Confidence regions in Cox proportional hazards model with measurement errors and unbounded parameter set

Oksana Chernova*[1]

[1] *Taras Shevchenko National University of Kyiv, Ukraine*

Cox proportional hazards model with measurement errors in covariates is considered. It is the ubiquitous technique in biomedical data analysis. In Kukush et al. (2011) and Chimisov & Kukush (2014) asymptotic properties of a simultaneous estimator $(\lambda_n; \beta_n)$ for the baseline hazard rate $\lambda(\cdot)$ and the regression parameter β were studied, at that the parameter set $\Theta = \Theta_\lambda \times \Theta_\beta$ was assumed bounded.

In Kukush & Chernova (2017) we dealt with the simultaneous estimator $(\lambda_n; \beta_n)$ in the case, where the Θ_λ was unbounded from above and not separated away from 0. The estimator was constructed in two steps: first we derived a strongly consistent estimator and then modified it to provide its asymptotic normality.

In this talk, we construct the confidence region for β. We reach our goal in each of the three cases: (a) the measurement error is bounded, (b) it is normally distributed, or (c) it is a shifted Poisson random variable. The censor is assumed to have a continuous pdf. In future research we intend to elaborate a method for heavy tailed error distributions and construct the confidence interval for an integral functional of $\lambda(\cdot)$.

Keywords: asymptotic normality, confidence region, consistent estimator, Cox proportional hazards model, measurement errors, simultaneous estimation of baseline hazard rate and regression parameter.

1 Model formulation and estimation

Consider the Cox proportional hazards model (Cox, 1972), where a lifetime T has the following intensity function

$$\lambda(t|X; \lambda, \beta) = \lambda(t) \exp(\beta^T X), \quad t \geq 0.$$

*Corresponding author: chernovaoksan@gmail.com

A covariate X is a given random vector distributed in $\mathbb{R}^m$, β is a parameter belonging to $\Theta_\beta \subset \mathbb{R}^m$, and $\lambda(\cdot) \in \Theta_\lambda \subset C[0,\tau]$, $\tau > 0$, is a baseline hazard function.

Instead of lifetime T one can usually observe a censored lifetime $Y := \min\{T, C\}$ and the censorship indicator $\Delta := I_{\{T \le C\}}$. The censor C is distributed on $[0,\tau]$. Its survival function $G_C(u) = 1 - F_C(u)$ is unknown, while we know τ. The conditional pdf of T given X is

$$f_T(t|X,\lambda,\beta) = \lambda(t|X;\lambda,\beta)\exp\left(-\int_0^t \lambda(t|X;\lambda,\beta)ds\right).$$

We consider an additive error model, i.e., instead of X a surrogate variable

$$W = X + U$$

is observed, where a random error U has known moment generating function $M_U(z) := \mathsf{E}e^{z^T U}$. A couple (T, X), censor C, and measurement error U are stochastically independent.

Consider independent copies of the model $(X_i, T_i, C_i, Y_i, \Delta_i, U_i, W_i)$, $i = 1, ..., n$. Based on triples (Y_i, Δ_i, W_i), $i = 1, ..., n$, we estimate true parameters β_0 and $\lambda_0(t)$, $t \in [0,\tau]$. Our model is presented in Augustin (2004) where baseline hazard function is assumed to belong to a parametric space, while we consider $\lambda_0(\cdot)$ from a closed convex subset of $C[0,\tau]$. Following the idea from Augustin (2004), we use the objective function

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

with

$$q(Y, \Delta, W; \lambda, \beta) := \Delta \cdot (\log \lambda(Y) + \beta^T W) - \frac{\exp(\beta^T W)}{M_U(\beta)} \int_0^Y \lambda(u)du.$$

Introduce the basic assumptions.

(i) $\Theta_\lambda \subset C[0,\tau]$ is the following closed convex set of nonnegative functions

$$\Theta_\lambda := \{\ f : [0,\tau] \to \mathbb{R} \mid f(t) \ge 0, \forall t \in [0,\tau] \text{ and } |f(t) - f(s)| \le L|t-s|, \forall t, s \in [0,\tau]\ \},$$

where $L > 0$ is a fixed constant.

(ii) $\Theta_\beta \subset \mathbb{R}^m$ is a compact set.

(iii) $\mathsf{E}U = 0$ and for some constant $\epsilon > 0$,

$$\mathsf{E}e^{D\|U\|} < \infty, \text{ where } D := \max_{\beta\in\Theta_\beta} \|\beta\| + \epsilon.$$

(iv) $\mathsf{E}e^{D\|X\|} < \infty$, with D defined in (iii).

(v) $\mathsf{P}(C > \tau) = 0$ and for all $\epsilon > 0$, $\mathsf{P}(C > \tau - \epsilon) > 0$.

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

Denote

$$\Theta = \Theta_\lambda \times \Theta_\beta. \tag{1}$$

(vii) True parameters (λ_0, β_0) belong to Θ, which is given in (1), and moreover $\lambda_0(t) > 0$, $t \in [0, \tau]$.

(viii) β_0 is an interior point of Θ_β.

(ix) $\lambda_0 \in \Theta_\lambda^\epsilon$ for some $\epsilon > 0$, where

$$\Theta_\lambda^\epsilon := \{\ f : [0, \tau] \to \mathbb{R} \mid f(t) \geq \epsilon,\ \forall t \in [0, \tau],$$
$$|f(t) - f(s)| \leq (L - \epsilon)|t - s|, \forall t, s \in [0, \tau]\ \}.$$

(x) $\mathsf{P}(C > 0) = 1$.

(xi) For some $\epsilon > 0$, $\mathsf{E}e^{2D\|U\|} < \infty$ where D is defined in (iii).

(xii) $\mathsf{E}e^{2D\|X\|} < \infty$ where D is defined in (iii).

Definition 1. Fix a sequence $\{\varepsilon_n\}$ of positive numbers, with $\varepsilon_n \downarrow 0$, as $n \to \infty$. The corrected estimator $\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right)$ of (λ, β) is a Borel measurable function of observations (Y_i, Δ_i, W_i), $i = 1, ..., n$, with values in Θ and such that

$$Q_n^{cor}\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right) \geq \sup_{(\lambda,\beta)\in\Theta} Q_n^{cor}(\lambda, \beta) - \varepsilon_n.$$

Theorem 3 from Kukush & Chernova (2017) proves that under conditions (i) to (vii), the couple $\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right)$ is a strongly consistent estimator of the true parameters (λ_0, β_0).

Based on $\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right)$, we derive a modified estimator $\left(\hat\lambda_n^{(2)}, \hat\beta_n^{(2)}\right)$ to be consistent and asymptotically normal.

Definition 2. The modified corrected estimator $\left(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)}\right)$ of (λ, β) is a Borel measurable function of observations (Y_i, Δ_i, W_i), $i = 1, ..., n$, with values in Θ and such that

$$\left(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)}\right) = \begin{cases} \arg\max \left\{ Q_n^{cor}(\lambda, \beta) \mid (\lambda, \beta) \in \Theta,\ \mu_\lambda \geq \frac{1}{2}\mu_{\hat{\lambda}_n^{(1)}} \right\}, & \text{if } \mu_{\hat{\lambda}_n^{(1)}} > 0, \\ \left(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)}\right), & \text{otherwise,} \end{cases}$$

with $\mu_\lambda := \min\limits_{t \in [0,\tau]} \lambda(t)$.

Introduce notations:

$$a(t) = \mathsf{E}[X e^{\beta_0^T X} G_T(t|X)], \quad b(t) = \mathsf{E}[e^{\beta_0^T X} G_T(t|X)], \quad \Lambda(t) = \int_0^t \lambda_0(t) dt,$$

$$p(t) = \mathsf{E}[X X^T e^{\beta_0^T X} G_T(t|X)], \quad T(t) = p(t)b(t) - a(t)a^T(t), \quad K(t) = \frac{\lambda_0(t)}{b(t)},$$

$$M = \int_0^\tau T(u) K(u) G_c(u) du.$$

For $i = 1, 2, \ldots$, introduce random variables

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

with

$$\frac{\partial q}{\partial \beta}(Y, \Delta, W; \lambda, \beta) = \Delta \cdot W - \frac{M_U(\beta) W - \mathsf{E}(U e^{\beta^T U})}{M_U(\beta)^2} \exp(\beta^T W) \int_0^Y \lambda(u) du.$$

Let

$$\Sigma_\beta = 4 \cdot \mathsf{Cov}(\zeta_1).$$

The following theorem from Kukush & Chernova (2017) states asymptotic normality of $\hat{\beta}_n^{(2)}$.

Theorem 1. *Assume conditions (i), (ii), and (v) – (xii). Then M is nonsingular and*

$$\sqrt{n}(\hat{\beta}_n^{(2)} - \beta_0) \xrightarrow{d} N_m(0, M^{-1} \Sigma_\beta M^{-1}). \tag{2}$$

2 Confidence region for the regression parameter

Based on Theorem 1, we construct a confidence region for the regression parameter. For $t \in [0, \tau]$ and $\beta \in \Theta_\beta$, denote

$$B(W, t; \lambda, \beta) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k! M_U((k+1)\beta)} \Lambda^k(t) e^{(k+1)\beta^T W},$$

$$A(W, t; \lambda, \beta) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k! M_U((k+1)\beta)} \Lambda^k(t) \left[W - \frac{\mathsf{E}[U e^{(k+1)\beta^T U}]}{M_U((k+1)\beta)} \right] e^{(k+1)\beta^T W},$$

$$P(W, t; \lambda, \beta) = \sum_{k=0}^{\infty} \frac{(-1)^k \Lambda^k(t)}{k! M_U((k+1)\beta)} \left[W W^T e^{(k+1)\beta^T W} - 2 \frac{\mathsf{E}[U e^{(k+1)\beta^T U}]}{M_U((k+1)\beta)} W^T e^{(k+1)\beta^T W} - \right.$$
$$\left. - \left(\frac{\mathsf{E}[U U^T e^{(k+1)\beta^T U}]}{M_U((k+1)\beta)} - 2 \frac{\mathsf{E}[U e^{(k+1)\beta^T U}] \mathsf{E}[U^T e^{(k+1)\beta^T U}]}{M_U^2((k+1)\beta)} \right) e^{(k+1)\beta^T W} \right].$$

Theorem 2. *Suppose that*

$$\sum_{k=0}^{\infty} \frac{\tilde{c}_{k+1}(\beta)}{k!} e^{k\beta^T z} < \infty, \qquad z \in \mathbb{R}^m, \qquad \beta \in \Theta_\beta,$$

where $\tilde{c}_{k+1}(\beta)$ is substituted with each of the following expressions

$$\frac{\mathsf{E}[||U|| e^{(k+1)\beta^T U}]}{M_U((k+1)\beta)}, \quad \frac{\mathsf{E}[||U||^2 e^{(k+1)\beta^T U}]}{M_U((k+1)\beta)}, \quad \left(\frac{\mathsf{E}[||U|| e^{(k+1)\beta^T U}]}{M_U((k+1)\beta)} \right)^2.$$

Then for all $t \in [0, \tau]$,

$$\hat{b}(t) = \frac{1}{n} \sum_{i=1}^{n} B(W_i, t; \hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)}),$$

$$\hat{a}(t) = \frac{1}{n} \sum_{i=1}^{n} A(W_i, t; \hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)}),$$

$$\hat{p}(t) = \frac{1}{n} \sum_{i=1}^{n} P(W_i, t; \hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)})$$

are consistent estimators of $b(t)$, $a(t)$ and $p(t)$, respectively.

We point out that the conditions of Theorem 2 are fulfilled in the next cases: (a) U is bounded, (b) it is normally distributed, or (c) it is a shifted Poisson random variable.

Denote

$$\hat{M} = \int_0^{Y_{(n)}} \hat{T}(u)\hat{K}(u)\hat{G}_C(u)du, \tag{3}$$

where $\hat{G}_C$ is the Kaplan-Meier estimator of the survival function of censor C, and

$$\hat{\Sigma}_\beta = \frac{4}{n-1}\sum_{i=1}^{n} \hat{\zeta}_i\hat{\zeta}_i^T, \quad \text{with} \tag{4}$$

$$\hat{\zeta}_i := -\frac{\Delta_i \cdot \hat{a}(Y_i)}{\hat{b}(Y_i)} + \frac{\exp(\hat{\beta}_n^{(2)T} W_i)}{M_U(\hat{\beta}_n^{(2)})} \int_0^{Y_i} \hat{a}(u)\hat{K}(u)du + \frac{\partial q}{\partial \beta}(Y_i, \Delta_i, W_i; \hat{\beta}_n^{(2)}, \hat{\lambda}_n^{(2)}).$$

Theorem 3. *Assume the conditions of Theorem 2. The estimators $\hat{M}$ and $\hat{\Sigma}_\beta$ defined in (3) and (4) are consistent estimators of matrices M and Σ_β, respectively.*

Given a confidence probability $1-\alpha$, $0<\alpha<1/2$, let $\left(\chi^2_m\right)_\alpha$ be the upper α-quantile of the χ^2_m distribution. Based on Theorem 2 and 3, we take as an asymptotic confidence ellipsoid for $\hat{\beta}_n^{(2)}$ the set

$$E_n := \left\{ z \in \mathbb{R}^m \mid \left(z - \hat{\beta}_n^{(2)}\right)^T \left(\hat{M}^{-1}\hat{\Sigma}_\beta\hat{M}^{-1}\right)^{-1} \left(z - \hat{\beta}_n^{(2)}\right) \le \frac{1}{n}\left(\chi^2_m\right)_\alpha \right\}.$$

It holds $\mathsf{P}(\beta_0 \in E_n) \to 1-\alpha$ as $n \to \infty$.

Acknowledgements: I would like to express my sincere gratitude to my advisor Prof. Alexander Kukush for the support of my PhD research.

References

[1] Augustin, T. (2004). An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scandinavian Journal of Statistics* **31**, 1, 43–50.

[2] Chimisov, C., and Kukush, A. (2014). Asymptotic normality of corrected estimator in Cox proportional hazards model with measurement error. *Modern Stochastics: Theory and Applications* **1**, 1, 13–32.

[3] Kukush, A., Baran, S., Fazekas, I., and Usoltseva, E. (2011). Simultaneous estimation of baseline hazard rate and regression parameters in Cox proportional hazards model with measurement error. *Journal of Statistical Research* **45**, 2, 77–94.

[4] Kukush, A., and Chernova, O. (2017). Consistent estimation in Cox proportional hazards model with measurement errors and unbounded parameter set. *Theory of Probability and Mathematical Statistics* **96**, 100–109.