Теорія Ймовір. та Матем. Статист.
Вип. 96, 2017

Theor. Probability and Math. Statist.
No. 96, 2018, Pages 101–110
https://doi.org/10.1090/tpms/1036
Article electronically published on October 5, 2018

CONSISTENT ESTIMATION IN COX PROPORTIONAL HAZARDS MODEL WITH MEASUREMENT ERRORS AND UNBOUNDED PARAMETER SET

UDC 519.21

A. G. KUKUSH AND O. O. CHERNOVA

Abstract. We study the Cox problem with proportional risks and measurement errors. The asymptotic properties of the simultaneous estimator $\lambda_n(\cdot)$, β_n of the baseline hazard function $\lambda(\cdot)$ and regression parameter β are considered in the papers [**6**] and [**3**] for the case of a bounded set of parameters $\Theta = \Theta_\lambda \times \Theta_\beta$. In the current paper, the set Θ_λ is unbounded from above and is not separated from zero. The estimator is constructed in the following two steps. First, one obtains a strictly consistent estimator and, second, this estimator is corrected in order to obtain an asymptotically normal estimator.

1. Introduction

We consider the Cox proportional hazards model [4]. According to this model, the intensity function T of the lifetime is of the following form:

$$\lambda(t|X;\lambda,\beta) = \lambda(t)\exp\left(\beta^T X\right), \qquad t \ge 0. \tag{1}$$

Here the regressor X is a random vector in $\mathbb{R}^m$, β is the regression parameter in the set $\Theta_\beta \subset \mathbb{R}^m$, and $\lambda(\cdot) \in \Theta_\lambda \subset C[0,\tau]$ is the baseline hazard function.

Observed are the censored data rather than the duration of life T; namely one observes the random variables $Y := \min\{T, C\}$ and censorship indicator $\Delta := I_{\{T \le C\}}$. The censor C is random and distributed in the interval $[0,\tau]$. The survival function of the censor $G_C(u) = 1 - F_C(u)$ is unknown, while τ is known. The conditional density of T given X is given by

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

The variable $W = X + U$ is observed instead of X, where the moment generating function $M_U(\beta) := \mathsf{E}\, e^{\beta^T U}$ of the random measurement error U is known. The pair (T, X), censor C, and error U are stochastically independent.

Consider independent copies of the model $(X_i, T_i, C_i, Y_i, \Delta_i, U_i, W_i)$, $i = 1,\dots,n$. Based on the observations (Y_i, Δ_i, W_i), $i = 1,\dots,n$, our aim is to estimate the true parameters β_0 and $\lambda_0(t)$, $t \in [0,\tau]$.

Following the paper [2], we use the corrected likelihood function

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

2010 *Mathematics Subject Classification.* Primary 62N02; Secondary 62N01.

Key words and phrases. Asymptotically normal estimator, consistent estimator, Cox proportional hazards model, simultaneous estimation of the baseline hazard function and regression parameter.

where

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

The corrected estimator is given by

$$\left(\hat\lambda_n,\hat\beta_n\right) = \underset{(\lambda,\beta)\in\Theta}{\arg\max}\, Q_n^{\mathrm{cor}}(\lambda,\beta), \tag{2}$$

where $\Theta := \Theta_\lambda \times \Theta_\beta$. If the sets of parameters are compact, then Θ is compact as well, and the maximum on the right hand side of equality (2) is attained at a point of Θ.

Many papers are devoted to the estimation of β_0 and cumulative risks

$$\Lambda(t) = \int_0^t \lambda(t|X;\lambda,\beta)\,dt.$$

In particular, Andersen and Gill [1] introduce general ideas of estimation by using the partial likelihood function. A model with measurement errors is considered in [5], where consistent and asymptotically normal estimators of β_0 and $\Lambda(t)$ are obtained with the help of the corrected score method. Royston in [8] describes the problems where the estimators of the risk function $\lambda(\cdot)$ rather than those of the cumulative risk $\Lambda(t)$ play a crucial role.

The model studied in the current paper is introduced in [2]. However, the risk function belongs to a finite set of parameters in [2], while we consider the risk function in $C[0,\tau]$.

The consistency of estimator (2) is proved in [6] for the case of a bounded set of parameters. An asymptotically normal estimator is obtained in [3]. Note that the set Θ_λ is defined in [6] without an assumption that the values $\lambda(0)$ are bounded, while the proof in [6] uses this property. We weaken this strong assumption as well as the assumption that $\lambda(\cdot)$ is separated from zero.

The paper is organized as follows. The estimator is introduced in Section 2 under the assumption that the set of parameters is unbounded. The consistency of the estimator is also proved, and the procedure for calculating the estimator is described in Section 2. The estimator considered in Section 2 is modified in Section 3 in such a way that the resulting estimator is asymptotically consistent. Some concluding remarks are given in Section 4.

2. Step 1: Consistent estimation

Below are the assumptions imposed on the set of parameters.

(i) $K_\lambda \subset C[0,\tau]$ is a closed convex set of non-negative functions,

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

where $L>0$ is a fixed constant.

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

The following assumptions (iii)–(vi) are introduced in the paper [6].

(iii) $\mathsf{E}\,U = 0$ and

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

for some $\varepsilon > 0$.

(iv) $\mathsf{E}\,e^{D\|X\|} < \infty$, where the number D is defined in assumption (iii).

(v) τ is the right end point of the distribution of the censor C, that is, a number such that $\mathsf{P}(C>\tau)=0$ and $\mathsf{P}(C>\tau-\varepsilon)>0$ for all $\varepsilon>0$.

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

Let

$$K = K_\lambda \times \Theta_\beta. \tag{3}$$

If $\lambda(Y) = 0$, then we put $\ln \lambda(Y) = -\infty$. For the same case of $\lambda(Y) = 0$, we also put

$$\Delta \cdot \ln \lambda(Y) = \begin{cases} 0, & \text{if } \Delta = 0, \\ -\infty, & \text{if } \Delta = 1. \end{cases}$$

Definition 1. Let $\{\varepsilon_n\}$ be a fixed sequence of positive numbers such that $\varepsilon_n \downarrow 0$, $n \to \infty$. The modified estimator $\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right)$ for (λ, β) is defined as a Borel function of observations (Y_i, Δ_i, W_i), $i = 1, \dots, n$, assuming values in K and such that

$$Q_n^{\mathrm{cor}}\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right) \geq \sup_{(\lambda,\beta)\in K} Q_n^{\mathrm{cor}}(\lambda, \beta) - \varepsilon_n. \tag{4}$$

Results of the paper [7] guarantee that such an estimator exists. A crucial property here is that the upper bound in (4) is finite.

Below is an additional assumption.

(vii) The true values of parameters (λ_0, β_0) belong to the set K defined in (3), and moreover

$$\lambda_0(t) > 0, \qquad t \in [0, \tau].$$

Definition 2. Let $A_n = A_n(\omega)$, $n \geq 1$, be a sequence of statements that depend on an elementary random event $\omega \in \Omega$. We say that statements A_n hold *eventually* if, for almost all $\omega \in \Omega$, there exists a positive integer number $n_0 = n_0(\omega)$ such that statement $A_n(\omega)$ holds for all $n \geq n_0(\omega)$.

Theorem 3. *Assume that assumptions* (i)–(vii) *hold. Then* $\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right)$ *is a strongly consistent estimator of the true parameters* (λ_0, β_0)*; that is,*

$$\max_{t\in[0,\tau]} \left|\hat\lambda_n^{(1)}(t) - \lambda_0(t)\right| \to 0, \qquad \hat\beta_n^{(1)} \to \beta_0$$

almost surely as $n \to \infty$.

Proof. For $R > 0$, denote

$$K_\lambda^R = K_\lambda \cap \bar B(0, R), \qquad K^R = K_\lambda^R \times \Theta_\beta,$$

where $\bar B(0, R)$ is the closed ball in $C[0, \tau]$ of radius R centered at the origin.

1. In the first part of the proof, we show that

$$\sup_{(\lambda,\beta)\in K^R} Q_n^{\mathrm{cor}}(\lambda, \beta) > \sup_{(\lambda,\beta)\in K\setminus K^R} Q_n^{\mathrm{cor}}(\lambda, \beta) \tag{5}$$

eventually for sufficiently large non-random numbers $R > \|\lambda_0\|$.

The Lipschitz condition for $\lambda \in K_\lambda$ implies that

$$\lambda(0) - L\tau \leq \lambda(t) \leq \lambda(0) + L\tau, \tag{6}$$

whence

$$q(Y_i, \Delta_i, W_i; \lambda, \beta) \leq \Delta_i \left(\ln(\lambda(0) + L\tau) + \beta^T W_i\right) - \frac{\exp\left(\beta^T W_i\right) Y_i}{M_U(\beta)}(\lambda(0) - L\tau).$$

Using the Lipschitz condition for $\lambda \in K_\lambda$, one can show that $\lambda(t) > R - L\tau$ for all $t \in [0, \tau]$ if $\lambda(t_1) > R$ for some $t_1 \in [0, \tau]$. On the other hand, $\lambda(0) > R$ implies that $\lambda(t) > R - L\tau$, $t \in [0, \tau]$. Thus the supremum on the right hand side of inequality (5) can be considered in the set $\{\lambda \in K_\lambda \colon \lambda(0) > R\} \times \Theta_\beta$.

Put

$$D_1 = \max_{\beta\in\Theta_\beta} \|\beta\|.$$

We have

$$\sup_{(\lambda,\beta)\in K\setminus K^R} Q_n^{\mathrm{cor}}(\lambda,\beta) \le I_1 + \sup_{\substack{\lambda\in K_\lambda:\\ \lambda(0)>R}} I_2 + I_3,$$

where

$$I_1 = -(R-L\tau)\frac{1}{n}\sum_{i:\Delta_i=0}\frac{\exp(-D_1\|W_i\|)Y_i}{\max\limits_{\beta\in\Theta_\beta} M_U(\beta)},$$

$$I_2 = \ln(\lambda(0)+L\tau)\frac{1}{n}\sum_{i:\Delta_i=1}\Delta_i - (\lambda(0)+L\tau)\frac{1}{n}\sum_{i:\Delta_i=1}\frac{\exp(-D_1\|W_i\|)Y_i}{\max\limits_{\beta\in\Theta_\beta} M_U(\beta)},$$

$$I_3 = \frac{1}{n}\sum_{i:\Delta_i=1} D_1\|W_i\| + 2L\tau\frac{1}{n}\sum_{i:\Delta_i=1}\frac{\exp(-D_1\|W_i\|)Y_i}{\max\limits_{\beta\in\Theta_\beta} M_U(\beta)}.$$

The strong law of large numbers yields

$$I_1 \to -(R-L\tau)\frac{\mathsf{E}\big[C\cdot I(\Delta=0)\exp\left(-D_1\|W\|\right)\big]}{\max\limits_{\beta\in\Theta_\beta} M_U(\beta)}$$

almost surely as $n\to\infty$. This means that

$$I_1 \le -(R-L\tau)D_2$$

eventually where $D_2>0$.

Let

$$A_n = \frac{1}{n}\sum_{i=1}^n \Delta_i, \qquad B_n = \frac{1}{n}\sum_{i=1}^n \frac{\exp(-D_1\|W_i\|)Y_i}{\max\limits_{\beta\in\Theta_\beta} M_U(\beta)} 1_{\{\Delta_i=1\}}.$$

Since $A_n>0$ and $B_n>0$ *eventually*, we obtain

$$I_2 \le \max_{z>0}(A_n\ln z - zB_n) = A_n\left(\ln\left(\frac{A_n}{B_n}\right)-1\right)$$

for $\lambda(0)>R$. By the strong law of large numbers,

$$A_n \to \mathsf{P}(\Delta=1)>0, \qquad B_n \to \frac{\mathsf{E}\big[\,T\cdot I(\Delta=1)\exp\left(-D_1\|W\|\right)\big]}{\max\limits_{\beta\in\Theta_\beta} M_U(\beta)} > 0$$

almost surely as $n\to\infty$. Hence I_2 is *eventually* bounded from above by some positive constant D_3.

Further, it follows from the strong law of large numbers that I_3 is *eventually* bounded from above by some positive constant D_4. Hence

$$\varlimsup_{n\to\infty}\sup_{(\lambda,\beta)\in K\setminus K^R} Q_n^{\mathrm{cor}}(\lambda,\beta) \le -(R-L\tau)D_2 + D_3 + D_4.$$

Note that the constants D_2, D_3, and D_4 introduced above do not depend on $\beta\in\Theta_\beta$.

Letting $R\to+\infty$, we get

$$\varlimsup_{n\to\infty}\sup_{(\lambda,\beta)\in K\setminus K^R} Q_n^{\mathrm{cor}}(\lambda,\beta) \to -\infty, \qquad R\to+\infty.$$

This proves that inequality (5) holds *eventually* for sufficiently large R. In particular, one can substitute K^R for K in Definition 1.

Therefore, we can assume that

$$Q_n^{\mathrm{cor}}\left(\hat\lambda_n^{(1)},\hat\beta_n^{(1)}\right) \ge \sup_{(\lambda,\beta)\in K^R} Q_n^{\mathrm{cor}}(\lambda,\beta) - \varepsilon_n \tag{7}$$

and $\left(\hat\lambda_n^{(1)},\hat\beta_n^{(1)}\right)\in K^R$ for all $n\ge 1$. Note that K^R is a compact set in $C[0,\tau]$.

2. Since $R > \|\lambda_0\|$, we have $(\lambda_0, \beta_0) \in K^R$. Then the inequality

$$Q_n^{\mathrm{cor}}\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right) \ge Q_n^{\mathrm{cor}}(\lambda_0, \beta_0) - \varepsilon_n \tag{8}$$

follows from (7). Fix $\omega \in A \subset \Omega$, where $\mathsf{P}(A) = 1$. In what follows we impose on A some extra assumption. Our aim is to show that

$$\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right) \to (\lambda_0, \beta_0)$$

at a point ω. We have

$$Q_n^{\mathrm{cor}}(\lambda_0, \beta_0) \to q_\infty(\lambda_0, \beta_0) := \mathsf{E}[q(Y, \Delta, W; \lambda_0, \beta_0)]. \tag{9}$$

This relation holds almost surely, and we can assume that (9) holds for a fixed ω. This means that the first extra condition on A is

$$Q_n^{\mathrm{cor}}(\lambda_0, \beta_0; \omega) \to q_\infty(\lambda_0, \beta_0), \qquad \omega \in A.$$

The sequence $\{(\hat\lambda_n^{(1)}(\omega), \hat\beta_n^{(1)}(\omega)), n \ge 1\}$ belongs to the compact set K^R. Consider an arbitrary convergent subsequence

$$\left(\hat\lambda_{n'}^{(1)}(\omega), \hat\beta_{n'}^{(1)}(\omega)\right) \to (\lambda_*, \beta_*) \in K^R. \tag{10}$$

It follows from (8) and (9) that

$$\begin{aligned} q_\infty(\lambda_0, \beta_0) &\le \varliminf_{n' \to \infty} Q_{n'}^{\mathrm{cor}}\left(\hat\lambda_{n'}^{(1)}, \hat\beta_{n'}^{(1)}\right) \\ &= \varliminf_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \ln \hat\lambda_{n'}^{(1)}(Y_i) \\ &\quad + \lim_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} \left(\Delta_i \cdot \hat\beta_{n'}^{(1)T} W_i - \frac{\exp(\hat\beta_{n'}^{(1)T} W_i)}{M_U(\hat\beta_{n'}^{(1)})} \int_0^{Y_i} \hat\lambda_{n'}^{(1)}(u)\, du \right). \end{aligned}$$

The second extra condition which is imposed on A reads as follows: for every $\omega \in A$, the sequence of random functions

$$\frac{1}{n} \sum_{i=1}^{n} \left(\Delta_i \beta^T W_i - \frac{\exp(\beta^T W_i)}{M_U(\beta)} \int_0^{Y_i} \lambda(u)\, du \right)$$

converges uniformly in $(\lambda, \beta) \in K^R$ to

$$\mathsf{E}\left[\Delta \beta^T W - \frac{\exp\left(\beta^T W\right)}{M_U(\beta)} \int_0^{Y} \lambda(u)\, du \right] =: q_\infty^2(\lambda, \beta).$$

This condition does not restrict the generality, since, for fixed $(\lambda, \beta) \in K^R$,

(1) the above sequence converges almost surely to q_∞^2,
(2) this sequence is equicontinuous almost surely in the compact set K^R,
(3) the limit function is continuous in K^R.

These three properties imply that the above sequence converges almost surely to q_∞^2 in the set K^R.

Note that the function q_∞^2 is continuous with respect to $(\lambda, \beta) \in K^R$, whence

$$q_\infty(\lambda_0, \beta_0) \le \varliminf_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \ln \hat\lambda_{n'}^{(1)}(Y_i) + q_\infty^2(\lambda_*, \beta_*).$$

For sufficiently large n',

$$\hat{\lambda}_{n'}^{(1)}(t) \le \lambda_*(t) + \varepsilon, \qquad t \in [0, \tau],$$

where $\varepsilon > 0$ is fixed.

In what follows we assume that

$$\frac{1}{n} \sum_{i=1}^{n} \Delta_i \ln \lambda(Y_i) \to \mathsf{E}[\Delta \lambda(Y)]$$

uniformly with respect to $(\lambda, \beta) \in \left(K_\lambda^{R+\delta_k} \cap \{\lambda\colon \lambda(t) \ge \delta_k\}\right) \times \Theta_\beta$ for all $k \ge 1$ and $\omega \in A$, where $\delta_k \downarrow 0$ and $\{\delta_k\}$ is a fixed sequence of positive numbers.

Then the strong law of large numbers implies that

$$\begin{aligned} \varliminf_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \ln \hat{\lambda}_{n'}^{(1)}(Y_i) &\le \varliminf_{n' \to \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \ln(\lambda_*(Y_i) + \varepsilon) \\ &= \mathsf{E}[\Delta \cdot \ln(\lambda_*(Y) + \varepsilon)] =: q_\infty^{1,\varepsilon}(\lambda_*), \end{aligned}$$

whence

$$q_\infty(\lambda_0, \beta_0) \le q_\infty^{1,\varepsilon}(\lambda_*) + q_\infty^2(\lambda_*, \beta_*)$$

for all $\varepsilon > 0$. Letting $\varepsilon \to 0$ we obtain

$$q_\infty^{1,\varepsilon}(\lambda_*) = \mathsf{E}\left[\Delta \cdot \ln(\lambda_*(Y) + \varepsilon) I\left(\lambda_*(Y) > \tfrac{1}{2}\right)\right] + \mathsf{E}\left[\Delta \cdot \ln(\lambda_*(Y) + \varepsilon) I\left(\lambda_*(Y) \le \tfrac{1}{2}\right)\right].$$

The first expectation converges to

$$\mathsf{E}\left[\Delta \cdot \ln(\lambda_*(Y)) I\left(\lambda_*(Y) > \tfrac{1}{2}\right)\right]$$

as $\varepsilon \to 0$ by the Lebesgue dominated convergence theorem. Accordingly, the second expectation converges to

$$\mathsf{E}\left[\Delta \cdot \ln(\lambda_*(Y)) I\left(\lambda_*(Y) \le \tfrac{1}{2}\right)\right]$$

as $\varepsilon \to 0$ by the Lebesgues monotone convergence theorem. Then

$$q_\infty^{1,\varepsilon}(\lambda_*) \to q_\infty^1(\lambda_*) := \mathsf{E}[\Delta \cdot \ln \lambda_*(Y)]$$

as $\varepsilon \to 0$. Thus

$$q_\infty(\lambda_0, \beta_0) \le q_\infty^1(\lambda_*) + q_\infty^2(\lambda_*, \beta_*) = q_\infty(\lambda_*, \beta_*).$$

According to [6],

$$q_\infty(\lambda_0, \beta_0) \ge q_\infty(\lambda_*, \beta_*).$$

Moreover, the inequality becomes an equality if and only if $\lambda_* = \lambda_0$ and $\beta_* = \beta_0$. Hence subsequence (10) converges to (λ_0, β_0). Since the whole sequence belongs to a compact set, we conclude that

$$\left(\hat{\lambda}_n^{(1)}(\omega), \hat{\beta}_n^{(1)}(\omega)\right) \to (\lambda_0, \beta_0), \qquad n \to \infty.$$

This relation holds for almost all $\omega \in \Omega$. The strong consistency is proved. □

Our next aim is to explain the procedure for calculating such an estimator. Similarly to [3] we prove that the function $\hat{\lambda}_n^{(1)}$ that minimizes Q_n^{cor} is a linear spline for a fixed $\beta \in \Theta_\beta$.

Theorem 4. *Assumptions* (i) *and* (ii) *imply that the function* $\hat{\lambda}_n^{(1)}$ *that minimizes* Q_n^{cor} *is a linear spline.*

Proof. Let $(Y_{i_1},\dots,Y_{i_n})$ denote the order statistics constructed from the observations $Y_1,\dots,Y_n$. Fix $\beta\in\Theta_\beta$ and assume that $\hat{\lambda}_n^{(1)}\in\Theta_\lambda$ maximizes $Q_n^{\mathrm{cor}}(\cdot,\beta)$. Along with $\big(\hat{\lambda}_n^{(1)},\beta\big)$ consider $\big(\bar{\lambda}_n,\beta\big)$, where $\bar{\lambda}_n$ is the function constructed below. Put

$$\bar{\lambda}_n(Y_{i_k})=\hat{\lambda}_n^{(1)}(Y_{i_k}),\qquad k=1,\dots,n.$$

On each interval $[Y_{i_k},Y_{i_{k+1}}]$, $k=1,\dots,n-1$, consider the segments of the two straight lines

$$L^1_{i_k}(t)=\hat{\lambda}_n^{(1)}(Y_{i_k})+L(Y_{i_k}-t),$$
$$L^2_{i_k}(t)=\hat{\lambda}_n^{(1)}\left(Y_{i_{k+1}}\right)+L\left(t-Y_{i_{k+1}}\right),$$

where L is defined in (i). By B_{i_k}, we denote the point of intersection of $L^1_{i_k}(t)$ and $L^2_{i_k}(t)$. We also agree that $B_{i_0}:=0$, $B_{i_n}:=\tau$ and $Y_{i_0}:=0$, $Y_{i_{n+1}}:=\tau$. Then the function $\bar{\lambda}_n(t)$ is constructed as follows:

$$\bar{\lambda}_n(t)=\begin{cases}L^2_{i_0}(t), & \text{if } t\in[0,Y_{i_1}],\\ \max\{L^1_{i_k}(t),0\}, & \text{if } t\in[Y_{i_k},B_{i_k}],\ k=1,\dots,n-1,\\ \max\{L^2_{i_k}(t),0\}, & \text{if } t\in[B_{i_k},Y_{i_{k+1}}],\ k=1,\dots,n-1,\\ L^1_{i_n}(t), & \text{if } t\in[Y_{i_n},\tau].\end{cases}\tag{11}$$

It is easy to see that $\bar{\lambda}_n\in\Theta_\lambda$. By construction, $\hat{\lambda}_n^{(1)}\geq\bar{\lambda}_n$. Thus

$$Q_n^{\mathrm{cor}}\left(\hat{\lambda}_n^{(1)},\beta\right)\leq Q_n^{\mathrm{cor}}\left(\bar{\lambda}_n,\beta\right),$$

whence $\hat{\lambda}_n^{(1)}=\bar{\lambda}_n$. This completes the proof. □

Note that $\bar{\lambda}_n(B_{i_k})>0$ *eventually*, whence we conclude that the minimum in (11) is not needed anymore.

Having constructed linear spline

$$\bar{\lambda}_n(\beta)=\operatorname*{arg\,max}_{\lambda:\,(\lambda,\beta)\in\Theta}Q_n^{\mathrm{cor}},$$

we minimize $Q(\beta):=Q_n^{\mathrm{cor}}\left(\bar{\lambda}_n(\beta),\beta\right)$ with respect to $\beta\in\Theta_\beta$. In other words, we search for $\hat{\beta}\in\Theta_\beta$ such that

$$Q\left(\hat{\beta}\right)\geq\sup_{\beta\in\Theta_\beta}Q(\beta)-\varepsilon_n.$$

Since $Q(\beta)$ is bounded, such a number $\hat{\beta}$ exists.

Now

$$Q_n^{\mathrm{cor}}\left(\bar{\lambda}_n(\hat{\beta}),\hat{\beta}\right)\geq\sup_{\beta\in\Theta_\beta}\max_{\lambda\in\Theta_\lambda}Q(\beta)-\varepsilon_n=\sup_{(\lambda,\beta)\in\Theta}Q_n^{\mathrm{cor}}(\lambda,\beta)-\varepsilon_n.$$

Hence the estimator $\big(\bar{\lambda}_n(\hat{\beta}),\hat{\beta}\big)$ satisfies conditions of Definition 1, and its evaluation is a parametric problem.

3. Step 2: Construction of an asymptotically normal estimator

Our aim in this section is to modify the estimator $\big(\hat{\lambda}_n^{(1)}(\omega),\hat{\beta}_n^{(1)}(\omega)\big)$ constructed in Definition 1 in order to obtain an asymptotically normal estimator.

Definition 5. A modified estimator $\left(\hat\lambda_n^{(2)}, \hat\beta_n^{(2)}\right)$ for (λ, β) is defined as a Borel function of observations (Y_i, Δ_i, W_i), $i = 1, \dots, n$, assuming values in K and such that

$$\left(\hat\lambda_n^{(2)}, \hat\beta_n^{(2)}\right) = \begin{cases} \arg\max\left\{Q_n^{\mathrm{cor}}(\lambda,\beta) \,\middle|\, (\lambda,\beta)\in K,\ \mu_\lambda \ge \frac{1}{2}\mu_{\hat\lambda_n^{(1)}}\right\}, & \text{if } \mu_{\hat\lambda_n^{(1)}} > 0,\\ \left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right), & \text{if } \mu_{\hat\lambda_n^{(1)}} \le 0,\end{cases}$$

where $\mu_\lambda := \min_{t\in[0,\tau]}\lambda(t)$.

The existence of such an estimator follows from some results of the paper [7].

According to Theorem 3, $\mu_{\hat\lambda_n^{(1)}} \to \mu_{\lambda_0} > 0$ almost surely and hence

$$\begin{aligned} K_1 := \left\{(\lambda,\beta)\in K \middle| \mu_\lambda \ge \tfrac{3}{4}\mu_{\lambda_0}\right\} &\subset \left\{(\lambda,\beta)\in K \middle| \mu_\lambda \ge \tfrac{1}{2}\mu_{\hat\lambda_n^{(1)}}\right\}\\ &\subset \left\{(\lambda,\beta)\in K \middle| \mu_\lambda \ge \tfrac{1}{4}\mu_{\lambda_0}\right\} =: K_2 \end{aligned}$$

eventually. The estimator

$$\left(\hat\lambda_n^{(3)}, \hat\beta_n^{(3)}\right) = \arg\max_{(\lambda,\beta)\in K_2} Q_n^{\mathrm{cor}}(\lambda,\beta) \tag{12}$$

is strongly consistent under the assumptions (i)–(vii), since, in view of Theorem 3, this estimator can *eventually* be chosen as the estimator $\left(\hat\lambda_n^{(1)}, \hat\beta_n^{(1)}\right)$. Thus $\left(\hat\lambda_n^{(3)}, \hat\beta_n^{(3)}\right) \in K_1$ *eventually*, and $\left(\hat\lambda_n^{(3)}, \hat\beta_n^{(3)}\right)$ can be chosen as the estimator $\left(\hat\lambda_n^{(2)}, \hat\beta_n^{(2)}\right)$. This implies the strong consistency of $\left(\hat\lambda_n^{(2)}, \hat\beta_n^{(2)}\right)$.

Below are some extra conditions needed for an estimator $\left(\hat\lambda_n^{(2)}, \hat\beta_n^{(2)}\right)$ to be asymptotically consistent.

(viii) β_0 is an inner point of Θ_β.

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

$$\begin{aligned}\Theta_\lambda^\varepsilon := \big\{f\colon [0,\tau]\to\mathbb{R} \,\big|\, & f(t)\ge\varepsilon,\ \forall t\in[0,\tau],\\ & |f(t)-f(s)| \le (L-\varepsilon)|t-s|, \forall t,s\in[0,\tau]\big\}.\end{aligned}$$

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

(xi) $\mathsf{E}\,U = 0$ and

$$\mathsf{E}\,e^{2D\|U\|} < \infty, \qquad D := \max_{\beta\in\Theta_\beta}\|\beta\| + \varepsilon$$

for some $\varepsilon > 0$.

(xii) $\mathsf{E}\,e^{2D\|X\|} < \infty$, where the number D is defined in condition (xi).

In what follows we use the notation introduced in [3]. Let

$$a(t) = \mathsf{E}\left[Xe^{\beta_0^TX}G_T(t|X)\right], \qquad b(t) = \mathsf{E}\left[e^{\beta_0^TX}G_T(t|X)\right],$$

$$p(t) = \mathsf{E}\left[XX^Te^{\beta_0^TX}G_T(t|X)\right], \qquad T(t) = p(t)b(t) - a(t)a^T(t), \qquad K(t) = \frac{\lambda_0(t)}{b(t)},$$

$$A = \mathsf{E}\left[XX^Te^{\beta_0^TX}\int_0^Y \lambda_0(u)\,du\right], \qquad M = \int_0^\tau T(u)K(u)G_c(u)\,du.$$

For $i \ge 1$, consider the random vectors

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

where

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

Let

$$\Sigma_\beta = 4\cdot \operatorname{Cov}(\zeta_1), \qquad m(\varphi_\lambda) = \int_0^\tau \varphi_\lambda(u)a(u)G_C(u)\,du,$$
$$\sigma_\varphi^2 = 4\cdot \operatorname{Var}\langle q'(Y,\Delta,W,\lambda_0,\beta_0),\varphi\rangle$$
$$= 4\cdot \operatorname{Var}\Bigg[\frac{\Delta\cdot\varphi_\lambda(Y)}{\lambda_0(Y)} - \frac{\exp\left(\beta_0^T W\right)}{M_U(\beta_0)}\int_0^Y \varphi_\lambda(u)\,du + \Delta\cdot\varphi_\beta^T W$$
$$+ \varphi_\beta^T \frac{M_U(\beta_0)W - \mathsf{E}\left(Ue^{\beta_0^T U}\right)}{M_U(\beta_0)^2}\exp\left(\beta_0^T W\right)\int_0^Y \lambda_0(u)\,du\Bigg]$$

with $\varphi=(\varphi_\lambda,\varphi_\beta)\in C[0,\tau]\times\mathbb{R}^m$, where the symbol q' stands for the Frechet derivative.

Now we are in position to apply Theorem 1 of [3] and obtain the asymptotic normality of $\hat\beta_n^{(2)}$ and $\hat\lambda_n^{(2)}$. Note that this result follows from the asymptotic normality of the consistent estimators $\hat\beta_n^{(3)}$ and $\hat\lambda_n^{(3)}$.

Theorem 6. *Let assumptions* (i), (ii), (v)–(xii) *hold. Then M is nonsingular and*

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

Moreover,

$$\sqrt{n}\int_0^\tau \left(\hat\lambda_n^{(2)} - \lambda_0\right)(u)f(u)G_C(u)\,du \xrightarrow{d} N\left(0,\sigma_\varphi^2(f)\right)$$

for all continuous functions f satisfying the Lipschitz condition in the interval $[0,\tau]$, where $\sigma_\varphi^2(f)=\sigma_\varphi^2$, $\varphi=(\varphi_\lambda,\varphi_\beta)$, $\varphi_\beta=-A^{-1}m(\varphi_\lambda)$, and φ_λ is a unique solution of the integral Fredholm equation

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

One can use the method of [3] to evaluate the estimator $\left(\hat\lambda_n^{(2)},\hat\beta_n^{(2)}\right)$.

4. Concluding remarks

In this paper, an estimator is constructed for the baseline hazard function $\lambda(\cdot)$ and parameter β in the Cox proportional risks model with measurement errors under comparatively weak assumptions. In contrast to the papers [6] and [3], the set of parameters is unbounded in our setting. The estimator considered is consistent and can be modified in such a way that the modified estimator is asymptotically normal. The procedure for evaluating the estimators is described. Further investigations will be devoted to constructions of confidence regions.

Bibliography

[1] P. K. Andersen and R. D. Gill, *Cox's regression model for counting processes: a large sample study*, Ann. Statist. **10** (1982), no. 4, 1100–1120. MR673646

[2] T. Augustin, *An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions*, Scand. J. Stat. **31** (2004), no. 1, 43–50. MR2042597

[3] C. Chimisov and A. Kukush, *Asymptotic normality of corrected estimator in Cox proportional hazards model with measurement error*, Modern Stoch. Theory Appl. **1** (2014), no. 1, 13–32. MR3314791

[4] D. R. Cox, *Regression models and life tables*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **34** (1972), 187–220. MR0341758

[5] M. Gu and F. H. Kong, *Consistent estimation in Cox proportional hazards model with covariate measurement errors*, Statist. Sinica **32** (1999), no. 9, 953–969. MR1744820

[6] A. Kukush, S. Baran, I. Fazekas, and E. Usoltseva, *Simultaneous estimation of baseline hazard rate and regression parameters in Cox proportional hazards model with measurement error*, J. Statist. Res. **45** (2011), no. 2, 77–94. MR2934363

[7] G. Pfanzagl, *On the measurability and consistency of minimum contrast estimates*, Metrika **14** (1969), 249–273.

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

Department of Mathematical Analysis, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue, 6, Kyiv 03127, Ukraine

Email address: alexander_kukush@univ.kiev.ua

Department of Probability Theory, Statistics, and Actuarial Mathematics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue, 6, Kyiv 03127, Ukraine

Email address: chernovaoksan@gmail.com

Received 07/MAR/2017

Translated by N. N. SEMENOV