Results on standard estimators in the Cox model

Cécile Durota, Eni Mustab,∗

aUniversité Paris Nanterre
bDelft University of Technology

Abstract

We consider the Cox regression model and prove some properties of the maximum partial likelihood estimator $\hat{\beta}_n$ and the empirical estimator $\Phi_n$. The asymptotic properties of these estimators have been widely studied in the literature but we are not aware of a reference where it is shown that they have uniformly bounded moments. These results are needed, for example, when studying global errors of shape restricted estimators of the baseline hazard function.

Keywords: Cox regression model, maximum partial likelihood estimator, uniformly bounded moments

1. Introduction

We consider the Cox proportional hazards model, which is commonly used to investigate the relationship between the survival times and the predictor variables in the presence of right censoring. Let $X$ be the event time and $C$ the censoring time for a subject with covariate vector $Z$. We terminate the study at time $T_0$ and collect $n$ i.i.d observations $(T_1, \Delta_1, Z_1), \ldots, (T_n, \Delta_n, Z_n)$, where $T_i = \min(X_i, C_i)$ is the follow up time, $T_i \leq T_0$, and $\Delta_i = 1\{X_i \leq C_i\}$ is censoring indicator.

The Cox regression model assumes that the hazard function at time $t$ for a subject with covariate vector $z \in \mathbb{R}^d$ has the form

$$\lambda(t|z) = \lambda_0(t) e^{\beta_0 z}, \quad t \in \mathbb{R}^+,$$

where $\lambda_0$ represents the baseline hazard function, corresponding to a subject with $z = 0$, and $\beta_0 \in \mathbb{R}^p$ is the vector of the regression coefficients.

The following assumptions are common when studying asymptotics in the Cox regression model (see for example [Tsiatis (1981), Lopuhaä and Nane (2013b)]). The variable $Z$ has density $f_Z(z)$. Given the covariate vector $Z$, the event time $X$ and the censoring time $C$ are assumed to be independent. Furthermore, conditionally on $Z = z$, the event time is a nonnegative r.v. with an
absolutely continuous distribution function $F(x|z)$ and density $f(x|z)$. Similarly the censoring time is a nonnegative r.v. with an absolutely continuous distribution function $G(x|z)$ and density $g(x|z)$.

The censoring mechanism is assumed to be non-informative, i.e. $F$ and $G$ share no parameters.

We will also need the following assumptions:

(A1) the end points $\tau_F$ and $\tau_G$ of the support of $F$ and $G$ satisfy

$$T_0 \leq \tau_G < \tau_F \leq \infty,$$

(A2) there exists $\epsilon > 0$ such that

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

(A3) for all $q \geq 1$, we have

$$\mathbb{E} \left[ e^{q\beta_0'Z} \right] < \infty,$$

(A4) for all $q \geq 1$ and $k = 1, \ldots, d$, we have

$$\mathbb{E} \left[ Z_k^{2q} e^{q\beta_0'Z} \right] < \infty.$$

Here $|.|$ denotes the Euclidean norm, $\beta'$ denotes the transpose of $\beta$ and $Z_k$ is the $k^{th}$ component of the vector $Z$. We will use the index $k = 1, \ldots, d$ when it corresponds to a component of a vector and indices $i, j = 1, \ldots, n$ when it corresponds to the different observations. The first assumption tells us that, at the end of the study, there is at least one subject alive while (A2) can be seen as conditions on the boundedness of the second moment of the covariates, for $\beta$ in a neighbourhood of $\beta_0$. The other two assumptions are additional ones needed for our analysis in order to get all moments of $\hat{\beta}_n$ and $\Phi_n$ bounded.

The proportional hazard property of the Cox model allows estimation of the effects $\beta_0$ of the covariates by the maximum partial likelihood estimator $\hat{\beta}_n$, while leaving the baseline hazard completely unspecified. $\hat{\beta}_n$ is defined as the maximizer of the partial likelihood function

$$L(\beta) = \prod_{i=1}^{m} \frac{e^{\beta'Z_i}}{\sum_{j=1}^{n} \mathbb{I}(T_j \geq X(i)) e^{\beta'Z_j}},$$

where $0 < X(1) < \cdots < X(m) < \infty$ denote the ordered, observed event times (see Cox (1972) and Cox (1975)). Note that, since we are considering observations on $[0, T_0]$, also $X(m) \leq T_0$. Moreover, the estimator $\hat{\beta}_n$ depends on $T_0$. Asymptotic properties of this estimator have been investigated, among other papers, in Tsiatis (1981), Andersen and Gill (1982). In particular, they show that

$$n^{1/2} (\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Sigma)$$
for some positive definite matrix $\Sigma$. The restriction on $[0, T_0]$ is common in asymptotic studies of the Cox model. If $T_0 < \tau_G$, it guarantees that $\Phi(t; \beta_0)$ is bounded away from zero on $[0, T_0]$, which is a condition assumed in [Andersen and Gill (1982), Tsiatis (1981) and Kalbfleisch and Prentice (2002)]. Here we need it in order to use their results and prove boundedness of the moments of $\hat{\beta}_n$ under the condition that $T_0 < \tau_G$ whereas the study of the empirical estimator $\Phi_n$ as defined below requires only that $T_0 \leq \tau_G$.

On the other hand, the nonparametric cumulative baseline hazard

$$\Lambda_0(t) = \int_0^t \lambda_0(u) \, du,$$

is usually estimated by the Breslow estimator

$$\Lambda_n(t) = \int \frac{\delta 1_{(u \leq t)}}{\Phi_n(u; \hat{\beta}_n)} \, dP_n(u, \delta, z). \tag{1}$$

where

$$\Phi_n(t; \beta) = \int 1_{(u \geq t)} e^{\beta^t z} \, dP_n(u, \delta, z), \tag{2}$$

and $P_n$ is the empirical measure of the triplets $(T_i, \Delta_i, Z_i)$ with $i = 1, \ldots, n$. $\Phi_n$ is an estimator of

$$\Phi(t; \beta) = \int 1_{(u \geq t)} e^{\beta^t z} \, dP(u, \delta, z), \tag{3}$$

where $P$ is the common distribution of the triplets $(T_i, \Delta_i, Z_i)$ and, in Lemma 4 of Lopuhaä and Nane (2013b) it is shown that

$$\sup_{t \in \mathbb{R}} |\Phi_n(t; \beta_0) - \Phi(t; \beta_0)| = O_p(n^{-1/2}). \tag{4}$$

In the next section, we show that $n^{1/2} |\hat{\beta}_n - \beta_0|$ and $n^{1/2} \sup_t |\Phi_n(t, \beta_0) - \Phi(t, \beta_0)|$ have uniformly bounded moments of any order. Such results are needed, for example, when studying global errors of the Grenander-type estimator of a monotone baseline hazard (see Appendix D in Durot and Musta (2018)).

2. Main results

**Theorem 1.** Suppose that (A3) holds and $T_0 \leq \tau_G$. Let $p \geq 1$. Then, there exists $K > 0$ such that

$$\sup_{n \geq 1} \mathbb{E} \left[ n^{p/2} \sup_{t \in \mathbb{R}} |\Phi_n(t, \beta_0) - \Phi(t, \beta_0)|^p \right] \leq K.$$

**Proof.** By definition we have

$$n^{1/2} \sup_{t \in \mathbb{R}} |\Phi_n(t, \beta_0) - \Phi(t, \beta_0)| = \sup_{t \in \mathbb{R}} \left| \int 1_{(u \geq t)} e^{\beta_0^t z} \, d\sqrt{n}(P_n - P)(u, \delta, z) \right|.$$

Let $\mathcal{F}$ be the class of functions

$$f_t(u, z) = 1_{(u \geq t)} e^{\beta_0^t z}, \quad t \in \mathbb{R},$$

for some positive definite matrix $\Sigma$. The restriction on $[0, T_0]$ is common in asymptotic studies of the Cox model. If $T_0 < \tau_G$, it guarantees that $\Phi(t; \beta_0)$ is bounded away from zero on $[0, T_0]$, which is a condition assumed in [Andersen and Gill (1982), Tsiatis (1981) and Kalbfleisch and Prentice (2002)]. Here we need it in order to use their results and prove boundedness of the moments of $\hat{\beta}_n$ under the condition that $T_0 < \tau_G$ whereas the study of the empirical estimator $\Phi_n$ as defined below requires only that $T_0 \leq \tau_G$.

On the other hand, the nonparametric cumulative baseline hazard

$$\Lambda_0(t) = \int_0^t \lambda_0(u) \, du,$$

is usually estimated by the Breslow estimator

$$\Lambda_n(t) = \int \frac{\delta 1_{(u \leq t)}}{\Phi_n(u; \hat{\beta}_n)} \, dP_n(u, \delta, z). \tag{1}$$

where

$$\Phi_n(t; \beta) = \int 1_{(u \geq t)} e^{\beta^t z} \, dP_n(u, \delta, z), \tag{2}$$

and $P_n$ is the empirical measure of the triplets $(T_i, \Delta_i, Z_i)$ with $i = 1, \ldots, n$. $\Phi_n$ is an estimator of

$$\Phi(t; \beta) = \int 1_{(u \geq t)} e^{\beta^t z} \, dP(u, \delta, z), \tag{3}$$

where $P$ is the common distribution of the triplets $(T_i, \Delta_i, Z_i)$ and, in Lemma 4 of Lopuhaä and Nane (2013b) it is shown that

$$\sup_{t \in \mathbb{R}} |\Phi_n(t; \beta_0) - \Phi(t; \beta_0)| = O_p(n^{-1/2}). \tag{4}$$

In the next section, we show that $n^{1/2} |\hat{\beta}_n - \beta_0|$ and $n^{1/2} \sup_t |\Phi_n(t, \beta_0) - \Phi(t, \beta_0)|$ have uniformly bounded moments of any order. Such results are needed, for example, when studying global errors of the Grenander-type estimator of a monotone baseline hazard (see Appendix D in Durot and Musta (2018)).

2. Main results

**Theorem 1.** Suppose that (A3) holds and $T_0 \leq \tau_G$. Let $p \geq 1$. Then, there exists $K > 0$ such that

$$\sup_{n \geq 1} \mathbb{E} \left[ n^{p/2} \sup_{t \in \mathbb{R}} |\Phi_n(t, \beta_0) - \Phi(t, \beta_0)|^p \right] \leq K.$$

**Proof.** By definition we have

$$n^{1/2} \sup_{t \in \mathbb{R}} |\Phi_n(t, \beta_0) - \Phi(t, \beta_0)| = \sup_{t \in \mathbb{R}} \left| \int 1_{(u \geq t)} e^{\beta_0^t z} \, d\sqrt{n}(P_n - P)(u, \delta, z) \right|.$$

Let $\mathcal{F}$ be the class of functions

$$f_t(u, z) = 1_{(u \geq t)} e^{\beta_0^t z}, \quad t \in \mathbb{R},$$
with envelope function $F(u, z) = e^{\beta_0 z}$. Then, we can write
\[
\mathbb{E} \left[ n^{p/2} \sup_{t \in \mathbb{R}} |\Phi_n(t, \beta_0) - \Phi(t, \beta_0)|^p \right] \leq \mathbb{E} \left[ \sup_{f \in \mathcal{F}} \left| \int f(u, z) d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z) \right|^p \right]
\]
From Theorem 2.14.1 in van der Vaart and Wellner (1996), it follows that
\[
\mathbb{E} \left[ \sup_{f \in \mathcal{F}} \left| \int f(u, z) d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z) \right|^p \right] \lesssim J(1, \mathcal{F}) \|F\|_{L_{2v}(\mathbb{P})},
\]
where
\[
J(1, \mathcal{F}) = \sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon \|F\|_{L_2(\mathcal{Q})}, \mathcal{F}, L_2(Q))} \, d\epsilon
\]
and the supremum is taken over all probability measures $Q$ such that $\|F\|_{L_2(\mathcal{Q})} > 0$. By Assumption (A3), $\|F\|_{L_{2v}(\mathbb{P})} < \infty$. Hence, it remains to show that $J(1, \mathcal{F})$ is bounded.

Let $Q$ be a probability measure on $\mathbb{R} \times \mathbb{R}^p$ such that $\|F\|_{L_2(\mathcal{Q})} > 0$. Let $Q'$ be the probability measure on $\mathbb{R}$ defined by
\[
Q'(S) = \frac{\int_{\mathbb{R} \times \mathbb{R}^p} e^{2\beta_0 z} dQ(u, z)}{\int_{\mathbb{R} \times \mathbb{R}^p} e^{2\beta_0 z} dQ(u, z)} = \frac{\int_{\mathbb{R} \times \mathbb{R}^p} e^{2\beta_0 z} dQ(u, z)}{\|F\|_{L_2(\mathcal{Q})}^2}, \quad S \subseteq \mathbb{R}.
\]
For a given $\epsilon > 0$ select an $\epsilon$-net $g_1, \ldots, g_N$ in the class $\mathcal{S}$ of monotone functions $\mathbb{R} \to [0, 1]$ with respect to $L_2(\mathcal{Q}')$. From Theorem 2.7.5 and in van der Vaart and Wellner (1996) and the relation between covering and bracketing numbers in page 84 of van der Vaart and Wellner (1996), we have $N(\epsilon \|F\|_{L_2(\mathcal{Q})}, \epsilon) \lesssim 1/\epsilon$ and the constant in the inequality $\lesssim$ does not depend on $Q'$. Next, we consider functions $f_i(u, z) = g_i(u)e^{\beta_0 z}$. Then $f_1, \ldots, f_N$ form an $\epsilon \|F\|_{L_2(\mathcal{Q})}$-net of the class $\mathcal{F}$ with respect to $L_2(\mathcal{Q})$.

Indeed, for each $t \in \mathbb{R}$, let $i$ be such that $g(u) = 1_{\{u \geq t\}}$ belongs in the $\epsilon$-ball around $g_i$. Then
\[
\|f_i - f_i\|_{L_2(\mathcal{Q})}^2 = \int_{\mathbb{R} \times \mathbb{R}^p} (1_{\{u \geq t\}} - g_i(u))^2 e^{2\beta_0 z} dQ(u, z)
\]
\[
= \|F\|_{L_2(\mathcal{Q})}^2 \int_{\mathbb{R}} (1_{\{u \geq t\}} - g_i(u))^2 dQ'(u)
\]
\[
= \|F\|_{L_2(\mathcal{Q})}^2 \|g - g_i\|_{L_2(\mathcal{Q}')}^2
\]
\[
\leq \epsilon^2 \|F\|_{L_2(\mathcal{Q})}^2.
\]
Therefore
\[
N(\epsilon \|F\|_{L_2(\mathcal{Q})}, \mathcal{F}, L_2(Q)) \leq \frac{K}{\epsilon}
\]
for some constant $K > 0$ independent of $Q$. It follows that $J(1, \mathcal{F})$ is bounded, which concludes the proof. \qed

**Theorem 2.** Assume that (A1), (A2) and (A4) hold , and that $T_0 < \tau_G$. Let $p \geq 1$. There exist an event $E_n$ that depends on $T_0$ with $\mathbb{P}(E_n) \to 1$, and $K > 0$ such that
\[
\limsup_{n \to \infty} \mathbb{E} \left[ 1_{E_n} n^{p/2} |\hat{\beta}_n - \beta_0|^p \right] \leq K.
\]
Then, we have
\[ S(\beta) = \log L(\beta) = \sum_{i=1}^{m} \beta' Z(i) - \sum_{i=1}^{m} \log \left( \sum_{j=1}^{n} I(T_j \geq X(i)) e^{\beta' Z_j} \right), \]
where \( X(1), \ldots, X(m) \) are the ordered observed event times. From Theorem 3.1 in Tsiatis (1981), \( \hat{\beta}_n \) is the solution of \( S'(\beta) = 0 \), where \( S' \) denotes the vector \( \left( \frac{\partial S(\beta)}{\partial \beta_1}, \ldots, \frac{\partial S(\beta)}{\partial \beta_p} \right) \). Note that, in Tsiatis (1981), it is written that \( \hat{\beta}_n \) is the solution to the equation (3.2) but actually it is a zero of the expression in (3.2). By a Taylor expansion we have
\[ S'(\hat{\beta}_n) = S'(\beta_0) - S''(\beta^*) \left( \hat{\beta}_n - \beta_0 \right) = 0, \]
where \( |\beta^* - \beta_0| \leq |\hat{\beta}_n - \beta_0| \) and the positive semi-definite matrix \( S'' \) is minus the matrix of the second derivatives \( S''_{ij}(\beta) = -\frac{\partial^2 S(\beta)}{\partial \beta_i \partial \beta_j} \). We also know that \( \frac{1}{n} S''(\beta^*) \) converges in probability to a non-singular matrix \( \Sigma \), see the second step of the proof of Theorem 3.2 in Andersen and Gill (1982). There \( S'' \) is denoted by \( J \). In this proof conditions A, B, D of Andersen and Gill (1982) are used. In our setting A is satisfied because we are assuming a continuous hazard rate. For B note that their \( S^{(0)}, S^{(1)} \) and \( S^{(2)} \) are our \( \Phi_n, D_1^n \) and \( D_2^n \) which converge uniformly to \( \Phi, D^1 \) and \( D^2 \) (See Lemma 1 in Lopuha¨ a and Nane (2013a) for the first two; in the same way one can also deal with \( D_2^n \)). The boundedness of \( D^1 \) and \( D^2 \) follows from our assumptions (A2) and (A4). They also consider observations in a compact interval away of the right boundary, for example on \([0, 1]\) such that \( 1 < \tau_C \), in order to have \( \inf_{t \in [0, 1]} \Phi(t) > 0 \). Here we consider observations on \([0, T_0]\) where \( T_0 < \tau_C \), so \( T_0 \) plays the role of 1. Hence
\[ \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) = \Sigma^{-1} n^{-1/2} S'(\beta_0) - \left( \Sigma^{-1} \frac{1}{n} S''(\beta^*) - I \right) \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right). \]
It follows that
\[ \sqrt{n} \left| \hat{\beta}_n - \beta_0 \right| \leq \left| \Sigma^{-1} n^{-1/2} S'(\beta_0) \right| + \left| \left( n^{-1} \Sigma^{-1} S''(\beta^*) - I \right) \sqrt{n} \left( \hat{\beta}_n - \beta_0 \right) \right| \]
\[ \leq \left\| \Sigma^{-1} \right\| \left| n^{-1/2} S'(\beta_0) \right| + \left| n^{-1} \Sigma^{-1} S''(\beta^*) - I \right\| \sqrt{n} \left| \hat{\beta}_n - \beta_0 \right| \]
where \( \cdot \) is the euclidian norm in \( \mathbb{R}^p \) and \( \cdot \) is the matrix norm induced by the euclidian vector norm, i.e.
\[ \| A \| = \sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{|Ax|}{|x|} = \sigma_{\max}(A) \leq \left( \sum_{i,j=1}^{p} A_{ij}^2 \right)^{1/2}, \quad A \in \mathbb{R}^{d \times d} \]
and \( \sigma_{\max}(A) \) is the largest singular value of \( A \). Let \( \epsilon < 1 \). Since \( n^{-1} S''(\beta^*) \to \Sigma \) in probability, we can take the event
\[ E_n = \left\{ \left| n^{-1} \Sigma^{-1} S''(\beta^*) - I \right\| \leq \epsilon \right\}. \]
Then, we have \( \mathbb{P}(E_n) \to 1 \) and
\[ \mathbb{I}_{E_n} \sqrt{n} \left| \hat{\beta}_n - \beta_0 \right| \leq \frac{1}{1 - \epsilon} \left\| \Sigma^{-1} \right\| \left| n^{-1/2} S'(\beta_0) \right|. \]
It suffices to show that $\mathbb{E} \left[ |n^{-1/2} S'(\beta_0)|^p \right]$ is uniformly bounded.

By definition we have

$$S'(\beta_0) = \sum_{i=1}^{n} \Delta_i Z_i - \sum_{i=1}^{n} \Delta_i \frac{D^1_n(T_i; \beta_0)}{\Phi_n(T_i; \beta_0)}$$

where

$$D^1_n(t; \beta) = \frac{\partial \Phi_n(t; \beta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(T_i \geq t)} Z_i e^{\beta Z_i}.$$  

We will follow the martingale approach of Kalbfleisch and Prentice (2002). For each $i = 1, \ldots, n$, let $N_i(t) = \Delta_i \mathbb{1}_{(T_i \leq t)}$ be the right-continuous counting process for the number of observed failures in $(0, t]$ and $Y_i(t) = \mathbb{1}_{(T_i \geq t)}$ be the at-risk process. From (5.49) in Kalbfleisch and Prentice (2002), the compensator of $N_i(t)$ is

$$A_i(t) = \int_0^t Y_i(u) \lambda_0(u) e^{\beta Z_i(u)} \, du,$$

and $M_i(t) = N_i(t) - A_i(t)$ is a mean zero martingale with respect to the filtration

$$\mathcal{F}_t = \{N_i(s), Y_i(s+), Z_1 : i = 1, \ldots, n, s \in [0, t]\}$$

(see Kalbfleisch and Prentice (2002), page 173). The score function $S'$ up to a certain time $t$ can be then written

$$S'(\beta_0, t) = \sum_{i=1}^{n} \int_0^t \left[ Z_i - \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \right] \, dN_i(u)$$

$$= \sum_{i=1}^{n} \int_0^t \left[ Z_i - \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \right] \, dM_i(u)$$

(see equations (5.50) and (5.51) in Kalbfleisch and Prentice (2002)). Note that we can replace $dN_i$ by $dM_i$ because

$$\sum_{i=1}^{n} \int_0^t \left[ Z_i - \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \right] \, dA_i(u)$$

$$= \sum_{i=1}^{n} \int_0^t \left[ Z_i - \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \right] Y_i(u) \lambda_0(u) e^{\beta Z_i(u)} \, du$$

$$= \int_0^t \lambda_0(u) \sum_{i=1}^{n} Z_i Y_i(u) e^{\beta Z_i(u)} \, du - \int_0^t \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \lambda_0(u) \sum_{i=1}^{n} Y_i(u) e^{\beta Z_i(u)} \, du$$

$$= n \int_0^t \lambda_0(u) \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \, du - n \int_0^t \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \lambda_0(u) \, du = 0.$$  

Being a sum of stochastic integrals of predictable processes with respect to a martingale, $S'(\beta_0, \cdot)$ is also an $\mathcal{F}_t$-martingale. Let

$$G_{i,n}(u) = \left[ Z_i - \frac{D^1_n(u; \beta_0)}{\Phi_n(u; \beta_0)} \right].$$

Then

$$n^{-1/2} S'(\beta_0, t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t G_{i,n}(u) \, dM_i(u)$$
It follows from properties of stochastic integrals that
\[
\langle n^{-1/2} S'(\beta_0) \rangle_t = \int_0^t \frac{1}{n} \sum_{i=1}^n \left\{ G_{i,n}(u)G^*_{i,n}(u)Y_i(u)e^{\beta_0 Z_i} \right\} \lambda_0(u) \, du
\]
(see proof of (5.58) in page 176 of [Kalbfleisch and Prentice 2002]). We have
\[
E \left[ n^{-1/2} S'(\beta_0, t) \right]^p \]
\[
= E \left[ n^{-p/2} \left( \sum_{i=1}^n \int_0^t G'_{i,n}(u) \, dM_i(u) \sum_{i=1}^n \int_0^t G_{i,n}(u) \, dM_i(u) \right)^{p/2} \right]
\]
\[
= E \left[ n^{-p/2} \left( \sum_{k=1}^d \left( \sum_{i=1}^n \int_0^t (G_{i,n}(u))_k \, dM_i(u) \right)^2 \right)^{p/2} \right]
\]
\[
\leq \sum_{k=1}^d E \left[ n^{-p/2} \left( \sum_{i=1}^n \int_0^t (G_{i,n}(u))_k \, dM_i(u) \right)^p \right],
\]
where again \( G'_{i,n}(u) \) denotes the transpose of the vector \( G_{i,n}(u) \) and \( (G_{i,n}(u))_k \) denotes its \( k \)th component. For the first and the second equalities we have used the definition of the euclidian norm of a vector in \( \mathbb{R}^d \), while for the last inequality we use that for positive numbers \( a_1, \ldots, a_d \) and all \( p \) we have \( (a_1 + \cdots + a_d)^p \leq d^p (a_1^p + \cdots + a_d^p) \). Each component
\[
\sum_{i=1}^n \int_0^t (G_{i,n}(u))_k \, dM_i(u)
\]
is a martingale with quadratic variation
\[
\left\langle \sum_{i=1}^n \int_0^t (G_{i,n}(u))_k \, dM_i(u) \right\rangle = \sum_{i=1}^n \int_0^t (G_{i,n}(u))_k^2 Y_i(u)e^{\beta_0 Z_i} \lambda_0(u) \, du.
\]
It follows from properties of stochastic integrals that
\[
E \left[ n^{-1/2} S'(\beta_0, t) \right]^p \]
\[
\leq \sum_{k=1}^d E \left[ n^{-p/2} \left( \sum_{i=1}^n \int_0^t (G_{i,n}(u))_k \, dM_i(u) \right)^p \right],
\]
Note that
\[
\frac{1}{n} \sum_{i=1}^n \int_0^t (G_{i,n}(u))_k^2 Y_i(u)e^{\beta_0 Z_i} \lambda_0(u) \, du
\]
\[
\leq \int_0^t \frac{1}{n} \sum_{i=1}^n (Z_i)_k^2 Y_i(u)e^{\beta_0 Z_i} \, du + \int_0^t \frac{(D_1^2(u; \beta_0))_k^2}{\Phi_n(u; \beta_0)} \frac{1}{n} \sum_{i=1}^n Y_i(u)e^{\beta_0 Z_i} \, du
\]
\[
\leq \int_0^t (D^2_n(u; \beta_0))_{kk} \, du + \int_0^t \frac{(D_1^2(u; \beta_0))_k^2}{\Phi_n(u; \beta_0)} \, du
\]
\[
\leq \sup_{u \in [0, t]} (D^2_n(u; \beta_0))_{kk} + \sup_{u \in [0, t]} \frac{(D_1^2(u; \beta_0))_k^2}{\Phi_n(u; \beta_0)},
\]
where
\[ D_n^2(u; \beta) = \frac{\partial^2 \Phi_n(u; \beta)}{\partial \beta^2} = \frac{1}{n} \sum_{i=1}^{n} Z_i Z_i u_i e^{Z_i}. \]

Hence, in order to have \( E \left[ |n^{-1/2} S'(\beta_0)|^p \right] \) uniformly bounded, it suffices to show that, for all \( p \geq 1 \),
\[ E \left[ \sup_{u \in [0, T(n)]} (D_n^2(u; \beta_0))_{kk} \right] \text{ and } E \left[ \sup_{u \in [0, T(n)]} \frac{(D_n^2(u; \beta_0))_{kk}^{2p}}{\Phi_n(u; \beta_0)^p} \right] \]
are uniformly bounded, where \( T(n) \) is the largest of the observations \( T_1, \ldots, T_n \). Note that \( S'(\beta_0) \)

By definition, we have
\[ \sup_{u \in [0, T(n)]} (D_n^2(u; \beta_0))_{kk} \leq \frac{1}{n} \sum_{i=1}^{n} (Z_i)^2 e^{\beta_0 Z_i}, \]

Also \( 1/\Phi_n \) is well defined up to \( T(n) \) and, from Titu’s lemma,
\[ \sup_{u \in [0, T(n)]} \frac{(D_n^2(u; \beta_0))_{kk}^2}{\Phi_n(u; \beta_0)^2} \leq \frac{1}{n} \sum_{i=1}^{n} (Z_i)^2 e^{\beta_0 Z_i}, \]

Hence, in order to show that the expectations in (5) are bounded, it suffices to show that
\[ E \left[ \left( \frac{1}{n} \sum_{i=1}^{n} (Z_i)^2 e^{\beta_0 Z_i} \right)^p \right] \]
is bounded. Let \( J = \{ a = (a_1, \ldots, a_n) \in \mathbb{Z}^n, a_i \geq 0 \text{ for all } i = 1, \ldots, n, \sum_{i=1}^{n} a_i = p \} \). Then, using linearity of the expectation, independence of the \( Z_i \)'s, it follows that
\[ \frac{1}{n^p} \sum_{a \in J} \left( a_1, \ldots, a_n \right) E \left[ \prod_{i=1}^{n} (Z_i)^{2a_i} e^{a_i \beta_0 Z_i} \right] \]
where \( (a_1, \ldots, a_n) \) are the multinomial coefficients. Using iteratively that, for a positive random variable \( Y \) and \( a, b \geq 0 \), we have \( E[Y^{a+b}] - E[Y^a]E[Y^b] = Cov(Y^a, Y^b) \geq 0 \), we obtain
\[ \prod_{i=1}^{n} E \left[ Z_k^{2a_i} e^{a_i \beta_0 Z_i} \right] \leq E \left[ Z_k^{2} e^{\beta_0 Z} \right] \leq E \left[ Z_k^{2p} e^{p \beta_0 Z} \right]. \]

Therefore, since \( \sum_{a \in J} (a_1, \ldots, a_n) = n^p \), we have
\[ \frac{1}{n^p} \sum_{a \in J} \left( a_1, \ldots, a_n \right) E \left[ Z_k^{2p} e^{p \beta_0 Z} \right] = E \left[ \frac{1}{n^p} \sum_{a \in J} \left( a_1, \ldots, a_n \right) e^{p \beta_0 Z} \right]. \]

By assumption (A4) it follows that \( E \left[ \left( \frac{1}{n} \sum_{i=1}^{n} (Z_i)^2 e^{\beta_0 Z_i} \right)^p \right] \), and as a result also \( E \left[ n^{-1/2} S'(\beta_0)^p \right] \),
are uniformly bounded. This concludes the proof of the theorem. \( \square \)
References

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

Cox, D.R., 1972. Regression models and life-tables. J. Roy. Statist. Soc. Ser. B 34, 187–220.

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

Durot, C., Musta, E., 2018. Supplement to: On the $l_p$ error of the grenander-type estimator in the cox model.

Kalbfleisch, J.D., Prentice, R.L., 2002. The statistical analysis of failure time data. Wiley Series in Probability and Statistics. second ed., Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Lopuhaä, H.P., Nane, G.F., 2013a. An asymptotic linear representation for the breslow estimator. Comm. Statist. Theory Methods 42, 1314–1324.

Lopuhaä, H.P., Nane, G.F., 2013b. Shape constrained non-parametric estimators of the baseline distribution in Cox proportional hazards model. Scand. J. Stat. 40, 619–646.

Tsiatis, A.A., 1981. A large sample study of Cox’s regression model. Ann. Statist. 9, 93–108.

van der Vaart, A.W., Wellner, J.A., 1996. Weak convergence and empirical processes. Springer Series in Statistics, Springer-Verlag, New York. With applications to statistics.