Published in final edited form as:
Stat Probab Lett. 2012 July ; 82(7): 1267–1272. doi:10.1016/j.spl.2012.03.023.

An Exponential Bound for Cox Regression☆

Y. Goldberg[a] and **M. R. Kosorok**[b]
Y. Goldberg: ygoldberg@stat.haifa.ac.il; M. R. Kosorok: kosorok@unc.edu
[a]Department of Statistics, University of Haifa, Israel, 31905
[b]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Abstract

We present an asymptotic exponential bound for the deviation of the survival function estimator of the Cox model. We show that the bound holds even when the proportional hazards assumption does not hold.

Keywords

Cox model; model misspecification; exponential bound

1. Introduction

The Cox proportional hazards model is often used to describe survival data in the presence of covariates. Under the Cox model, the hazard of an individual is given as the multiplication of a baseline hazard with the effect of the individual's covariates Z. The effect of the individual's covariates is given by the exponent of the regression of the covariates on some vector of coefficients $\boldsymbol{\beta}_0$. The survival function of each individual is then obtained from the hazard using the product integral of $-\exp(\beta_0' Z)\Lambda_0$ where Λ_0 is the cumulative hazard.

Standard results (Fleming and Harrington, 1991) show that when the proportional hazards assumption holds, the maximizer of the partial likelihood for the regression vector $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$, and that the Breslow estimator $\hat{\Lambda}$, which maximizes the profile likelihood at $\hat{\boldsymbol{\beta}}$, is consistent for the baseline cumulative hazard. Combining these two result, one can show that the estimated survival function converges to the underlying true survival function for each individual. Moreover, it can be shown that for every individual, the difference between the estimated survival function $\hat{G}$, and the true survival function G_0, in the root-n scale, converges to a mean zero tight Gaussian process.

Even when the proportional hazards assumption does not hold, it was shown by Lin and Wei (1989) and Sasieni (1993) that the maximizer of the partial likelihood for the regressor vector $\hat{\boldsymbol{\beta}}$ converges to some vector $\boldsymbol{\beta}^*$. Moveover, $\sqrt{n}(\widehat{\beta} - \beta^*)$ converges to some mean zero normal random vector. However, the convergence of the survival function, under the misspecified model was not discussed.

☆The authors are grateful to the anonymous reviewer for helpful suggestions and comments. The first author thanks Jason Fine for encouragement to write this paper. The second author was funded in part by grant CA142538 from the National Cancer Institute.

Correspondence to: Y. Goldberg, `ygoldberg@stat.haifa.ac.il`.

In this work we show that even when the proportional hazards assumption is violated, the survival function converges to some well defined limit G_P. Here, we regard the survival function as a function of both time and covariates. Moreover, the difference between the estimated and the limiting survival function in the root-n scale converges to a mean zero tight Gaussian process. Finally, we provide an exponential bound for the tail difference between the estimated survival function $\hat{G}$ and G_P. More formally, Theorem 3.2 below states that under some regularity conditions, for every $\varepsilon > 0$, and all n large enough,

$$P(\sup_{z\in\mathscr{Z},t\in[0,\tau]} \sqrt{n}|\widehat{G}(t|z) - G_p(t|z)|>\varepsilon)<2\exp\{-D_1\varepsilon^2+D_2\varepsilon\}, \quad (1)$$

where D_1 and D_2 are universal constants that depend only on the set of covariates $\mathscr{Z}$, and where $\tau > 0$ is some constant (see Section 2 below). Note that this bound does not depend on the proportional hazards assumption.

Maximal inequality bounds, such as the bound (1), play an important role in machine-learning proofs of universal consistency (Steinwart and Chirstmann, 2008, Chapter 6). Here universal consistency means that asymptotically, an estimator achieves the minimal risk with respect to some loss function, regardless of the probability measure that generates that data. However, when the data is subject to censoring, the usual maximal inequality bounds do not apply and bounds like (1) are needed in order to derive universal consistency. For example, Goldberg and Kosorok (2012a) use a maximal inequality bound to prove universal consistency for multistage decision problems with censored observations. More recently, Goldberg and Kosorok (2012b) proposed support vector regression estimators for right censored data. In that work, when the censoring distribution is assumed to follow the Cox model, a maximal inequality like the one proved here is used to bound the deviation of the estimator from the Bayes function.

Exponential bounds for the tails of distribution functions are well known. The Dvoretzky, Kiefer and Wolfowitz bound states that

$$P(\sqrt{n}\sup_{t\in\mathbb{R}}|\mathbb{F}_n(t) - F(t)|>\varepsilon) \le 2\exp\{-2\varepsilon^2\},$$

where $\mathbb{F}_n$ and F are the empirical and true distribution functions, respectively (Kosorok, 2008, Theorem 11.6). In the context of survival analysis, the literature is limited. Bitouzé et al. (1999) proved a non-asymptotic exponential bound for the deviation of the Kaplan-Meier estimator from the survival function for right censored data, when no covariates are present. We note that the bound (1) differs from these two bounds, since it is asymptotic in nature. The difficulty of obtaining a non-asymptotic bound follows from two main reasons. First, the finite-sample difference $\sqrt{n}(\widehat{\beta} - \beta^*)$ involves the inverse of some empirical covariance matrix (Lin and Wei, 1989). Second, the difference between the survival functions depends on the distribution of covariates Z. Finding such a non-asymptotic exponential bound for the Cox model is an important challenge which is beyond the scope of this paper.

The paper is organized as follows. The notation and some standard results for the Cox model are presented in Section 2. The main results are presented in Section 3. Proofs are provided in Section 4.

2. Preliminaries

Assume that the observed data consist of n independent and identically distributed random triplets $\{(Z_1, U_1, \delta_1), \ldots, (Z_n, U_n, \delta_n)\}$ drawn from some probability distribution P. The random vector Z is a covariate vector that takes its values in $\mathcal{Z} \subset \mathbb{R}^d$. The observed time U equals $T \wedge C$, where $T \geq 0$ is a failure time, C is a right-censoring time, and where $a \wedge b = \min(a, b)$. Let $\delta = 1_{\{T \leq C\}}$ be the failure indicator where $1_{\{A\}}$ is 1 if A is true and is 0 otherwise, i.e., $\delta = 1$ whenever a failure time is observed. Let $G(\cdot|Z) = P(T > t|Z)$ denote the survival functions of T given Z.

Let $\tau > 0$ be such that $P(U \geq \tau) > 0$. For every $t \in [0, \tau]$, define $N(t) = 1_{\{U \leq t; \delta=1\}}$ and $Y(t) = 1_{\{U \geq t\}}$. For a cadlag function A on $[0, \tau]$, define the product integral $\varphi(A)(t) = \prod_{0<s\leq t}(1 + dA(s)) \equiv \exp(A^c(t)) \prod_{0<s\leq t}(1 + \Delta A(s))$ where $\Delta A(t) = A(t) - A(t-)$ and $A^c(t) = A(t) - \Sigma_{0<s\leq t} \Delta A(s)$ (see, for example, Kosorok, 2008, Chapter 12). Furthermore, the product integral can be generalized, in a natural way to functions from $[0, \tau] \times \mathcal{Z}$ to $\mathbb{R}$, that are cadlag in the first variable. Denote this space of functions with the supremum norm $D[0, \tau] \times \mathcal{Z}$ (Billingsley, 1999, Chapter 3). For an integrable function f, denote its expectation as $Pf \equiv \int f dP$. Define $\mathbb{P}_n$ to be the empirical measure, i.e., $\mathbb{P}_n f(X) = n^{-1} \sum_{i=1}^{n} f(X_i)$.

The proportional hazards assumption states that the integrated hazard of $T|Z$ is of the form $e^{\beta_0' Z} \Lambda_0$ for some unknown vector $\beta_0 \in \mathbb{R}^d$ and some continuous unknown nondecreasing function Λ_0 with $\Lambda_0(0) = 0$ and $0 < \Lambda(\tau) < \infty$. Estimation of β can be done by maximizing the partial likelihood. Define the estimating equation

$$U_n(\beta) = \mathbb{P}_n \int_0^{\tau} \left(Z - \frac{\mathbb{P}_n Z Y(s) e^{\beta' Z}}{\mathbb{P}_n Y(s) e^{\beta' Z}} \right) dN(s).$$

Under the proportional hazards assumption, the zero of this estimating equation, $\hat{\beta}$, is consistent for β_0, and $\sqrt{n}(\hat{\beta} - \beta_0)$ converges to a Gaussian random variable (Fleming and Harrington, 1991, Chapter 8). Define $\hat{\Lambda}$ to be an estimator for the cumulative hazard function, where

$$\hat{\Lambda}(t) = \int_0^t \frac{\mathbb{P}_n dN(s)}{\mathbb{P}_n Y(s) e^{\hat{\beta}' Z}}. \quad (2)$$

It can be shown (Fleming and Harrington, 1991, Chapter 8), that $\hat{\Lambda}$ is consistent for Λ_0, and $\sqrt{n}(\hat{\Lambda} - \Lambda_0)$ converges to a mean zero tight Gaussian process. The estimator for the survival function can be obtained using the product integral operator, i.e., $\hat{G}(t, z) = \varphi(-e^{\hat{\beta}' z} \hat{\Lambda}(t))$. By the delta method (Kosorok, 2008, Theorem 2.8), it can be shown that $\hat{G}$ is consistent for G and that $\sqrt{n}(\hat{G} - G)$ weakly converges in $D[0, \tau] \times \mathcal{Z}$ to a tight Gaussian process.

In the discussion above, it was assumed that the hazard model assumption holds. However, even when the model is misspecified, it was shown by Lin and Wei (1989) and Sasieni (1993) that under some regularity conditions, $\sqrt{n}(\hat{\beta} - \beta^*)$ is asymptotically normal, where β^* is the zero of the estimating equation

$$U(\beta)=P\int_0^{\tau}\left(Z-\frac{PZY(s)e^{\beta' Z}}{PY(s)e^{\beta' Z}}\right)dN(s). \quad (3)$$

Estimation for the cumulative hazard can be obtained as in (2).

3. Main Results

In the following, we first discuss the limit G_P of the survival hazard estimator $\hat{G}$ defined above. We note that $G_P \equiv G$, the true survival function, only when the model is correctly specified. Then we present an exponential bound on the deviation of $\hat{G}$ from its limit.

We need the following assumptions, adapted from Sasieni (1993):

(C1) $\mathcal{Z}$ is a compact set.

(C2) There is no pair (α, φ) with $\alpha \neq 0 \in \mathbb{R}^d$ and $\varphi: \mathbb{R} \mapsto \mathbb{R}$, a monotone decreasing function, such that for $P^{(2)}$-almost all $t < \tau$, $\alpha' Z dN(t) = \varphi(t) dN(t)$ P-almost surely, and $\alpha' Z Y(t) \leq \varphi(t)$ P-almost surely, where $P^{(2)}$ is the marginal subdistribution of U with $\delta = 1$.

(C3) $P(Y(\tau) = 1) \geq K > 0$.

The assumption that $\mathcal{Z}$ is compact can be relaxed to an assumption on the moments of Z at the price of complicating the proofs (compare to Assumption (C1) in Sasieni, 1993). Assumptions (C2)–(C3) ensure that β^*, the zero of the estimating equation (3), exists and is unique (Sasieni, 1993, Corollary 3.1).

Define $\Lambda^*(t)=\int_0^t PdN(s)/P(Y(s)e^{\beta^{*\prime} Z})$. Define $G_P(t|z) = \varphi(-e^{\beta^{*\prime} z}\Lambda^*)(t)$. In the following, we show that G_P is the limit of $\hat{G}$:

Lemma 3.1

Assume (C1)–(C3). Then $\widehat{G}(t|z) \xrightarrow{\text{a.s.}} G_P(t|z)$. Moreover, $\sqrt{n}(\widehat{G}-G_P)$ converges to a mean zero tight Gaussian process on $D[0, \tau] \times \mathcal{Z}$.

It can be shown, under some conditions, that the limiting survival function G_P is the minimizer of the Kullback-Leibler divergence between the functions within the Cox family and the true distribution (see Kosorok et al., 2004, Theorem 6 and its conditions).

Standard results for the supremum of a mean zero tight Gaussian processes ensure that $\limsup P(\sqrt{n}\|(\widehat{G}-G_P)\|_\infty > \varepsilon) < 2\exp\{-C\varepsilon^2\}$. However, the constant in the exponent depends on the distribution P. The main result stated below shows that under some regularity conditions, an exponential bound can be given for which the constants are universal and do not depend on the distribution P. Before we state this result, we need to strengthen our assumptions:

(C4) $\alpha' Var(Z)\, \alpha \geq K_1 > 0$ for all $\alpha \in \mathbb{R}^d$ such that $\|\alpha\|_2 = 1$.

(C5) $\inf_{z \in Z} P(Y(\tau) = 1|Z) \geq K_2 > 0$.

(C6) $\beta^* \in B_R$ where B_R is the open ball of radius R around the origin.

Theorem 3.2

Assume (C1) and (C4)–(C6). Then for all $\varepsilon > 0$, and all n large enough,

$$P(\sqrt{n}\left\|\widehat{G}-G_P\right\|_\infty>\varepsilon)\le 2\exp\{-D_1\varepsilon^2+D_2\varepsilon\},$$

where D_1, D_2 are universal constants that depend on the set $\mathcal{Z}$, the radius R, and the constants K_1 and K_2, but otherwise do not depend on the distribution P nor on ε.

4. Proofs

Proof of Lemma 3.1

By Theorem 2.1 of Lin and Wei (1989) and Corollary 4.1 of Sasieni (1993), $\sqrt{n}(\widehat{\beta}-\beta^*)\rightsquigarrow V_1$ where $V_1\sim N(\mathbf{0},\Sigma)$ for some positive definite matrix Σ. In the following we use arguments similar to Kosorok (2008, Chapter 4.2) but without the assumption that the proportional hazards model holds. Write

$$\begin{aligned}
&e^{\widehat{\beta}'Z}\widehat{\Lambda}(t)-e^{\beta^{*\prime}Z}\Lambda^*(t)\\
&=e^{\widehat{\beta}'Z}\left(\int_0^t 1_{\{\mathbb{P}_nY(s)>0\}}\left\{\frac{(\mathbb{P}_n-P)dN(s)}{\mathbb{P}_nY(s)e^{\widehat{\beta}'Z}}\right\}-\int_0^t 1_{\{\mathbb{P}_nY(s)=0\}}\left\{\frac{PdN(s)}{\mathbb{P}_nY(s)e^{\widehat{\beta}'Z}}\right\}\right.\\
&\quad-\int_0^t\frac{(\mathbb{P}_n-P)Y(s)e^{\widehat{\beta}'Z}}{PY(s)e^{\widehat{\beta}'Z}}\left\{\frac{PdN(s)}{\mathbb{P}_nY(s)e^{\widehat{\beta}'Z}}\right\}\\
&\quad\left.-\int_0^t\frac{PY(s)\left(e^{\widehat{\beta}'Z}-e^{\beta^{*\prime}Z}\right)}{PY(s)e^{\widehat{\beta}'Z}}\frac{PdN(s)}{PY(s)e^{\beta^{*\prime}Z}}\right)-\left(e^{\widehat{\beta}'Z}-e^{\beta^{*\prime}Z}\right)\int_0^t\frac{PdN(s)}{PY(s)e^{\beta^{*\prime}Z}}\\
&\equiv A_n(t)-B_n(t)-C_n(t)-D_n(t)-E_n(t).
\end{aligned}\tag{4}$$

It follows from Assumption (C3) and Lemma 7.4 of Sasieni (1993) that $B_n(t)$ converges to zero uniformly in t. Since both N and Y are Donsker (see Kosorok, 2008, Lemma 4.1), $\widehat{\beta}\xrightarrow{\text{a.s.}}\beta^*$, $\mathcal{Z}$ is compact, and the processes A_n, C_n, D_n, and E_n are smooth in $\boldsymbol{\beta}$, we conclude that each of these processes converges uniformly to zero. Since the product integral is continuous, we obtain that $\left\|\widehat{G}-G_P\right\|_\infty\xrightarrow{\text{a.s.}}0$.

We now move to show that $\sqrt{n}(e^{\widehat{\beta}'z}\widehat{\Lambda}(t)-e^{\widehat{\beta}'z}\Lambda^*(t))\to V(t,z)$ where V is a tight, mean zero Gaussian process on $D[0,\tau]\times\mathcal{Z}$. It follows from (4) that

$$\begin{aligned}
&\sqrt{n}(e^{\widehat{\beta}'z}\widehat{\Lambda}(t)-e^{\beta^{*\prime}z}\Lambda^*(t))\\
&=e^{\beta^{*\prime}z}\sqrt{n}(\mathbb{P}_n-P)\int_0^t\left(\frac{dN(s)}{PY(s)e^{\beta^{*\prime}Z}}-\frac{Y(s)e^{\beta^{*\prime}z}PdN(s)}{\{P(Y(s)e^{\beta^{*\prime}Z})\}^2}\right)\\
&\quad-\sqrt{n}(\widehat{\beta}-\beta^*)e^{\beta^{*\prime}z}\int_0^t\left(\frac{PZY(s)e^{\beta^{*\prime}Z}}{PY(s)e^{\beta^{*\prime}Z}}+z\right)d\Lambda^*(s)+o_P(1)\\
&\equiv\tilde{A}_n+\tilde{B}_n+o_P(1),
\end{aligned}\tag{5}$$

where the reminder term is uniform in t and z. The joint asymptotic tightness of $(\tilde{A}_n,\tilde{B}_n)$ follows, since the marginals are asymptotically tight (Kosorok, 2008, Lemma 7.14). The joint convergence is established using the Cramer-Wold device (Kosorok, 2008, Theorem 7.17). Using the continuous mapping theorem for the sum, we conclude that the limiting process V is a tight, mean zero Gaussian process in $D[0,\tau]\times\mathcal{Z}$. Finally, it follows from the delta method that $\sqrt{n}(\widehat{G}-G_P)\equiv\sqrt{n}(\varphi(-e^{\widehat{\beta}'z}\widehat{\Lambda})-\varphi(-e^{\beta^{*\prime}z}\Lambda^*))$ converges to a mean zero, tight Gaussian process.

Proof of Theorem 3.2

By Assumption (C5) and the fact that $\mathcal{Z}$ is compact, we obtain that there is a universal constant M_0, that depends only on K_2 and $\mathcal{Z}$, such that $\Lambda^*(\tau) < M_0$. For all n large enough, and by the consistency of $\hat{\Lambda}$, we thus have $\hat{\Lambda}(\tau) < 2M_0$. Using Lemma 12.6 of Kosorok (2008), on the event $\Omega = \{\hat{\Lambda}(\tau) < 2M_0, \|\hat{\beta}\|_2 < 2R\}$, for every $z \in \mathcal{Z}$, we have

$$\begin{aligned} \left\|\widehat{G}(\cdot|z) - G_P(\cdot|z)\right\|_\infty &\le e^{3M_0}(1+3M_0)^2\left\|e^{\widehat{\beta}'z}\widehat{\Lambda} - e^{\beta^{*\prime}z}\Lambda^*\right\|_\infty \\ &\le e^{3M_0}(1+3M_0)^2\left(\left\|\widehat{\Lambda}\right\|_\infty |e^{\widehat{\beta}'z} - e^{\beta^{*\prime}z}| + |e^{\widehat{\beta}'z}|\left\|\widehat{\Lambda} - \Lambda^*\right\|_\infty\right) \\ &\le e^{3M_0}(1+3M_0)^2\left(2M_0M_1e^{2M_1R}\left\|\widehat{\beta} - \beta^*\right\|_2 + e^{2M_1R}\left\|\widehat{\Lambda} - \Lambda^*\right\|_\infty\right). \end{aligned} \quad (6)$$

where $M_1 = \max_{z\in\mathcal{Z}}\{\|z\|_2\}$. Note that using similar arguments as in (5) on the event Ω,

$$\left\|\widehat{\Lambda}(t) - \Lambda^*(t)\right\|_\infty \le \left\|(\mathbb{P}_n - P)\int_0^t\left(\frac{dN(s)}{PY(s)e^{\beta^{*\prime}Z}} - \frac{Y(s)e^{\beta^{*\prime}z}PdN(s)}{\{P(Y(s)e^{\beta^{*\prime}Z})\}^2}\right)\right\|_\infty + \left\|(\widehat{\beta} - \beta^*)\right\|_2\left\|\int_0^t\frac{PZY(s)e^{\beta^{*\prime}Z}d\Lambda^*(s)}{PY(s)e^{\beta^{*\prime}Z}}\right\|_\infty. \quad (7)$$

By the compactness of both $\mathcal{Z}$ and the closure of B_R, and Assumption (C5), we obtain that there is a constant K_3 that does not depend on the distribution P, such that $PY(s)e^{\beta^{*\prime}Z} > K_3$. Hence

$$\left\|\int_0^t\left(\frac{PZY(s)e^{\beta^{*\prime}Z}}{PY(s)e^{\beta^{*\prime}Z}}\right)d\Lambda^*(s)\right\|_\infty \le M_0K_3^{-1}M_1e^{M_1R}. \quad (8)$$

The influence function representation of $\sqrt{n}(\widehat{\beta} - \beta^*)$ in Theorem 4.1 of Sasieni (1993), together with (6)–(8), yield that

$$\sqrt{n}\left\|\widehat{G} - G_P\right\|_\infty \le \left\|\mathbb{G}_n\tilde{l}_1(Z, U, \delta)\right\|_\infty + \left\|\mathbb{G}_n\tilde{l}_2(Z, U, \delta)(t)\right\|_\infty + \mathrm{Rem}_n, \quad (9)$$

where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$, Rem_n is $o_P(1)$,

$$\begin{aligned} \tilde{l}_1(Z, U, \delta) &= C_1A(\beta^*)^{-1}\left\{\int_0^\tau\left(Z - \frac{PY(s)Ze^{\beta^{*\prime}Z}}{PY(s)e^{\beta^{*\prime}Z}}\right)dN(s) + \int_0^\tau\frac{Y(s)e^{\beta^{*\prime}Z}}{PY(s)e^{\beta^{*\prime}Z}}\left(Z - \frac{PY(s)Ze^{\beta^{*\prime}Z}}{PY(s)e^{\beta^{*\prime}Z}}\right)PdN(s)\right\} \\ &\equiv C_1A(\beta^*)^{-1}\tilde{l}_3, \\ \tilde{l}_2(Z, U, \delta)(t) &= C_2\int_0^t\left(\frac{dN(s)}{PY(s)e^{\beta^{*\prime}Z}} - \frac{Y(s)e^{\beta^{*\prime}Z}PdN(s)}{\{P(Y(s)e^{\beta^{*\prime}Z})\}^2}\right), \end{aligned}$$

and

$$\begin{aligned} C_1 &= 2M_0M_1(1+3M_0)^2e^{3M_0+2M_1R} + M_0K_3^{-1}M_1e^{M_1R}, \\ C_2 &= (1+3M_0)^2e^{3M_0+2M_1R}, \\ A(\beta) &= \int_0^\tau\left\{\frac{PY(s)ZZ'e^{\beta'Z}}{PY(s)e^{\beta'Z}} - \frac{PY(s)Ze^{\beta'Z}PY(s)Z'e^{\beta'Z}}{\{PY(s)e^{\beta'Z}\}^2}\right\}PdN(s). \end{aligned}$$

The inverse of $A(\boldsymbol{\beta}^*)$ exists by Lemma 7.3 of Sasieni (1993). Note that the constants C_1 and C_2 do not depend on the distribution P. We now show that for every $\alpha \in \mathbb{R}^d$, such that $\|\alpha\|_2 = 1$, $\alpha' A(\boldsymbol{\beta}^*)\alpha > C_4 K_1$, where $C_4 > 0$ does not depend on the distribution P. Write

$$A(\beta^*)=\frac{1}{2}\int_0^\tau\int_0^\tau P\left[\frac{(Z_1-Z_2)(Z_1-Z_2)' Y_1(s_1)Y_2(s_2)e^{\beta^{*\prime}Z_1}e^{\beta^{*\prime}Z_2}}{PY_1(s_1)Y_2(s_2)e^{\beta^{*\prime}Z_1}e^{\beta^{*\prime}Z_2}}\right]PdN_1(s_1)PdN_2(s_2),$$

where (Z_1, Y_1, N_1) and (Z_2, Y_2, N_2) are independent copies, distributed according to P. Note that the denominator is bounded from above. Since $\mathcal{Z}$ is compact and $\boldsymbol{\beta}^*$ is bounded, $e^{\boldsymbol{\beta}^{*\prime} Z_1}e^{\boldsymbol{\beta}^{*\prime} Z_2}$ in the nominator is bounded from below by some constant. Hence, for every α, such that $\|\alpha\|_2 = 1$,

$$\begin{aligned}\alpha' A(\beta^*)\alpha &\geq \tfrac{C_3}{2}\alpha'\left\{\int_0^\tau\int_0^\tau P[(Z_1-Z_2)(Z_1-Z_2)' Y_1(s_1)Y_2(s_2)]PdN_1(s_2)PdN_2(s_2)\right\}\alpha\\ &\geq \tfrac{C_3}{2}\alpha'\left\{P[(Z_1-Z_2)(Z_1-Z_2)' Y_1(\tau)Y_2(\tau)]\int_0^\tau\int_0^\tau PdN_1(s_1)PdN_2(s_2)\right\}\alpha,\end{aligned}$$

where the last inequality follows from the monotonicity of Y and $C_3 > 0$ is some constant depending on K_1, K_2, R and M_1. By Assumption (C5), both $\int_0^\tau PdN(s_1)$ and $\inf_{z\in\mathcal{Z}} P(Y(\tau) = 1|Z = z)$ are bounded from below. Thus we conclude that there is a constant $C_4 > 0$ such that

$$\alpha' A(\beta^*)\alpha \geq \frac{C_4}{2}\alpha'\{P(Z_1-Z_2)(Z_1-Z_2)'\}\alpha = C_4\alpha' Var(Z)\alpha > K_1 C_4.$$

Since $A(\boldsymbol{\beta}^*)$ is symmetric and positive definite, we conclude that $\|A(\boldsymbol{\beta}^*)^{-1}\|_2 \equiv \max_{\{\|\alpha\|_2=1\}} \alpha' A(\boldsymbol{\beta}^*)^{-1}\alpha < 1/(K_1C_4)$ (Golub and Loan, 1983).

We would like to investigate the empirical processes defined by the functions $\tilde{l}_1$ and $\tilde{l}_2$. First, note that $\tilde{l}_1$ is a random vector such that each of its component is bounded by some constat C_5 that depends only on K_1, K_2, R and M_1. To see this, recall that $\tilde{l}_1 \equiv C_1 A(\boldsymbol{\beta}^*)^{-1}\tilde{l}_3$. Since both $\mathcal{Z}$ and $\boldsymbol{\beta}^*$ are bounded and $PY(s)Ze^{\boldsymbol{\beta}^{*\prime}Z} > K_3$, each of the components of $\tilde{l}_3$ is bounded in absolute value by a constant, say C_6. Since $\|A^{-1}\|_2 < (K_1C_4)^{-1}$, $\|A^{-1}\|_\infty < (K_1C_4)^{-1}d^{1/2}$ (where $\|A\|_\infty$ is the maximum absolute row sum of the matrix, see Golub and Loan, 1983), and thus

$$\|\tilde{l}_1\|_\infty \leq \|A^{-1}\|_\infty\|\tilde{l}_3\|_\infty \leq (K_1C_4)^{-1}d^{1/2}C_6 \equiv C_5.$$

Let e_i be the i-th member of the standard basis of $\mathbb{R}^d$. Then

$$P\left(\|\mathbb{G}_n\tilde{l}_1\|_\infty > \frac{\varepsilon}{3}\right) \leq \sum_{i=1}^d P\left(|\mathbb{G}_n e_i'\tilde{l}_1| > \frac{\varepsilon}{3d}\right) \leq 2d\exp\left\{\frac{-2\varepsilon^2}{(dC_5)^2}\right\}, \quad (10)$$

for any $\varepsilon > 0$, where the last inequality follows from Hoeffding's inequality.

Define the ε-bracketing entropy to be the log of the ε-bracketing covering number $N_{[]}(\varepsilon, \mathcal{F}, L_2(P))$ (see Kosorok, 2008, Chapter 2). Define the class of sample pathes $\mathcal{F}_2 = \{\tilde{l}_2(Z, U, \delta)(t): t \in [0, \tau]\}$. We would like to show that for every $0 < \varepsilon$,

$$\log N_{[\,]}(\varepsilon, \mathscr{F}_2, L_2(P)) < C_o \varepsilon^{-1}. \quad (11)$$

The sample pathes of the random functions $\int_0^t \{PY(s)\exp(\beta^{*\prime} Z)\}^{-1} dN(s)$ and $\int_0^t \{PY(s)\exp(\beta^{*\prime} Z)\}^{-2} Y(s) e^{\beta^{*\prime} Z} P dN(s)$ are monotone in t from $[0, \tau]$ to $[0, C_7]$ for some constant C_7 that does not depend on P. Hence, by Theorem 9.23 of Kosorok (2008), the ε-bracketing entropy number of these classes is bounded by K_o/ε for some universal constant K_o that depends only on C_7. By Lemma 9.25 of Kosorok (2008), the ε-bracketing entropy of the sum of these two classes is bounded by $4K_o/\varepsilon$, where the sum of two classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$ is defined as $\mathcal{G}_1 + \mathcal{G}_2 \equiv \{g_1 + g_2: g_i \in G_i\}$. By changing variables we obtain that (11) holds with the constant $C_o = 4K_o C_2$. Fix an $\varepsilon > 0$, then using the bound (11), and Corollary 2 of Bitouzé et al. (1999), we obtain that

$$P(\|\mathbb{G}_n \tilde{l}_1(Z, U, \delta)\|_\infty > \varepsilon/3) < 2.5\exp\{-2(\varepsilon/3C_o)^2 + B\varepsilon\}, \quad (12)$$

where B is some universal constant.

We are now ready to bound $\sqrt{n}\|\widehat{G} - G_P\|_\infty$. Fix $\varepsilon > 0$. It then follows from (9), (10), and (12), that

$$\begin{aligned} P(\sqrt{n}\|\widehat{G} - G_P\|_\infty > \varepsilon) &\le P\left(\|\mathbb{G}_n \tilde{l}_1(Z, U, \delta)\|_\infty > \tfrac{\varepsilon}{3}\right) + P\left(\|\mathbb{G}_n \tilde{l}_2(Z, U, \delta)\|_\infty > \tfrac{\varepsilon}{3}\right) + P\left(\mathrm{Rem}_n > \tfrac{\varepsilon}{3}\right) \\ &\le 2.5\exp\left\{-2\frac{\varepsilon^2}{(3C_o)^2} + B\varepsilon\right\} + 2d\exp\left\{\frac{-2\varepsilon^2}{(dC_5)^2}\right\} + P\left(\mathrm{Rem}_n > \tfrac{\varepsilon}{3}\right) \\ &\le 2\exp\{-D_1\varepsilon^2 + D_2\varepsilon\} + P(\mathrm{Rem}_n > \tfrac{\varepsilon}{3}), \end{aligned}$$

where $D_1 = 2\min\{\log(2.5)(3C_o)^{-2}, \log(2d)(dC_5)^{-2}\}$ and $D_2 = \log(2.5)B$.

References

Billingsley, P. Convergence of probability measures. John Wiley & Sons, Ltd; 1999.

Bitouzé D, Laurent B, Massart P. A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. Ann Inst H Poincaré Probab Statist. 1999; 35(6):735–763.

Fleming, TR.; Harrington, DP. Counting processes and survival analysis. John Wiley & Sons, Ltd; 1991.

Goldberg Y, Kosorok MR. Q-learning with censored data, the Annals of Statistics. 2012a (to appear).

Goldberg, Y.; Kosorok, MR. Support vector regression for right censored data. Unpublished maniscript. 2012b. avaliable at http://arxiv.org/abs/1202.5130

Golub, GH.; Loan, CFV. Matrix Computations. Johns Hopkins University Press; 1983.

Kosorok, MR. Introduction to Empirical Processes and Semiparametric Inference. Springer; 2008.

Kosorok MR, Lee BL, Fine JP. Robust inference for univariate proportional hazards frailty regression models. The Annals of Statistics. 2004; 32 (4):1448–1491.

Lin DY, Wei LJ. The robust inference for the cox proportional hazards model. Journal of the American Statistical Association. 1989; 84 (408):1074–1078.

Sasieni P. Some new estimators for cox regression. The Annals of Statistics. 1993; 21 (4):1721–1759.

Steinwart, I.; Chirstmann, A. Support Vector Machines. Springer; 2008.

NIH-PA Author Manuscript

NIH-PA Author Manuscript