Modern Stochastics: Theory and Applications 1 (2014) 13–32
DOI: 10.15559/vmsta-2014.1.1.3

Asymptotic normality of corrected estimator in Cox proportional hazards model with measurement error

C. Chimisov*, **A. Kukush**

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

kirill.chimisov@gmail.com (C. Chimisov), alexander_kukush@univ.kiev.ua (A. Kukush)

Received: 16 March 2014, Revised: 29 May 2014, Accepted: 5 June 2014,
Published online: 27 June 2014

Abstract Cox proportional hazards model is considered. In Kukush et al. (2011), *Journal of Statistical Research*, Vol. **45**, No. 2, 77–94 simultaneous estimators $\lambda_n(\cdot)$ and β_n of baseline hazard rate $\lambda(\cdot)$ and regression parameter β are studied. The estimators maximize the objective function that corrects the log-likelihood function for measurement errors and censoring. Parameter sets for $\lambda(\cdot)$ and β are convex compact sets in $C[0,\tau]$ and $\mathbb{R}^k$, respectively. In present paper the asymptotic normality for β_n and linear functionals of $\lambda_n(\cdot)$ is shown. The results are valid as well for a model without measurement errors. A way to compute the estimators is discussed based on the fact that $\lambda_n(\cdot)$ is a linear spline.

Keywords Asymptotic normality of estimators, classical measurement error, Corrected Maximum Likelihood Estimator, Cox proportional hazards model, estimator of baseline hazards function
2010 MSC 62N02, 62N01, 62J12

1 Introduction

We deal with Cox proportional hazards model where a lifetime $T \geq 0$ has the following intensity function

$$\lambda(t|X;\lambda,\beta) = \lambda(t)\exp(\beta^{\mathsf{T}}X), \quad t \geq 0. \tag{1.1}$$

*Corresponding author.

Here we say that positive random variable ξ has *intensity function* $\tilde{\lambda}(\cdot)$ if

$$\tilde{\lambda}(t) = \lim_{h \to 0_+} h^{-1} \mathbf{P}\{t \le \xi < t+h | \, \xi \ge t\}, \quad t \ge 0.$$

In (1.1) covariate X is a random vector distributed in $\mathbb{R}^k$, $\lambda(\cdot) \in \Theta_\lambda \subset C[0,\tau]$ is the baseline hazard function and β is a parameter from $\Theta_\beta \subset \mathbb{R}^k$. We observe only censored value $Y := \min\{T, C\}$, where censor C is distributed in $[0,\tau]$. Survival function of C, $G_C(u) = 1 - F_C(u)$, is unknown but we know τ. Censorship indicator $\Delta := \mathbb{I}_{\{T \le C\}}$ is observed as well. X is not observed directly, instead a surrogate data $W = X + U$ is observed, where U has known and finite moment generating function $M_U(\beta) := \mathbf{E} e^{\beta^\top U}$. Here $\mathbf{E}$ stands for expectation. A couple (T, X), censor C and measurement error U are stochastically independent. We mention that recently measurement error models become quite popular, e.g., in [9] an autoregressive model with measurement error was studied.

Consider independent copies of the model $(X_i, T_i, C_i, Y_i, \Delta_i)$, $i = 1, \dots, n$. Based on (Y_i, Δ_i, W_i), $i = 1, \dots, n$, we estimate true values of β and $\lambda(\cdot)$ that we denote by β_0 and $\lambda_0(\cdot)$, respectively. The latter is estimated on $[0,\tau]$ only.

There are a lot of papers on estimation of β_0 and cumulative hazard $\Lambda(t) = \int_0^t \lambda(t)\,\mathrm{d}t$. In [1] general ideas are presented based on partial likelihood. Same model but with measurement errors is considered in [4], where, based on Corrected Score method, consistent and asymptotically normal estimators are constructed for regression parameter and cumulative hazard function. Another approach is proposed in [6] where doubly censored data are considered without measurement error. Here cumulative hazard is estimated, and strong consistency and asymptotic normality of maximum likelihood estimators are proven. However, sometimes it is necessary to know the behaviour of baseline hazard function $\lambda(\cdot)$ itself, not cumulative hazard (see [10]). Our model is presented in [2] and [5] where baseline hazard function is assumed to belong to a parametric space while we consider $\lambda(\cdot)$ from a compact set of $C[0,\tau]$.

If values of X_i were measured without measurement error, we could use Maximum Likelihood Estimator (MLE) which maximizes the log-likelihood function

$$\tilde{Q}_n(\lambda, \beta) := \frac{1}{n} \sum_{i=1}^n q(Y_i, \Delta_i, X_i; \lambda, \beta),$$

where

$$\tilde{q}(Y, \Delta, X; \lambda, \beta) = \Delta\big(\log \lambda(Y) + \beta^\top X\big) - e^{\beta^\top X} \int_0^Y \lambda(u)\,\mathrm{d}u.$$

Since X_i is contaminated, we have to correct our objective function for measurement error. Due to suggestion of Augustin [2] we construct a new objective function q such that

$$\mathbf{E}\big[q(Y_i, \Delta_i, W_i; \lambda, \beta) | \, Y_i, \Delta_i, X_i\big] = \tilde{q}(Y_i, \Delta_i, X_i; \lambda, \beta) \quad \text{a.s.}$$

Then the corrected log-likelihood function is

$$Q_n(\lambda, \beta) := \frac{1}{n} \sum_{i=1}^n q(Y_i, \Delta_i, W_i; \lambda, \beta), \tag{1.2}$$

where

$$q(Y, \Delta, W; \lambda, \beta) = \Delta(\log \lambda(Y) + \beta^\top W) - \frac{e^{\beta^\top W}}{M_U(\beta)} \int_0^Y \lambda(u)\, \mathrm{d}u. \tag{1.3}$$

As an estimator of true parameters (λ_0, β_0), we use a couple (λ_n, β_n) which maximizes (1.2).

Introduce further assumptions.

(i) $\Theta_\lambda = \{f : [0, \tau] \to \mathbb{R} | f(t) \geq a,\ \forall t \in [0, \tau],\ f(0) \leq A,\ |f(t) - f(s)| \leq L|t - s|,\ \forall t, s \in [0, \tau]\}$, where $a > 0$, $A > a$ and $L > 0$ are fixed constants.

(ii) Θ_β is a compact and convex set in $\mathbb{R}^k$.

(iii) $\mathbf{E}U = 0$ and for some $\varepsilon > 0$,

$$\mathbf{E}\big[e^{2D\|U\|}\big] < \infty \quad \text{where } D := \max_{\beta \in \Theta_\beta} \|\beta\| + \varepsilon.$$

(iv) $\mathbf{E}[e^{2D\|X\|}] < \infty$ where $D > 0$ is defined in (iii).

(v) τ is right endpoint of the distribution of C, i.e., $\mathbf{P}\{C > \tau\} = 0$ and for all $\varepsilon > 0$, $\mathbf{P}\{C > \tau - \varepsilon\} > 0$.

(vi) The covariance matrix of random vector X is positive definite.

(vii) β_0 is an interior point of Θ_β.

(viii) $\lambda_0 \in \Theta_\lambda^\varepsilon$ for some $\varepsilon > 0$, where $\Theta_\lambda^\varepsilon := \{f : [0, \tau] \to \mathbb{R} | f(t) \geq a + \varepsilon,\ \forall t \in [0, \tau],\ f(0) \leq A - \varepsilon,\ |f(t) - f(s)| \leq (L - \varepsilon)|t - s|,\ \forall t, s \in [0, \tau]\}$.

(ix) $\mathbf{P}\{C > 0\} = 1$.

Remark. Assumptions (i) to (ix) allow us to consider model without measurement error. One just has to set $U_i = 0$ and $M_U(\beta) = 1$. All results of the article are valid for this case as well.

In [7] the strong consistency of (λ_n, β_n) is proven and the rate of convergence is presented. Our goal is to provide asymptotic normality for β_n and λ_n. The paper is organised as follows. Section 2 states the main results on the asymptotic normality. Section 3 suggests the procedure for computation of the estimates. Section 4 proves the stochastic boundedness results. Section 5 proves auxiliary results, Section 6 gives the proof of the main result, and Section 7 concludes.

For a sequence of random variables $\{x_n\}$, notation $x_n = O_p(1)$ means that $\{x_n\}$ is *stochastically bounded*. We assume that censor C has pdf f_C (this is a technical assumption that can be easily avoided). According to [7], Section 3, conditional density of (Y, Δ) given X at point (λ_0, β_0) equals

$$f(y, \delta|X) = f_T^\delta(y|X) G_T^{1-\delta}(y|X) f_C^{1-\delta} G_C^\delta(y), \tag{1.4}$$

where f_T is conditional pdf of T given X and G_T is conditional survival function:

$$f_T(t|X) = \Lambda(t|X; \lambda_0, \beta_0) \exp\Big(-\int_0^t \lambda(s|X; \lambda_0, \beta_0)\, \mathrm{d}s\Big),$$
$$G_T(t|X) = \exp\Big(-\int_0^t \lambda(s|X; \lambda_0, \beta_0)\, \mathrm{d}s\Big).$$

Let Z be a normed linear space. For a function $f: Z \to \mathbb{R}$ we denote $f^{(n)}(x_0)$ its n-th Fréchet derivative at a point $x_0 \in Z$. $f^{(n)}(x_0)$ is n-linear form and for $h_1, \dots, h_n \in Z$ we denote $\langle f^{(n)}(x_0), (h_1, \dots, h_n)\rangle$ the action of $f^{(n)}(x_0)$. If $h_1 = \cdots = h_n$ we simply write $\langle f^{(n)}(x_0), (h_1)^n\rangle$ where it does not cause ambiguity. If a functional F acts on a product space $Z_1 \times Z_2$ then elements of this space are denoted as $(h_1, h_2) \in Z_1 \times Z_2$ and $\langle F, (h_1, h_2)\rangle$ stands for the action of F on (h_1, h_2). For $x, y \in Z$, the following set is called an *interval* that connects x and y

$$[x, y] = \{\alpha x + (1-\alpha) y |\ \alpha \in [0,1]\}.$$

2 Main result

We make some more notations. Let

$$\begin{aligned} a(u) &= \mathbf{E}[X e^{\beta_0^\top X} G_T(u|X)], \qquad b(u) = \mathbf{E}[e^{\beta_0^\top X} G_T(u|X)],\\ p(u,x) &= \exp(\beta_0^\top X) G_T(u|X),\\ T(u) &= \mathbf{E}[XX^\top p(u,x)]\mathbf{E}[p(u,x)] - \mathbf{E}[Xp(u,x)]\mathbf{E}[X^\top p(u,x))]. \end{aligned}$$

Denote

$$A = \mathbf{E}\Big[XX^\top \exp(\beta_0^\top X) \int_0^Y \lambda_0(u)\,\mathrm{d}u\Big], \qquad M = \int_0^\tau T(u)K(u)G_C(u)\,\mathrm{d}u,$$

where $K(u) = \frac{\lambda_0(u)}{b(u)}$. Also introduce a sequence of random vectors

$$\xi_n := \sum_{i=1}^n \zeta_i,$$

with i.i.d. summands

$$\zeta_i = -\frac{\Delta_i a(Y_i)}{b(Y_i)} + \frac{\exp(\beta_0^\top W_i)}{M_U(\beta_0)} \int_0^{Y_i} a(u)K(u)\,\mathrm{d}u + \frac{\partial q}{\partial \beta}(Y_i, \Delta_i, W_i, \beta_0, \lambda_0).$$

Let $\Sigma_\beta = 4\mathrm{Cov}(\zeta_1)$, $m(\varphi_\lambda) = \int_0^\tau \varphi_\lambda(u)a(u)G_C(u)\,\mathrm{d}u$, $\Sigma_\varphi^2 = 4\mathrm{Var}[\langle q'(Y, \Delta, W, \lambda_0, \beta_0), \varphi\rangle]$ with $\varphi = (\varphi_\lambda, \varphi_\beta) \in C[0,\tau] \times \mathbb{R}^k$.

Theorem 1. *Assume conditions (i) to (ix). Then M is invertible and*

$$\sqrt{n}(\beta_n - \beta_0) \xrightarrow{d} N_k(0, M^{-1}\Sigma_\beta M^{-1}). \tag{2.1}$$

Moreover, for any Lipschitz continuous function f on $[0,\tau]$,

$$\sqrt{n}\int_0^\tau (\lambda_n - \lambda_0)(u) f(u) G_C(u)\,\mathrm{d}u \xrightarrow{d} N(0, \sigma_\varphi^2(f)) \tag{2.2}$$

where $\sigma_\varphi^2(f) = \sigma_\varphi^2$ with $\varphi = (\varphi_\lambda, \varphi_\beta)$, $\varphi_\beta = -A^{-1}m(\varphi_\lambda)$, and φ_λ is a unique solution to the Fredholm's integral equation

$$\frac{\varphi_\lambda}{K(u)} - a^\top(u)A^{-1}m(\varphi_\lambda) = f(u). \tag{2.3}$$

Corollary 2. *Let* $0 < \varepsilon < \tau$. *Assume that* $\frac{1}{G_C}$ *is Lipschitz continuous on* $[0, \tau - \varepsilon]$. *Under conditions (i) to (ix), for any Lipschitz continuous function* f *on* $[0, \tau]$ *with support on* $[0, \tau - \varepsilon]$,

$$\sqrt{n} \int_0^{\tau-\varepsilon} (\lambda_n - \lambda_0)(u) f(u)\,\mathrm{d}u \xrightarrow{d} N(0, \sigma_\varphi^2(f)) \tag{2.4}$$

where $\sigma_\varphi^2(f) = \sigma_\varphi^2$ *with* $\varphi = (\varphi_\lambda, \varphi_\beta)$, $\varphi_\beta = -A^{-1} m(\varphi_\lambda)$, *and* φ_λ *is a unique solution to the Fredholm's integral equation*

$$\frac{\varphi_\lambda}{K(u)} - a^\top(u) A^{-1} m(\varphi_\lambda) = \frac{f(u)}{G_C(u)}.$$

Here by definition $\frac{f(\tau)}{G_C(\tau)} = 0$.

Note that the corollary immediately follows from the theorem after f is substituted by $\frac{f}{G_C}$.

3 Computation of estimators

Since Θ_λ is infinite-dimensional, computation of (λ_n, β_n) is not a parametric problem in general setting. We refer to the ideas of I.J. Schoenberg [11]. We will show that maximum of (1.2) is attained on a linear spline with nodes located at points Y_i, $i = 1, \dots, n$ and some other points that can be calculated.

Let $i_1, \dots, i_n \in 1, \dots, n$ be such a numbering that $Y_{i_1} \le \dots \le Y_{i_n}$, i.e., $(Y_{i_1}, \dots, Y_{i_n})$ is a variational series of $(Y_1, \dots, Y_n)$. Alongside with (λ_n, β_n) we consider $(\overline{\lambda}_n, \beta_n)$, where $\overline{\lambda}_n$ is the following function. We set $\overline{\lambda}_n(Y_{i_k}) = \lambda(Y_{i_k})$, $k = 1, \dots, n$. For each interval $[Y_{i_k}, Y_{i_{k+1}}]$, $k = 1, \dots, n-1$, perform the next procedure. Draw straight lines

$$L_{i_k}^1(t) = \lambda(Y_{i_k}) + L(Y_{i_k} - t) \tag{3.1}$$

and

$$L_{i_k}^2(t) = \lambda(Y_{i_{k+1}}) + L(t - Y_{i_{k+1}}), \tag{3.2}$$

where L is defined in (i).

Denote B_{i_k} the intersection of $L_{i_k}^1(t)$ and $L_{i_k}^2(t)$. $B_{i_0} := 0$, $B_{i_n} := \tau$, $Y_{i_0} := 0$, $Y_{i_{n+1}} := \tau$. We set

$$\overline{\lambda}_n(t) = \begin{cases} \max\{L_{i_k}^1(t), a\} & \text{if } t \in [Y_{i_k}, B_{i_k}], \\ \max\{L_{i_k}^2(t), a\} & \text{if } t \in [B_{i_k}, Y_{i_{k+1}}]. \end{cases} \tag{3.3}$$

Note that $\lambda_n \ge \overline{\lambda}_n$ because $\lambda_n \in \Theta_\lambda$. Then

$$\int_{Y_{i_k}}^{Y_{i_{k+1}}} \lambda_n(u)\,\mathrm{d}u \ge \int_{Y_{i_k}}^{Y_{i_{k+1}}} \overline{\lambda}_n(u)\,\mathrm{d}u.$$

Thus, one can easily see that

$$Q_n(\lambda_n, \beta_n) \le Q_n(\overline{\lambda}_n, \beta_n)$$

implying $\lambda_n = \overline{\lambda}_n$ so that we conclude with the following statement.

Theorem 3. *Under conditions (i) and (ii), function λ_n that maximizes Q_n is a linear spline constructed in (3.3).*

Using maximization in (3.3) makes computation of (λ_n, β_n) inconvenient. Thus, we propose to modify the estimators. As soon as condition (viii) is satisfied and estimator (λ_n, β_n) is strongly consistent, one can induce that eventually $\overline{\lambda}(B_{i_k}) > a$, and thus, eventually there is no need in finding maximum in (3.3). Therefore, instead of (λ_n, β_n) we propose to consider a couple $(\widehat{\lambda}_n, \widehat{\beta}_n)$ with $\widehat{\beta}_n \in \Theta_\beta$ that maximizes Q_n under restrictions:

(1) $\widehat{\lambda}_n(0) \le A$.
(2) $\widehat{\lambda}_n(Y_{i_k}) \ge a, k = 1, \dots, n$.
(3) $\widehat{\lambda}_n(Y_{i_k}) + L(Y_{i_k} - Y_{i_{k+1}}) \le \widehat{\lambda}_n(Y_{i_{k+1}}) \le \widehat{\lambda}_n(Y_{i_k}) - L(Y_{i_k} - Y_{i_{k+1}}), k = 1, \dots, n-1$.
(4) $\widehat{\lambda}_n(t) := \begin{cases} L^1_{i_k}(t) \text{ if } t \in [Y_{i_k}, B_{i_k}], \\ L^2_{i_k}(t) \text{ if } t \in [B_{i_k}, Y_{i_{k+1}}], \end{cases} \quad k = 1, \dots, n-1.$
(5) $\widehat{\lambda}_n(t) := \begin{cases} L^2_{i_0}(t) \text{ if } t \in [0, Y_{i_1}], \\ L^1_{i_n}(t) \text{ if } t \in [Y_{i_n}, \tau]. \end{cases}$

Evaluating $(\widehat{\lambda}_n, \widehat{\beta}_n)$ is a parametric problem. We mention that eventually $(\widehat{\lambda}_n, \widehat{\beta}_n) = (\lambda_n, \beta_n)$. We summarise with the next statement.

Theorem 4. *Assume conditions (i) to (ix). Then estimator $(\widehat{\lambda}_n, \widehat{\beta}_n)$ is strongly consistent and statements of Theorem 1 and Corollary 2 hold true for that estimator.*

4 Stochastic boundedness of transformed and normalized estimators

Theorem 5. *Assume (i) to (vi). Then*

$$\sqrt[4]{n}\|\beta_n - \beta_0\| = O_p(1),$$
$$\sqrt{n}\int_0^\tau (\lambda_n(u) - \lambda_0(u))^2 G_C(u)\,\mathrm{d}u = O_p(1).$$

The proof is based on the three lemmas. Using integration by parts one can easily prove the following.

Lemma 6. *For all $u \in [0, \tau]$*

$$\int_u^\tau (f_C(y)G_T(y|X) + f_T(y|X)G_C(y))\,\mathrm{d}y = G_T(u|X)G_C(u) =: G(u|X).$$

Crucial step of the proof of Theorem 5 is the following.

Lemma 7. *There exists a closed bounded set A such that $\mu_X(A) := P(X \in A) > 0$ and that the identity $(v^\top x - c)I_A(x) \equiv 0$, for some $v \in \mathbb{R}^k$, $c \in \mathbb{R}$, implies $v = 0$ and $c = 0$.*

Proof of Lemma 7. Denote by M the support of μ_X, so that M is minimal closed set with $\mu_X(M) = \mu_X(\mathbb{R}^k)$. Since μ_X is not concentrated on a hyperplane due to the condition (vi), there are at least $k + 1$ distinct points $m_1, \dots, m_{k+1}$ that belong to

M and do not lie on a hyperplane. Consider a closed ball $\overline{B}(0,r)$ with radius $r > \max\{\|m_1\|,\dots,\|m_{k+1}\|\}$. Now one can take $A = M \cap \overline{B}(0,r)$ and make sure that A has all desired properties. □

Let $A_n(\omega)$ be a collection of assertions (here ω stands for elementary event). We say that $\{A_n\}$ hold *eventually* if for almost all ω there exists N_ω such that for all $n > N_\omega$, $A_n(\omega)$ holds.

Lemma 8. *Let η_n, ξ_n be two sequences of random variables, η_n be stochastically bounded, and eventually $|\xi_n| \le |\eta_n|$. Then ξ_n is stochastically bounded as well.*

Proof of Theorem 5.

Step 1. Denote $q_\infty(\lambda,\beta) = \mathbf{E}[\tilde{q}(Y,\Delta,W,\lambda,\beta)] = \mathbf{E}[\tilde{q}(Y,\Delta,X,\lambda,\beta)]$. Let us show that $(q_\infty)'$ exists for $(\lambda,\beta) \in B$ and equals zero at the true point (λ_0,β_0), where B is some open set in $\mathbb{R}^k \times C[0,\tau]$ that contains $\Theta_\beta \times \Theta_\lambda$.

Using (iv) one can easily obtain that

$$\frac{\partial q_\infty}{\partial \beta}(\lambda,\beta) = \mathbf{E}\Big[\Delta X - X\exp(\beta^\top X)\int_0^Y \lambda(u)\,\mathrm{d}u\Big],$$
$$\Big\langle \frac{\partial q_\infty}{\partial \lambda}(\lambda,\beta), h\Big\rangle = \mathbf{E}\Big[\frac{\Delta h(Y)}{\lambda(Y)} - \exp(\beta^\top X)\int_0^Y h(u)\,\mathrm{d}u\Big],$$

where $h \in C[0,\tau]$. Hence, $(q_\infty)'$ exists. According to [7], Section 3 $q_\infty(\lambda,\beta) < q_\infty(\lambda_0,\beta_0)$ for all $(\lambda,\beta) \ne (\lambda_0,\beta_0)$, $(\lambda,\beta) \in B$. Hence,

$$(q_\infty)'(\lambda_0,\beta_0) = 0.$$

In fact, condition (iv) implies that $(q_\infty)''$ and $(q_\infty)'''$ exist. Hence, third order Taylor's formula holds,

$$q_\infty(\lambda_n,\beta_n) - q_\infty(\lambda_0,\beta_0) = \frac{1}{2}\big\langle (q_\infty)''(\lambda_0,\beta_0), (\lambda_n-\lambda_0,\beta_n-\beta_0)^2\big\rangle + \frac{1}{6}\big\langle (q_\infty)'''(\tilde{\lambda}_n,\tilde{\beta}_n), (\lambda_n-\lambda_0,\beta_n-\beta_0)^3\big\rangle, \quad (4.1)$$

where $(\tilde{\lambda}_n,\tilde{\beta}_n)$ belongs to interval $[(\lambda_n,\beta_n),(\lambda_0,\beta_0)]$.

Step 2. We transform $(q_\infty)''$ and show that $-(q_\infty)''(\lambda_0,\beta_0)$ is a positive definite operator. We have

$$\Big\langle \frac{\partial^2 q_\infty(\lambda_0,\beta_0)}{\partial\lambda^2}, (h_1,h_2)\Big\rangle = -\mathbf{E}\Big[\frac{\Delta}{\lambda_0^2(Y)} h_1(Y)h_2(Y)\Big],$$
$$\frac{\partial^2 q_\infty}{\partial\beta^2}(\lambda_0,\beta_0) = -\mathbf{E}\Big[XX^\top \exp(\beta_0^\top X)\int_0^Y \lambda_0(u)\,\mathrm{d}u\Big],$$
$$\Big\langle \frac{\partial^2 q_\infty(\lambda_0,\beta_0)}{\partial\lambda\partial\beta}, (h_\lambda,h_\beta)\Big\rangle = -\mathbf{E}\Big[(h_\beta^\top X)\exp(\beta_0^\top X)\int_0^Y h_\lambda(u)\,\mathrm{d}u\Big],$$

where $h_1,h_2,h_\lambda \in C[0,\tau]$, $h_\beta \in \mathbb{R}^k$.

We use (1.4) and Lemma 6 for further transformations:

$$\Big\langle \frac{\partial^2 q_\infty(\lambda_0,\beta_0)}{\partial \lambda^2}, (h_\lambda, h_\lambda)\Big\rangle$$
$$= -\mathbf{E}\bigg[\frac{\Delta}{\lambda_0^2(Y)} h_\lambda^2(Y)\bigg] = \mathbf{E}\bigg(\int_0^\tau \frac{h_\lambda^2(u)}{\lambda_0^2(u)} f_T(u|X) G_C(u)\,\mathrm{d}u\bigg)$$
$$= \mathbf{E}\bigg(\int_0^\tau \frac{h_\lambda^2(u)}{\lambda_0^2(u)} \lambda_0(u)\exp(\beta_0^\top X)\exp\bigg(-\int_0^u \Lambda(s|X;\lambda_0,\beta_0)\,\mathrm{d}s\bigg) G_C(u)\,\mathrm{d}u\bigg)$$
$$= \mathbf{E}\bigg(\int_0^\tau \frac{h_\lambda^2(u)}{\lambda_0(u)} \exp(\beta_0^\top X) G_T(u|X) G_C(u)\,\mathrm{d}u\bigg). \tag{4.2}$$

Next,

$$\Big\langle \frac{\partial^2 q_\infty(\lambda_0,\beta_0)}{\partial \beta^2}, (h_\beta, h_\beta)\Big\rangle$$
$$= -\mathbf{E}\bigg[(h_\beta^\top X)^2 \exp(\beta_0^\top X)\int_0^Y \lambda_0(u)\,\mathrm{d}u\bigg]$$
$$= -\mathbf{E}\bigg[(h_\beta^\top X)^2 \exp(\beta_0^\top X)\bigg(\int_0^\tau\bigg(\int_0^y \lambda_0(u)\,\mathrm{d}u\, f_T(y|X)G_C(y)$$
$$+ \int_0^y \lambda_0(u)\,\mathrm{d}u\, f_C(y) G_T(y|X)\bigg)\,\mathrm{d}y\bigg)\bigg]$$
$$= -\mathbf{E}\bigg[(h_\beta^\top X)^2 \exp(\beta_0^\top X)\int_0^\tau \lambda_0(u)\int_u^\tau (f_C(y)G_T(y|X) + f_T(y|X)G_C(y))\,\mathrm{d}y\,\mathrm{d}u\bigg]$$
$$= -\mathbf{E}\bigg[(h_\beta^\top X)^2 \exp(\beta_0^\top X)\int_0^\tau \lambda_0(u) G_T(u|X) G_C(u)\,\mathrm{d}u\bigg]. \tag{4.3}$$

At last,

$$\Big\langle \frac{\partial^2 q_\infty(\lambda_0,\beta_0)}{\partial \lambda\partial\beta}, (h_\lambda, h_\beta)\Big\rangle$$
$$= -\mathbf{E}\bigg[(h_\beta^\top X)\exp(\beta_0^\top X)\int_0^Y h_\lambda(u)\,\mathrm{d}u\bigg]$$
$$= -\mathbf{E}\bigg[(h_\beta^\top X)\exp(\beta_0^\top X)\int_0^\tau h_\lambda(u)\int_u^\tau (f_C(y)G_T(y|X) + f_T(y|X)G_C(y))\,\mathrm{d}y\,\mathrm{d}u\bigg]$$
$$= -\mathbf{E}\bigg[(h_\beta^\top X)\exp(\beta_0^\top X)\int_0^\tau h_\lambda(u) G_T(u|X) G_C(u)\,\mathrm{d}u\bigg]. \tag{4.4}$$

Hence, from (4.2) to (4.4) it follows that

$$\big\langle (q_\infty)''(\lambda_0,\beta_0), (h_\lambda, h_\beta)^2\big\rangle$$
$$= -\mathbf{E}\bigg[\exp(\beta_0^\top X)\int_0^\tau \bigg((h_\beta^\top X)\sqrt{\lambda_0(u)G(u|X)} + h_\lambda(u)\frac{\sqrt{G(u|x)}}{\sqrt{\lambda_0(u)}}\bigg)^2\,\mathrm{d}u\bigg]. \tag{4.5}$$

Now, condition (vi) implies that $-(q_\infty)''$ is positive definite at (λ_0,β_0), i.e.,

$$\big\langle (q_\infty)''(\lambda_0,\beta_0), (h_\lambda, h_\beta)^2\big\rangle = 0 \quad\Longleftrightarrow\quad (h_\lambda, h_\beta) = (0,0).$$

Indeed, if to assume that $\langle (q_\infty)''(\lambda_0, \beta_0), (h_\lambda, h_\beta)^2\rangle = 0$ and $(h_\lambda, h_\beta) \neq (0,0)$ then (4.5) implies that $h_\beta \neq 0$ and $(h_\beta^\top X) = const\ a.s.$ We get a contradiction with (vi).

Step 3. We show that there exist such $C > 0$ and $\delta > 0$ that, whenever $\max\{\|h_\beta\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\} > 0$, it holds

$$\mathbf{E}\Bigg[\frac{-\langle (\tilde{q})''(Y, \Delta, X, \lambda_0, \beta_0), (h_\lambda, h_\beta)^2\rangle}{\max\{\|h_\beta\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\Bigg] \geq \delta. \tag{4.6}$$

Note that $G(u|X)$ is continuous in X. Denote $G_0(u) = \min_{X\in A} G(u|X)$, where A is a set from Lemma 7. Note that $G_0(u) = G(u|X_0) > 0$, for all $u \in [0, \tau)$ and some X_0.

Assume that $\|h_\beta\|^2 \geq \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u$. Jensen's inequality and (4.5) yield

$$\begin{aligned}
&-\big\langle (q_\infty)''(\lambda_0, \beta_0), (h_\lambda, h_\beta)^2\big\rangle \\
&\geq \frac{1}{\tau}\mathbf{E}\Bigg[I_{X\in A}\exp(\beta_0^\top X)\Bigg(\int_0^\tau (h_\beta^\top X)\sqrt{\lambda_0(u)G_0(u)} + h_\lambda(u)\frac{\sqrt{G_0(u)}}{\sqrt{\lambda_0(u)}}\,\mathrm{d}u\Bigg)^2\Bigg] \\
&= \frac{1}{\tau}\mathbf{E}\Bigg[I_{X\in A}\exp(\beta_0^\top X)\Bigg((h_\beta^\top X)\int_0^\tau \sqrt{\lambda_0(u)G_0(u)}\mathrm{d}u + \int_0^\tau h_\lambda(u)\frac{\sqrt{G_0(u)}}{\sqrt{\lambda_0(u)}}\,\mathrm{d}u\Bigg)^2\Bigg].
\end{aligned} \tag{4.7}$$

Denote

$$a_0 = \min_{X\in A}\frac{1}{\tau}\exp(\beta_0^\top X), \qquad a_1 = \int_0^\tau \sqrt{\lambda_0(u)G_0(u)}\,\mathrm{d}u,$$
$$K_\lambda(h_\beta) = \frac{\int_0^\tau h_\lambda(u)\frac{\sqrt{G_0(u)}}{\sqrt{\lambda_0(u)}}\,\mathrm{d}u}{\|h_\beta\|}.$$

Inequality (4.7) implies that

$$\mathbf{E}\Bigg[\frac{-\langle (\tilde{q})''(Y, \Delta, X, \lambda_0, \beta_0), (h_\lambda, h_\beta)^2\rangle}{\max\{\|h_\beta\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\Bigg] \geq a_0\mathbf{E}[I_{X\in A}((\hat{h}_\beta^\top X)a_1 + K_\lambda)^2],$$

where $\hat{h}_\beta = \frac{h_\beta}{\|h_\beta\|}$. Fix $T \in \mathbb{R}$. Equality

$$\mathbf{E}[I_{X\in A}((\hat{h}_\beta^\top X)a_1 + T)^2] = 0$$

implies that $I_{X\in A}((\hat{h}_\beta^\top X) + \frac{T}{a_1}) = const\ a.s.$, which contradicts to the choice of A.

It is easy to see that for a fixed $\hat{h}_\beta$, minimum of

$$\mathbf{E}[I_{X\in A}((\hat{h}_\beta^\top X)a_1 + T)^2]$$

is attained at a unique point $T = T(\hat{h}_\beta, A)$. Moreover, $T(\hat{h}_\beta, A)$ is a continuous function of $\hat{h}_\beta$. Hence, we have

$$\mathbf{E}\Big[\frac{-\langle(\tilde{q})''(Y,\Delta,X,\lambda_0,\beta_0),(h_\lambda,h_\beta)^2\rangle}{\max\{\|h_\beta\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\Big] \geq a_0\mathbf{E}\big[I_{X\in A}\big((\hat{h}_\beta^\top X)a_1 + T(\hat{h}_\beta, A)\big)^2\big] > 0. \tag{4.8}$$

Due to $\|\hat{h}_\beta\| = 1$, the right hand side of (4.8) attains its minimum at some point $\hat{h}_{\beta_0}$. Now one can take

$$\delta_1 = a_0\mathbf{E}\big[I_{X\in A}\big((\hat{h}_{\beta_0}^\top X)a_1 + T(\hat{h}_{\beta_0}, A)\big)^2\big] > 0.$$

Consider the second case, where inequality $\|h_\beta\|^2 < \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0(u)}\,\mathrm{d}u$ holds. Transform right hand side of (4.5):

$$\begin{aligned}\mathbf{E}\Big[&\exp(\beta_0^\top X)\int_0^\tau\Big((h_\beta^\top X)\sqrt{\lambda_0(u)G(u|X)} + h_\lambda(u)\frac{\sqrt{G(u|x)}}{\sqrt{\lambda_0(u)}}\Big)^2\,\mathrm{d}u\Big]\\ &\geq \mathbf{E}\Big[I_{X\in A}\exp(\beta_0^\top X)\Big((h_\beta^\top X)^2\int_0^\tau \lambda_0(u)G_0(u)\,\mathrm{d}u + 2\int_0^\tau (h_\beta^\top X)h_\lambda(u)G_0(u)\,\mathrm{d}u\\ &\quad + \int_0^\tau h_\lambda^2(u)\frac{G_0(u)}{\lambda_0(u)}\,\mathrm{d}u\Big)\Big].\end{aligned}$$

Denote $\Phi = \int_0^\tau h_\lambda^2(u)\frac{G_0(u)}{C\lambda_0(u)}\,\mathrm{d}u$. Hence, the left hand side of (4.6) is transformed to

$$\begin{aligned}\mathbf{E}\Big[&\frac{-\langle(\tilde{q})''(Y,\Delta,X,\lambda_0,\beta_0),(h_\lambda,h_\beta)^2\rangle}{\Phi}\Big]\\ &\geq \mathbf{E}\Big[I_{X\in A}\exp(\beta_0^\top X)\Big((\tilde{h}_\beta^\top X)^2 a_2 + \frac{2}{\Phi}\int_0^\tau ({h_\beta}^\top X)h_\lambda(u)G_0(u)\,\mathrm{d}u + C\Big)\Big],\end{aligned}$$

where

$$a_2 = \int_0^\tau \lambda_0(u)G(u|X)\,\mathrm{d}u, \qquad \tilde{h}_\beta = \frac{h_\beta}{\sqrt{\int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0(u)}\,\mathrm{d}u}}.$$

Jensen's inequality implies

$$\Phi^{1/2} \geq \sqrt{\frac{1}{\tau C}}\Big(\int_0^\tau |h_\lambda(u)|\sqrt{\frac{G_0(u)}{\lambda_0(u)}}\,\mathrm{d}u\Big).$$

Since $\sqrt{\Phi} > \|h_\beta\|$, $G_0(u) \in [0,1]$ and λ_0 is bounded away from 0, we have

$$\begin{aligned}\Big|\frac{\int_0^\tau ({h_\beta}^\top X)h_\lambda(u)G_0(u)\,\mathrm{d}u}{\Phi}\Big| &\leq \frac{\|h_\beta\|}{\sqrt{\Phi}}\|X\|\Big|\frac{\tau^{1/2}\int_0^\tau h_\lambda(u)\sqrt{G_0(u)}\,\mathrm{d}u}{\int_0^\tau |h_\lambda(u)|\sqrt{\frac{G_0(u)}{\lambda_0(u)}}\,\mathrm{d}u}\Big|\sqrt{C}\\ &\leq \sqrt{C}\|X\|D,\end{aligned}$$

for some constant $D > 0$ which depends only on τ and λ_0. Since $\|\tilde{h}_\beta\| < 1$, there exist constants $K_1 > 0,\ K_2 > 0$ that satisfy

$$\mathbf{E}\Big[\frac{-\langle(\tilde{q})''(Y,\Delta,X,\lambda_0,\beta_0),(h_\lambda,h_\beta)^2\rangle}{\Phi}\Big] \ge \tau a_0(-K_1 - \sqrt{C}K_2 + C).$$

Choosing C large enough, we get (4.6).

Step 4. Now transform Taylor's decomposition (4.1):

$$\begin{aligned}&q_\infty(\lambda_n,\beta_n) - q_\infty(\lambda_0,\beta_0)\\ &\quad= \mathbf{E}\bigg(\max\bigg\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h^2_{\lambda_n}(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\bigg\}\bigg[\frac{1}{2}\frac{\langle(\tilde{q})''(Y,\Delta,X,\lambda_0,\beta_0),(h_{\lambda_n},h_{\beta_n})^2\rangle}{\max\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h^2_{\lambda_n}(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\\ &\qquad+\frac{1}{6}\frac{\langle(\tilde{q})'''(Y,\Delta,X,\tilde{\lambda}_n,\tilde{\beta}_n),(h_{\lambda_n},h_{\beta_n})^3\rangle}{\max\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h^2_{\lambda_n}(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\bigg]\bigg),\end{aligned}\tag{4.9}$$

where we denote $h_{\lambda_n} = \lambda_n - \lambda_0$ and $h_{\beta_n} = \beta_n - \beta_0$. Remember that $G_T(u|X) \in (0,1]$ for all X, so that $G_0(u) \ge K_3 G_C(u)$ for some $K_3 > 0$. One can see that $\frac{\partial^3 q_\infty}{\partial\lambda^2\partial\beta} = 0$. Using the same technique as in (4.2)–(4.4) and the assumptions, we get

$$\begin{aligned}&\Big\langle\frac{\partial^3 q_\infty(\tilde{\lambda}_n,\tilde{\beta}_n)}{\partial\lambda^3},(h_1,h_2,h_3)\Big\rangle\\ &\quad= \frac{1}{2}\mathbf{E}\Big[\frac{\Delta}{\tilde{\lambda}_n^3(Y)}h_1(Y)h_2(Y)h_3(Y)\Big]\\ &\quad= \frac{1}{2}\mathbf{E}\bigg(\int_0^\tau \frac{h_1(u)h_2(u)h_3(u)}{\tilde{\lambda}_n^2(u)}\exp(\tilde{\beta}_n^\top X)G_T(u|X)G_C(u)\,\mathrm{d}u\bigg)\\ &\quad\le K_4\|h_1\|\int_0^\tau h_2(u)h_3(u)G_0(u)\,\mathrm{d}u,\end{aligned}\tag{4.10}$$

$$\Big\langle\frac{\partial^3 q_\infty(\tilde{\lambda}_n,\tilde{\beta}_n)}{\partial\beta^3},h_\beta\Big\rangle = -\mathbf{E}\Big[(h_\beta^\top X)^3\exp(\tilde{\beta}_n^\top X)\int_0^Y\tilde{\lambda}_n\,\mathrm{d}u\Big] \le K_5\|h_\beta\|^3,\tag{4.11}$$

$$\begin{aligned}&\Big\langle\frac{\partial^3 q_\infty(\tilde{\lambda}_n,\tilde{\beta}_n)}{\partial\lambda\partial\beta^2},(h_\lambda,h_\beta,h_\beta)\Big\rangle\\ &\quad= -\mathbf{E}\Big[(h_\beta^\top X)^2\exp(\tilde{\beta}_n^\top X)\int_0^Y h_\lambda(u)\,\mathrm{d}u\Big] \le K_6\|h_\beta\|^2\|h_\lambda\|\end{aligned}\tag{4.12}$$

where K_4 to K_6 are positive constants. We note that all constants K_3 to K_6 depend only on $\Theta = \Theta_\lambda \times \Theta_\beta$. Kukush et al. [7] prove strong consistency of the estimator (λ_n,β_n), that is $\max_{t\in[0,\tau]}|\lambda_n(t) - \lambda_0(t)| \to 0$ and $\beta_n \to \beta_0$ a.s., as $n \to \infty$. One can conclude that

$$\lim_{n\to\infty}\mathbf{E}\Big[\frac{\langle(\tilde{q})'''(Y,\Delta,X,\tilde{\lambda}_n,\tilde{\beta}_n),(h_{\lambda_n},h_{\beta_n})^3\rangle}{\max\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h^2_{\lambda_n}(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\Big] = 0 \quad a.s.\tag{4.13}$$

Step 5. Set $S_n(\lambda,\beta) = n(Q_n(\lambda,\beta) - q_\infty(\lambda,\beta))$. Kukush et al. [7] prove that under assumptions (i) to (vi) $\frac{S_n(\lambda,\beta)}{\sqrt{n}}$ converges in distribution in $C(\Theta)$ to a Gaussian measure. Hence,

$$\begin{aligned}0 &\le \sqrt{n}(q_\infty(\lambda_0,\beta_0) - q_\infty(\lambda_n,\beta_n))\\ &\le \sqrt{n}(Q_n(\lambda_n,\beta_n) - q_\infty(\lambda_n,\beta_n) - Q_n(\lambda_0,\beta_0) + q_\infty(\lambda_0,\beta_0))\\ &\le 2\sqrt{n} \sup_{(\lambda,\beta)\in\Theta_\lambda\times\Theta_\beta} |Q_n(\lambda,\beta) - q_\infty(\lambda,\beta)| = O_p(1),\end{aligned}$$

because $q_\infty(\lambda,\beta)$ and $Q_n(\lambda,\beta)$ attain their maximums at (λ_0,β_0) and (λ_n,β_n), respectively.

Now, (4.9) yields

$$\begin{aligned}&\sqrt{n}\max\biggl\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\biggr\}\mathbf{E}\biggl(\biggl[\frac{1}{2}\frac{\langle(\tilde{q})''(Y,\Delta,X,\lambda_0,\beta_0),(h_{\lambda_n},h_{\beta_n})^2\rangle}{\max\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\\ &\quad + \frac{1}{6}\frac{\langle(\tilde{q})'''(Y,\Delta,X,\tilde{\lambda}_n,\tilde{\beta}_n),(h_{\lambda_n},h_{\beta_n})^3\rangle}{\max\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\}}\biggr]\biggr)\\ &\quad = \sqrt{n}(q_\infty(\lambda_0,\beta_0) - q_\infty(\lambda_n,\beta_n)) = O_p(1).\end{aligned}$$

Step 6. Equations (4.6), (4.9) and (4.13) imply that eventually

$$\sqrt{n}\max\biggl\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h_\lambda^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\biggr\} < \frac{\sqrt{n}(q_\infty(\lambda_0,\beta_0) - q_\infty(\lambda_n,\beta_n))}{\delta/3}.$$

Lemma 8 proves that $\sqrt{n}\max\{\|h_{\beta_n}\|^2, \int_0^\tau \frac{h_{\lambda_n}^2(u)G_0(u)}{C\lambda_0}\,\mathrm{d}u\} = O_p(1)$. Hence the first equation of Theorem 5 is proved:

$$\sqrt{n}\|\beta_n - \beta_0\|^2 = \sqrt{n}\|h_{\beta_n}\|^2 = O_p(1).$$

Finally, $G_0(u) \ge K_3 G_C(u)$. Note that λ_0 is bounded away from 0 on $[0,\tau]$. Hence

$$\sqrt{n}\int_0^\tau h_{\lambda_n}^2(u)G_C(u)\,\mathrm{d}u = O_p(1).$$

Thus, Theorem 5 is proved. □

5 Auxiliary results

We use the ideas of [3].

Let $\theta_n = (\lambda_n,\beta_n)$, $\theta_0 = (\lambda_0,\beta_0)$, $\Theta = \Theta_\lambda \times \Theta_\beta$. Denote $\varphi = (\varphi_\lambda,\varphi_\beta)$ an admissible shift such that there exists $\delta > 0$ with $\theta_0 \pm \delta\varphi \in \Theta$. We demand that (vii)–(viii) hold. Note that φ can be a random element and depend on n. However, $\|\varphi\|$ should be bounded from above a.s.

Consider the function $f(t) = Q_n(\theta_n + t(\theta_0 - \theta_n \pm \delta\varphi))$, $0 \le t \le 1$. It is well-defined (due to the convexity of Θ) and attains its maximum at point $t = 0$. Therefore, $\langle Q_n'(\theta_n), \theta_0 - \theta_n \pm \delta\varphi\rangle \le 0$ and

$$|\langle Q_n'(\theta_n), \varphi\rangle| \le \frac{1}{\delta}\langle Q_n'(\theta_n), \Delta\theta_n\rangle,$$

where $\Delta\theta_n := \theta_n - \theta_0$.

Taylor's expansion at point (λ_0, β_0) implies

$$\begin{aligned} &\Big|\big\langle Q_n'(\theta_0), \varphi\big\rangle + \frac{1}{2}\big\langle Q_n''(\theta_0), (\Delta\theta_n, \varphi)\big\rangle + \frac{1}{6}\big\langle Q_n'''(\tilde{\theta}_n), (\Delta\theta_n^2, \varphi)\big\rangle\Big| \\ &\quad \le \frac{1}{\delta}\Big(\big\langle Q_n'(\theta_0), \Delta\theta_n\big\rangle + \frac{1}{2}\big\langle Q_n''(\theta_0), \Delta\theta_n^2\big\rangle + \frac{1}{6}\big\langle Q_n'''(\widehat{\theta}_n), \Delta\theta_n^3\big\rangle\Big), \end{aligned} \tag{5.1}$$

for some $\widehat{\theta}_n$ and $\tilde{\theta}_n$ from interval $[\theta_0, \theta_n]$.

Proposition 9. *Under conditions (i) to (viii) for every admissible shift φ, one has that $\sqrt{n}\langle Q_n''(\theta_0), (\Delta\theta_n, \varphi)\rangle$ and $\sqrt{n}\langle q_\infty''(\theta_0), (\Delta\theta_n, \varphi)\rangle$ are stochastically bounded.*

Relying on this proposition we will be able to show that $\sqrt{n}\|\beta_n - \beta_0\|$ and $\sqrt{n}\int_0^\tau(\lambda_n - \lambda_0)(u)G_C(u)\,du$ are stochastically bounded and then prove the asymptotic normality of $\sqrt{n}\langle Q_n''(\theta_0), (\Delta\theta_n, \varphi)\rangle$.

Denote $\Theta_- = \Theta - \Theta$. It is clear that it is compact and convex. Before proving the proposition, we show the following.

Lemma 10. *Under conditions (i) to (viii), $\sqrt{n}Q_n'(\theta_0)$ and $\sqrt{n}(Q_n''(\theta_0) - q_\infty''(\theta_0))$ converge in distribution in $C(\Theta_-)$ and $C(\Theta_-^2)$, respectively. Moreover, for all $\theta = (\lambda, \beta) \in \Theta$, $\sqrt{n}(\frac{\partial^3 Q_n}{\partial\lambda^3}(\theta) - \frac{\partial^3 q_\infty}{\partial\lambda^3}(\theta))$ converges in distribution in $C(\Theta_-^3)$.*

Proof of Lemma 10. Here only convergence for $\sqrt{n}Q_n'(\theta_0)$ will be shown, because for $\sqrt{n}(Q_n''(\theta_0) - q_\infty''(\theta_0))$ and $\sqrt{n}(\frac{\partial^3 Q_n}{\partial\lambda^3}(\theta) - \frac{\partial^3 q_\infty}{\partial\lambda^3}(\theta))$ the proof is similar. We note that $q_\infty'(\theta_0) = 0$ and due to conditions (iii)–(iv) we have $\mathbf{E}[\sup_{\beta\in\Theta_\beta} e^{2\beta^\top X}\|X\|^k] < \infty$ and $\mathbf{E}[\sup_{\beta\in\Theta_\beta} e^{2\beta^\top U}\|U\|^k] < \infty$, for any $k \in \mathbb{N}$.

For $(\lambda, \beta) \in \Theta_-$ let

$$g(\lambda, \beta, Y, \Delta, W) = \big\langle q'(Y, \Delta, W, \lambda_0, \beta_0), (\lambda, \beta)\big\rangle$$

and

$$\rho\big((\lambda_1, \beta_1), (\lambda_2, \beta_2)\big) = \sup_{u\in[0,\tau]} |\lambda_1(u) - \lambda_2(u)| + \|\beta_1 - \beta_2\|.$$

(Θ_-, ρ) is a compact metric space. We denote by $Lip(\rho)$ a subspace of Lipschitz continuous functions on Θ_- with respect to the metric ρ and by $\|\cdot\|_\rho$ the norm induced by ρ, that is for some fixed point $(\lambda^*, \beta^*) \in \Theta_-$ and for all $l \in Lip(\rho)$ we define:

$$\|l\|_\rho := \sup_{(\lambda_1,\beta_1)\ne(\lambda_2,\beta_2)} \frac{|l(\lambda_1, \beta_1) - l(\lambda_2, \beta_2)|}{\rho((\lambda_1, \beta_1), (\lambda_2, \beta_2))} + l(\lambda^*, \beta^*).$$

We apply Theorem 2 from [12]. It states that $\sqrt{n}Q_n'(\theta_0)$ converges in distribution in $C(\Theta_-)$ under the following conditions:

(1) $\mathbf{P}(g \in Lip(\rho)) = 1$.
(2) $\mathbf{E}\|g\|_\rho^2 < \infty$.
(3) $\int_{0^+} H^{\frac{1}{2}}(\Theta_-, u)\,du < \infty$, where H is ε-entropy on (Θ_-, ρ), i.e. $H(\Theta_-, u) = \log_2 N(\Theta_-, u)$, where N is a minimal number of balls with diameter not exceeding 2ε that cover Θ_-.

Let $\Theta_{\lambda_-} = \Theta_\lambda - \Theta_\lambda$ and $\Theta_{\beta_-} = \Theta_\beta - \Theta_\beta$, so that $\Theta_- = \Theta_{\lambda_-} \times \Theta_{\beta_-}$. Consider Θ_{λ_-} and Θ_{β_-} as compact metric spaces with uniform and Euclidean norm, respectively. Then for $N(\Theta_-, 2u) \le N(\Theta_{\beta_-}, u)N(\Theta_{\lambda_-}, u)$, (3) is equivalent to

(3.1) $\int_{0^+} H^{\frac{1}{2}}(\Theta_{\lambda_-}, u)\,du < \infty$, and
(3.2) $\int_{0^+} H^{\frac{1}{2}}(\Theta_{\beta_-}, u)\,du < \infty$.

Since $\Theta_{\beta_-} \subset \mathbb{R}^k$, we have $N(\Theta_{\beta_-}, u) < Cu^k$ for some constant $C > 0$, and (3.2) is fulfilled. Note that Θ_{λ_-} can be considered as a set of Lipschitz continuous functions that map compact connected space $[0, \tau]$ into some interval in $\mathbb{R}$. Lemma 1 from [8] implies

$$H(\Theta_{\lambda_-}, u) \ge 1 + H(\Theta_{\lambda_-}, 4u),$$

so that Θ_{λ_-} is of "uniform type" (see [8]). According to Theorem 1 from [8] there exists such constant C that

$$H(\Theta_{\lambda_-}, 4L\varepsilon) \le CN([0, \tau], \varepsilon).$$

For the space $\mathbb{R}^1$ we have that $N([0, \tau], u) < \tilde{C}\frac{1}{u}$ for some constant $\tilde{C}$. Hence (3.1) holds.

To verify (1) and (2) note that

$$g(\lambda, \beta, Y, \Delta, W) = \frac{\Delta\lambda(Y)}{\lambda_0(Y)} - \frac{e^{\beta_0^\top W}}{M_U(\beta_0)} \int_0^Y \lambda\,du + \Delta\beta^\top W + \beta^\top \frac{(M_U(\beta_0)W - \mathbf{E}(Ue^{\beta_0^\top U}))e^{\beta_0^\top W}}{M_U^2(\beta_0)} \int_0^Y \lambda_0\,du,$$

and conditions (i)–(ii) imply

$$\sup_{(\lambda,\beta)\in\Theta_-} \|g'(\lambda, \beta, Y, \Delta, W)\| < \infty,$$

where g' is considered as a bilinear operator on $C[0, \tau] \times \mathbb{R}^k$. Hence, condition (1) is fulfilled. Moreover, there exists such a constant $K > 0$ that

$$\|g(\lambda, \beta, Y, \Delta, W)\|_\rho < K\big(1 + \|W\| + e^{D\|W\|} + \|W\|e^{D\|W\|}\big)$$

and due to conditions (iii) and (iv), condition (2) is also satisfied. Thus, lemma is proved. □

Returning to inequality (5.1), because $\Delta\theta_n$ converges to zero a.s., one can conclude the following.

(a) $\sqrt{n}\langle Q_n'(\theta_0), \varphi\rangle = O_p(1)$ and $\langle \sqrt{n}Q_n'(\theta_0), \Delta\theta_n\rangle = o_p(1)$, where $o_p(1)$ means convergence to zero in probability.

(b) $\sqrt{n}(\frac{\partial^3 Q_n}{\partial\lambda^3}(\theta) - \frac{\partial^3 q_\infty}{\partial\lambda^3}(\theta))$ converges in probability in $C(\Theta_-^3)$. Inequality (4.10) implies that $\sqrt{n}\langle\frac{\partial^3 q_\infty}{\partial\lambda^3}(\theta), (\Delta\theta_n^2, \varphi)\rangle$ is stochastically bounded, so is $\sqrt{n}\langle\frac{\partial^3 Q_n}{\partial\lambda^3}(\theta), (\Delta\theta_n^2, \varphi)\rangle$.

(c) $\sqrt{n}\langle(Q_n''(\theta_0) - q_\infty''(\theta_0)), \Delta\theta_n^2\rangle$ and $\sqrt{n}\langle(Q_n''(\theta_0) - q_\infty''(\theta_0)), (\Delta\theta_n, \varphi)\rangle$ converge to zero in probability. Note that $\langle\sqrt{n}Q_n''(\theta_0), \Delta\theta_n^2\rangle = O_p(1)$ if and only if $\sqrt{n}\langle q_\infty''(\theta_0), \Delta\theta_n^2\rangle = O_p(1)$. The latter equality can be easily derived from Theorem 5, formula (4.1) and convergence (4.13).

Proof of Proposition 9. To prove the first part of the proposition one has to show that

$$\langle Q_n'''(\widehat{\theta}_n), \Delta\theta_n^3\rangle = \frac{O_p(1)}{\sqrt{n}} \tag{5.2}$$

and

$$\langle Q_n'''(\tilde{\theta}_n), (\Delta\theta_n^2, \varphi)\rangle = \frac{O_p(1)}{\sqrt{n}}. \tag{5.3}$$

It is clear that (5.3) yields (5.2). After a series of computations one can induce that for some constants $C_1 > 0,\ C_2 > 0$

$$\begin{aligned}\left|\left\langle\frac{\partial^3 q(Y,\Delta,W,\lambda,\beta)}{\partial\beta^3}, (h_\beta)^3\right\rangle\right| &\le C_1 e^{D\|W\|}\|h_\beta\|^3\\ \left|\left\langle\frac{\partial^3 q(Y,\Delta,W,\lambda,\beta)}{\partial\lambda\partial\beta^2}, (h_\beta, h_\beta, h_\lambda)\right\rangle\right| &\le C_2 e^{D\|W\|}\|h_\beta\|^2\|h_\lambda\|.\end{aligned} \tag{5.4}$$

Expectations of right hand sides of inequalities in (5.4) are finite. Together with $\sqrt{n}\|\beta_n - \beta_0\|^2 = O_p(1)$ and SLLN, this implies that $\langle\frac{\partial Q_n^3(\tilde{\theta}_n)}{\partial\beta^3}, (\Delta\theta_n^2, \varphi)\rangle$ and $\langle\frac{\partial Q_n^3(\tilde{\theta}_n)}{\partial\beta^2\partial\lambda}, (\Delta\theta_n^2, \varphi)\rangle$ are $\frac{O_p(1)}{\sqrt{n}}$. Noting that $\frac{\partial Q_n^3(\tilde{\theta}_n)}{\partial\beta\partial\lambda^2} = 0$, one can conclude that the first part of the proposition will be proven if one shows that

$$\begin{aligned}\left\langle\frac{\partial Q_n^3(\tilde{\theta}_n)}{\partial\lambda^3}, (\Delta\theta_n^2, \varphi)\right\rangle &= \left\langle\frac{\partial Q_n^3(\tilde{\lambda}_n)}{\partial\lambda^3}, ((\lambda_n - \lambda_0), (\lambda_n - \lambda_0), \varphi_\lambda)\right\rangle\\ &= \frac{1}{n}\sum_{i=1}^n \Delta_i \frac{(\lambda_n - \lambda_0)^2(Y_i)\varphi_\lambda(Y_i)}{\tilde{\lambda}_n^3(Y_i)} = \frac{O_p(1)}{\sqrt{n}}.\end{aligned} \tag{5.5}$$

From the definition of admissible shifts, φ_λ belongs to Θ. Since $\tilde{\lambda}_n$ is bounded away from zero, there is a constant C such that for the second summand we have

$$\left|\left\langle\frac{\partial Q_n^3(\tilde{\theta}_n)}{\partial\lambda^3}, (\Delta\theta_n^2, \varphi)\right\rangle\right| \le C\left|\frac{1}{n}\sum_{i=1}^n \Delta_i(\lambda_n - \lambda_0)^2(Y_i)\right| = \frac{O_p(1)}{\sqrt{n}},$$

where the last equality holds due to the conclusion (b). Thus, (5.5) holds. This completes the proof of the first part of the proposition. The second part is easily derived from conclusion (c). □

Corollary 11.

$$\sqrt{n}\|\beta_n - \beta_0\| = O_p(1),$$
$$\sqrt{n}\int_0^\tau (\lambda_n - \lambda_0)(u) G_C(u)\,\mathrm{d}u = O_p(1).$$

Proof. Let $h_\beta = \beta_n - \beta_0$, $h_\lambda = \lambda_n - \lambda_0$. Take some admissible shift $\varphi := (\varphi_\lambda, \varphi_\beta)$. For this shift one has

$$\begin{aligned}
-\langle q''_\infty(\theta_0), (\Delta\theta_n, \varphi)\rangle &= \mathbf{E}\Big[({h_\beta}^\top X)(\varphi_\beta^\top X)\exp(\beta_0^\top X)\int_0^Y \lambda_0(u)\,\mathrm{d}u\Big] \\
&\quad + \mathbf{E}\Big[({\varphi_\beta}^\top X)\exp(\beta_0^\top X)\int_0^Y h_\lambda(u)\,\mathrm{d}u\Big] \\
&\quad + \mathbf{E}\Big[\frac{\Delta}{\lambda_0^2(Y)} h_\lambda(Y)\varphi_\lambda(Y)\Big] \\
&\quad + \mathbf{E}\Big[({h_\beta}^\top X)\exp(\beta_0^\top X)\int_0^Y \varphi_\lambda(u)\,\mathrm{d}u\Big] \\
&= O_p\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned} \tag{5.6}$$

The idea is to find such φ_λ that

$$E\Big[({\varphi_\beta}^\top X)\exp(\beta_0^\top X)\int_0^Y h_\lambda(u)\,\mathrm{d}u\Big] + \mathbf{E}\Big[\frac{\Delta}{\lambda_0^2(Y)} h_\lambda(Y)\varphi_\lambda(Y)\Big] = 0. \tag{5.7}$$

Then after some calculations (using Lemma 6) one can see that (5.7) is equivalent to

$$\int_0^\tau h_\lambda \varphi_\beta^\top a(u) G_C(u)\,\mathrm{d}u + \int_0^\tau \frac{h_\lambda}{\lambda_0}\varphi_\lambda b(u) G_C(u)\,\mathrm{d}u = 0. \tag{5.8}$$

One can take

$$\varphi_\lambda(u) := -\frac{\varphi_\beta^\top a(u)}{b(u)}\lambda_0(u)$$

as a solution to (5.8). Since $G_T(u|X)$ is differentiable function of u, one can conclude that φ_λ is an admissible shift for $\|\varphi_\beta\|$ small enough.

Equation (5.6) is now equivalent to

$$\int_0^\tau h_\beta^\top T(u)\varphi_\beta \frac{\lambda_0(u)G_C(u)}{b(u)}\,\mathrm{d}u = O_p\Big(\frac{1}{\sqrt{n}}\Big). \tag{5.9}$$

Using Hölder's inequality and condition (vi), one can easily see that $T(u)$ is positive definite. Now let $\tilde{h}_\beta = \frac{\beta_n - \beta_0}{\|\beta_n - \beta_0\|}$ and take $\varphi_\beta = \frac{\tilde{h}_\beta}{C_1}$, where $C_1 > 0$ such that $\varphi = (\varphi_\lambda, \varphi_\beta)$ is an admissible shift. Then (5.6) can be transformed to

$$\|h_\beta\|\int_0^\tau \tilde{h}_\beta^\top T(u)\tilde{h}_\beta \frac{\lambda_0(u)G_C(u)}{b(u)}\,\mathrm{d}u = O_p\Big(\frac{1}{\sqrt{n}}\Big). \tag{5.10}$$

Since $\|\tilde{h}_\beta\| = 1/C_1$, left hand side of (5.10) is greater than $\delta\|h_\beta\|$ for some $\delta > 0$. Using Lemma 8 the first part of the corollary is proved.

If now in (5.6) one takes $\varphi = (\frac{1}{C_2}, 0)$ for large enough $C_2 > 0$, then (5.6) takes form

$$\mathbf{E}\left[\frac{\Delta}{\lambda_0^2(Y)} h_\lambda(Y) \frac{1}{C_2}\right] + \mathbf{E}\left[(h_\beta{}^\top X) \exp(\beta_0^\top X) \int_0^Y \frac{1}{C_2} \,\mathrm{d}u\right] = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Due to $\sqrt{n}\|\beta_n - \beta_0\| = O_p(1)$, the latter equality implies

$$\mathbf{E}\left[\frac{\Delta}{\lambda_0^2(Y)} h_\lambda(Y) \frac{1}{C_2}\right] = O_p\left(\frac{1}{\sqrt{n}}\right)$$

and the second part of the corollary holds. □

We present the main result of this section.

Theorem 12. *Under conditions (i) to (ix), for all admissible shifts the following convergence in probability holds*

$$\sqrt{n}\langle Q_n'(\theta_0), \varphi\rangle + \frac{1}{2}\sqrt{n}\langle q_\infty''(\theta_0), (\Delta\theta_n, \varphi)\rangle \xrightarrow{P} 0. \tag{5.11}$$

Moreover, if φ is a non-random admissible shift then $\langle Q_n'(\theta_0), \varphi\rangle \xrightarrow{d} N(0, \sigma_\varphi^2)$, where $\sigma_\varphi^2 = 4\mathrm{Var}[\langle q'(Y, \Delta, W, \lambda_0, \beta_0), \varphi\rangle]$, and

$$\sqrt{n}\langle Q_n''(\theta_0), (\Delta\theta_n, \varphi)\rangle \xrightarrow{d} N(0, \sigma_\varphi^2),$$
$$\sqrt{n}\langle q_\infty''(\theta_0), (\Delta\theta_n, \varphi)\rangle \xrightarrow{d} N(0, \sigma_\varphi^2).$$

Proof. Using Corollary 11 and inequality (5.1), one can repeat the proof of Proposition 9 with a remark that stochastic boundedness should be changed for a convergence to zero in probability. We use (4.2) to (4.4) to show $\sqrt{n}\langle q_\infty''(\theta_0), \Delta\theta_n^2\rangle = o_p(1)$. Thus, the convergence (5.11) is proved. The rest of the proof is trivial. □

6 Proof of Theorem 1

We assume that condition (ix) is satisfied. Thus A is positive definite and, consequently, invertible. Since $T(u)$ is positive definite, M is positive definite as well and therefore, invertible.

Note that due to conditions (vii) and (viii), Theorem 12 is valid for all non-random shifts $\varphi \in \mathbb{R}^k \times Lip_1([0, \tau])$, where $Lip_1([0, \tau])$ is a class of Lipschitz continuous functions. Take the shift as follows: $\varphi = (\varphi_\beta^\top a(u)K(u), \varphi_\beta)$, where φ_β is a fixed vector of $\mathbb{R}^k$. For this shift one can rewrite $\langle Q_n'(\theta_0), \varphi\rangle$ as

$$\langle Q_n'(\theta_0), \varphi\rangle = \varphi_\beta^\top \xi_n.$$

By the CLT applied to ξ_n one can see that the limit distribution of $\sqrt{n}\xi_n$ is in fact $N_k(0, \frac{1}{4}\Sigma_\beta)$. Note that we have already faced with the shift φ in Corollary 11. In particular, (5.9) yields that $\langle q''_\infty(\theta_0), (\Delta\theta_n, \varphi)\rangle$ can be rewritten as

$$\int_0^\tau h_\beta^\top T(u)\varphi_\beta K(u) G_C(u)\,\mathrm{d}u = h_\beta^\top M\varphi_\beta.$$

Theorem 12 and Cramér-Wold's theorem yield that

$$h_\beta^\top M \xrightarrow{d} N_k(0, \Sigma_\beta).$$

Since M is invertible, the convergence (2.1) is proved.

Now, for a fixed shift φ_λ take such φ_β that

$$\mathbf{E}\Big[(h_\beta{}^\top X)(\varphi_\beta^\top X)\exp(\beta_0^\top X)\int_0^Y \lambda_0(u)\,\mathrm{d}u\Big] + \mathbf{E}\Big[(h_\beta{}^\top X)\exp(\beta_0^\top X)\int_0^Y \varphi_\lambda(u)\,\mathrm{d}u\Big]$$
$$= h_\beta^\top(A\varphi_\beta + m(\varphi_\lambda)) = 0.$$

Hence, $\varphi_\beta = -A^{-1}m(\varphi_\lambda)$. From (5.6) it follows that

$$\begin{aligned}
&-\langle q''_\infty(\theta_0), (\Delta\theta_n, \varphi)\rangle\\
&\quad= \mathbf{E}\Big[(\varphi_\beta{}^\top X)\exp(\beta_0^\top X)\int_0^Y h_\lambda(u)\,\mathrm{d}u\Big] + \mathbf{E}\Big[\frac{\Delta}{\lambda_0^2(Y)}h_\lambda(Y)\varphi_\lambda(Y)\Big]\\
&\quad= \int_0^\tau h_\lambda(u)\varphi_\beta^\top a(u)G_C(u)\,\mathrm{d}u + \int_0^\tau h_\lambda(u)\frac{\varphi_\lambda}{K(u)}G_C(u)\,\mathrm{d}u\\
&\quad= \int_0^\tau h_\lambda(u)\Big(-a(u)^\top\varphi_\beta + \frac{\varphi_\lambda}{K(u)}\Big)G_C(u)\,\mathrm{d}u.
\end{aligned}$$

In view of Theorem 12 and the remark at the beginning of the proof, in order to show the convergence (2.4), one should show that the equation (2.3) has a Lipschitz continuous solution φ_λ. But if φ_λ is a solution to (2.3) then

$$\varphi_\lambda(u) = K(u)f(u) + K(u)a^\top(u)C \tag{6.1}$$

for some constant $C \in \mathbb{R}^k$ and thus, is Lipschitz continuous. After substitution (6.1) in (2.3) we obtain

$$a^\top(u)\Big[C - \int_0^\tau (f(u) + a^\top(u)C)A^{-1}a(u)K(u)G_C(u)\,\mathrm{d}u\Big] = 0.$$

Let $S = \int_0^\tau f(u)A^{-1}a(u)K(u)G_C(u)\,\mathrm{d}u$ and $P(u) = \mathbf{E}[XX^\top\exp(\beta_0^\top X)G_T(u|X)]$. We show that it is possible to choose C so that

$$C - \int_0^\tau a^\top(u)CA^{-1}a(u)K(u)G_C(u)\,\mathrm{d}u = S.$$

After transposing both sides and multiplying by A, we have

$$C^\top\Big(A - \int_0^\tau a(u)a^\top(u)K(u)G_C(u)\,\mathrm{d}u\Big) = S^\top A.$$

Transformation of $R := A - \int_0^\tau a(u)a^\top(u)K(u)G_C(u)\,\mathrm{d}u$ leads to

$$R = \int_0^\tau \lambda_0(u)\biggl(P(u) - \frac{a(u)a^\top(u)}{b(u)}\biggr)G_C(u)\,\mathrm{d}u.$$

In the proof of Corollary 11 it was shown that $P(u) - \frac{a(u)a^\top(u)}{b(u)} = \frac{T(u)}{b(u)}$ is a positive definite matrix. Therefore, R is positive definite and invertible. Hence, (2.3) has a unique solution and convergence (2.4) holds. This completes the proof.

7 Conclusion

Here we studied properties of the Corrected MLE (λ_n, β_n) proposed by Kukush et al. [7] in Cox proportional hazards model with measurement error. Asymptotic normality was obtained for β_n and integral functionals of λ_n. We also present estimator $(\widehat{\lambda}_n, \widehat{\beta}_n)$ that inherits properties of (λ_n, β_n) and transforms the maximization problem to a parametric one.

In future we intend to provide simulations in this model.

References

[1] Andersen, P.K., Gill, R.D.: Cox's regression model for counting processes: a large sample study. Ann. Stat. **10**(4), 1100–1120 (1982). MR0673646

[2] Augustin, T.: An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. Scand. J. Stat. **31**, 43–50 (2004). MR2042597

[3] Dorogovtsev, A.Ya., Kukush, A.G.: Asymptotic properties of a nonparametric intensity estimator of a heterogeneous Poisson process. Cybern. Syst. Anal. **32**(1), 74–85 (1996). MR1409534

[4] Gu, M., Kong, F.H.: Consistent estimation in Cox proportional hazards model with covariate measurement errors. Stat. Sin. **32**(9), 953–969 (1999). MR1744820

[5] Hu, P., Tsiatis, A.A., Davidian, M.: Estimating the parameters in the Cox model when covariate variables are measured with error. Biometrics **54**(4), 1407–1419 (1998). MR1671598

[6] Kim, Y., Kim, B., Jang, W.: Asymptotic properties of the maximum likelihood estimator for the proportional hazards model with doubly censored data. J. Multivar. Anal. **101**(6), 1339–1351 (2010). MR2609496

[7] Kukush, A.G., Baran, S., Fazekas, I., Usoltseva, E.: Simultaneous estimation of baseline hazard rate and regression parameters in Cox proportional hazards model with measurement error. J. Stat. Res. **45**(2), 77–94 (2011). MR2934363

[8] Potapov, V.N.: ε-Entropy of compact sets in C and tabulation of continuous functions. Sib. Math. J. **38**, 759–771 (1997). MR1474919

[9] Pupashenko, D.S., Shklyar, S.V., Kukush, A.G.: Asymptotic properties of Corrected Score estimator in autoregressive model with measurement errors. Teor. Imovir. Mat. Stat. **89**, 156–166 (2013)

[10] Royston, P.: Estimating a smooth baseline hazard function for the Cox model. Research Report No. 314, Department of Statistical Science, University College, London (2011)

[11] Schoenberg, I.J.: Spline functions and the problem of graduation. Proc. Natl. Acad. Sci. USA **52**(4), 947–950 (1964). MR0167768

[12] Zinn, J.: A note on the central limit theorem in Banach spaces. Ann. Probab. **5**, 283–286 (1977). MR0443026