1)}(\beta_0, u)}{\Phi(\beta_0, u)} \right| + \sup_{0 \leq x \leq M} \left| \int \delta\{ u \leq x \} \frac{D^{(1)}(\beta_0, u)}{\Phi(\beta_0, u)} \, d(P_n - P)(u, \delta, z) \right|, $$

By (3.5) and (3.6), the first term on the right hand side is bounded by

$$ \frac{1}{K_1^2} \sup_{0 \leq x \leq M} \left| D_n^{(1)}(\beta_0, x) - D^{(1)}(\beta_0, x) \right| + \frac{2K_2}{K_1^2} \sup_{0 \leq x \leq M} |\Phi_n(\beta_0, x) - \Phi(\beta_0, x)|, $$

and the second by

$$ \frac{4K_1}{K_1^2} \sup_{0 \leq x \leq M} |\Phi_n(\beta_0, x) - \Phi(\beta_0, x)|. $$
which is of the order $O_p(n^{-1/2})$, by Lemma 3.1. For the second term on the right hand side, for each $j = 1, 2, \ldots, p$ fixed, consider the class $G_j = \{g_j(u, \delta; x) : x \in [0, M]\}$, consisting of functions
\[ g_j(u, \delta; x) = \delta \{ u \leq x \} \frac{D_j^{(1)}(\beta_0, u)}{\Phi^2(\beta_0, u)}, \]
where $D_j^{(1)}$ denotes the $j$th coordinate of $D^{(1)}$. Now, each $g_j(u, \delta; x)$ is the product of indicators and a fixed uniformly bounded function. Standard results from empirical process theory (van der Vaart and Wellner, 1996) give that the class $G_j$ is Donsker. As in the proof of Lemma 3.1, we find that for every $j = 1, 2, \ldots, p$,
\[ \sqrt{n} \sup_{0 \leq x \leq M} \left| \int g_j(u, \delta; x) d(\mathbb{P}_n - P)(u, \delta, z) \right| = O_p(1). \]
It follows that
\[ \sup_{0 \leq x \leq M} |A_n(x) - A_0(x)| = O_p(n^{-1/2}). \]
and we can conclude that
\[ (\hat{\beta} - \beta_0)' A_n(x) = (\hat{\beta} - \beta_0)' A_0(x) + R_{n2}(x), \]
where $R_{n2}(x) = O_p(n^{-1})$, uniformly for $x \in [0, M]$, since it is well-known that $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$, see for example Tsiatis (1981). For the term containing $R_{n1}$, first observe that, according to (3.6), for $n$ sufficiently large,
\[ \sup_{u \in [0, M]} \left| \frac{2D_n^{(1)}(\beta^*, u)D_n^{(1)}(\beta^*, u)' - D_n^{(2)}(\beta^*, u)\Phi_n(\beta^*, u)}{\Phi_n^2(\beta^*, u)} \right| = O(1), \]
aalmost surely, so that
\[ \sup_{0 \leq x \leq M} \left| \frac{1}{2} (\hat{\beta} - \beta_0)' R_{n1}(x) (\hat{\beta} - \beta_0) \right| = O_p(n^{-1}). \]
Concluding,
\[ T_{n1}(x) = (\hat{\beta} - \beta_0)' A_0(x) + O_p(n^{-1}), \quad (3.8) \]
uniformly in $x \in [0, M]$. Proceeding with $T_{n2}$, write
\[ T_{n2}(x) = A_n(\beta_0, x) - A_0(x) = B_n(x) + C_n(x) + R_{n3}(x) + R_{n4}(x), \]
where
\[ B_n(x) = \int \delta \{ u \leq x \} \frac{\Phi(\beta_0, u) - \Phi_n(\beta_0, u)}{\Phi^2(\beta_0, u)} dP(u, \delta, z), \]
\[ C_n(x) = \int \delta \{ u \leq x \} d(\mathbb{P}_n - P)(u, \delta, z), \]
\[ R_{n3}(x) = \int \delta \{ u \leq x \} \left( \frac{1}{\Phi_n(\beta_0, u)} - \frac{1}{\Phi(\beta_0, u)} \right) d(\mathbb{P}_n - P)(u, \delta, z), \]
\[ R_{n4}(x) = \int \delta \{ u \leq x \} [\Phi(\beta_0, u) - \Phi_n(\beta_0, u)]^2 \frac{1}{\Phi^2(\beta_0, u)\Phi_n(\beta_0, u)} dP(u, \delta, z). \]
For the dominating term in $T_{n2}$, we can write

$$B_n(x) + C_n(x) = -\int \delta\{u \leq x\} \frac{\Phi_n(\beta_0, u)}{\Phi'(\beta_0, u)} \, dP(u, \delta, z) + \int \delta\{u \leq x\} \frac{dP_n(u, \delta, z)}{\Phi(\beta_0, u)}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \xi(T_i, \Delta_i, Z_i; x),$$

where

$$\xi(t, \delta, z; x) = -\int \gamma\{u \leq x\} \frac{\{t \geq u\} e^{\beta_0'z}}{\Phi(\beta_0, u)} \, dP(u, \gamma, y) + \delta\{t \leq x\} \frac{\Phi(\beta_0, t)}{\Phi(\beta_0, t)}.$$ 

Moreover, together with (3.7), we conclude that

$$\xi(t, \delta, z; x) = -e^{\beta_0'z} \int_0^{t \wedge M} \frac{\lambda_0(u)}{\Phi(\beta_0, u)} \, du + \delta\{t \leq x\} \frac{\Phi(\beta_0, t)}{\Phi(\beta_0, t)}.$$ 

For the remainder terms, it follows immediately, by Lemma 3.2, that for any sequence $a_n \to 0$,

$$\sup_{0 \leq x \leq M} |R_{n3}(x)| = O_p(n^{-1}a_n^{-1}). \quad (3.9)$$

To treat $R_{n4}$, note that

$$|R_{n4}(x)| \leq \frac{1}{\Phi'(\beta_0, M)} \frac{1}{\Phi_n(\beta_0, M)} \sup_{x \in \mathbb{R}} |\Phi_n(\beta_0, x) - \Phi(\beta_0, x)|^2,$$

so that by (3.2) and (3.6),

$$\sup_{0 \leq x \leq M} |R_{n4}(x)| = O_p(n^{-1}).$$

Together with (3.8) and (3.9), this proves the theorem. \qed

In the special case of no covariates, i.e., $\beta_0 = \hat{\beta} = 0$, it follows that

$$\Phi(\beta_0, x) = 1 - H(x)$$

and

$$\xi(t, \delta, z; x) = -e^{\beta_0'z} \int_0^{t \wedge M} \frac{\lambda_0(u)}{\Phi(\beta_0, u)} \, du + \delta\{t \leq x\} \frac{\Phi(\beta_0, t)}{\Phi(\beta_0, t)}$$

$$= -\int_0^{t \wedge M} \frac{dH^{uc}(u)}{[1 - H(u)]^2} + \frac{\delta\{t \leq x\}}{1 - H(t)}.$$ 

This means that Theorem 3.1 retrieves a result similar to that in Lemma 2.1 in Lo et al. (1989).

The rate at which the error term $R_n$ tends to zero becomes faster as $a_n$ tends to zero more slowly. If $a_n = 1/\log n$, one obtains the same rate as the error term in Lemma 2.1 in Lo et al. (1989). However, they obtain order $O(n^{-1} \log n)$ almost surely, whereas Theorem 3.1, with the choice $a_n = 1/\log n$, only provides this order in probability. On the other hand, the sequence $(a_n)$ may be chosen to converge to zero arbitrarily slow. This means that the order at which the error term $R_n$ tends to zero (in probability) is arbitrarily close to $O_p(n^{-1})$. 

9
References


