Cox proportional hazards model with heteroscedastic errors in covariates

Alexander Kukush, Oksana Chernova
Taras Shevchenko National University of Kyiv, Ukraine

02/06/21

Contents

Survival Analysis

Survival Analysis (SA) is a comprehensive set of statistical methods for analysis of data representing life times (times to the occurrence of some specified event), the time being measured from some well-defined time origin until the event of interest.

Examples:

- biological sciences: time from start of a clinical trial to death
- engineering: time from instalment of a device to failure
- insurance: time from the beginning of contract to occurrence of claim

Let a nonnegative r.v. T be the time to the event (survival time, lifetime).
Aim: draw inference about the distribution of T.

Distribution of T is characterized by

- survival function $G_T(t) = \mathrm{P}(T > t) = 1 - F_T(t)$, $t \geq 0$,

  or equivalently,

- by intensity function (the so-called hazard)

$$\tilde{\lambda}(t) = \lim_{h \to 0} \frac{\mathrm{P}(T \leq t + h \mid T > t)}{h} = -\frac{d}{dt} \ln \left(G_T(t) \right).$$

Let a nonnegative r.v. T be the time to the event (survival time, lifetime).
Aim: draw inference about the distribution of T.

Distribution of T is characterized by

- survival function $G_T(t) = \mathrm{P}(T > t) = 1 - F_T(t)$, $t \geq 0$,

  or equivalently,

- by intensity function (the so-called hazard)

$$\tilde{\lambda}(t) = \lim_{h \to 0} \frac{\mathrm{P}(T \leq t + h \mid T > t)}{h} = -\frac{d}{dt} \ln \left(G_T(t) \right).$$

Model description

Cox, D.R. (1972), J. R. Stat. Soc., Ser. B **34**, 187–220.

In the CPH model a lifetime T has the following intensity function

$$\lambda(t|X;\lambda,\beta) = \lambda(t)\exp(\beta^T X), \quad t \ge 0, \quad \text{where}$$

the covariate X is a random vector distributed in $\mathbb{R}^k$,
$\beta \in \Theta_\beta \subset \mathbb{R}^k$ is the regression parameter,
and the baseline hazard function $\lambda(\cdot) \in \Theta_\lambda$ is continuous.

The conditional pdf of T given X is

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

Model description

Cox, D.R. (1972), J. R. Stat. Soc., Ser. B **34**, 187–220.

In the CPH model a lifetime T has the following intensity function

$$\lambda(t|X;\lambda,\beta)=\lambda(t)\exp(\beta^T X),\quad t\geq 0,\quad \text{where}$$

the covariate X is a random vector distributed in $\mathbb{R}^k$,
$\beta\in\Theta_\beta\subset\mathbb{R}^k$ is the regression parameter,
and the baseline hazard function $\lambda(\cdot)\in\Theta_\lambda$ is continuous.

The conditional pdf of T given X is

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

Let a r.v. C distributed on $[0, \tau]$ denote the censoring time. C is independent of T. The exact event time T may not be available, instead we observe the right-censored lifetime

$$Y = \min\{T, C\}$$

and the indicator of non-censoring

$$\Delta = I_{T \le C}.$$

Survival function G_C is unknown, while we know τ.

Therefore, the baseline hazard function is estimated only on $[0, \tau]$ and we assume $\lambda(\cdot) \in \Theta_\lambda \subset C[0, \tau]$.

In practice, some covariates could be measured with error.

We use the classical measurement error model, i.e.

$$W_i = X_i + U_i$$

is observed.

First, we assume that measurement errors U_i are i.i.d. and moment generating function $M_U(z) = \mathrm{E}e^{z^T U}$ is known.

The couple (T_i, X_i), censor C, and U_i are independent. We collect observations (Y_i, Δ_i, W_i), $i = 1, \ldots, n$.

All the results are valid for the model without measurement errors!

In that case we set $U_i = 0, M_{U_i} \equiv 1$.

In practice, some covariates could be measured with error.

We use the classical measurement error model, i.e.

$$W_i = X_i + U_i$$

is observed.

First, we assume that measurement errors U_i are i.i.d. and moment generating function $M_U(z) = \mathsf{E}e^{z^T U}$ is known.

The couple (T_i, X_i), censor C, and U_i are independent. We collect observations $(Y_i, \Delta_i, W_i)\,, i = 1, \ldots, n$.

All the results are valid for the model without measurement errors!

In that case we set $U_i = 0, M_{U_i} \equiv 1$.

Consider independent copies $(X_i, T_i, C_i, Y_i, \Delta_i, U_i, W_i),\ i = 1, ..., n$.

Aim: Based on triples $(Y_i, \Delta_i, W_i),\ i = 1, ..., n$, estimate true parameters β, $\lambda(t),\ t \in [0, \tau]$ over a parameter set $\Theta = \Theta_\lambda \times \Theta_\beta$.

Joint pdf of $(Y, \Delta) \mid X$ on $\mathcal{X} = (0, \tau] \times \{0, 1\}$ w.r.t. $\mu = \lambda_1 \times \delta_1 + \mu_C \times \delta_0$ is as follows

$$f(y, \delta | X; \lambda, \beta) = f_T^{\delta}(y | X; \lambda, \beta) G_T^{1-\delta}(y | X; \lambda, \beta) G_C^{\delta}(y), \quad (y, \delta) \in \mathcal{X}.$$

An elementary objective function q is derived from $\ln f(Y, \Delta \mid X)$,

$$q(Y, \Delta, X; \lambda, \beta) = \Delta \cdot (\ln \lambda(Y) + \beta^T X) - e^{\beta^T X} \int_0^Y \lambda(u) du.$$

Joint pdf of $(Y, \Delta) \mid X$ on $\mathcal{X} = (0, \tau] \times \{0, 1\}$ w.r.t. $\mu = \lambda_1 \times \delta_1 + \mu_C \times \delta_0$ is as follows

$$f(y, \delta | X; \lambda, \beta) = f_T^{\delta}(y|X; \lambda, \beta) G_T^{1-\delta}(y|X; \lambda, \beta) G_C^{\delta}(y), \quad (y, \delta) \in \mathcal{X}.$$

An elementary objective function q is derived from $\ln f(Y, \Delta \mid X)$,

$$q(Y, \Delta, X; \lambda, \beta) = \Delta \cdot (\ln \lambda(Y) + \beta^T X) - e^{\beta^T X} \int_0^Y \lambda(u) du.$$

In the presence of measurement errors, due to Augustin (2004),

$$\mathrm{E}[q^{cor}(Y, \Delta, W; \lambda, \beta) | Y, \Delta, X] = q(Y, \Delta, X; \lambda, \beta).$$

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

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

Joint pdf of $(Y, \Delta) \mid X$ on $\mathcal{X} = (0, \tau] \times \{0, 1\}$ w.r.t. $\mu = \lambda_1 \times \delta_1 + \mu_C \times \delta_0$ is as follows

$$f(y, \delta | X; \lambda, \beta) = f_T^{\delta}(y|X; \lambda, \beta) G_T^{1-\delta}(y|X; \lambda, \beta) G_C^{\delta}(y), \quad (y, \delta) \in \mathcal{X}.$$

An elementary objective function q is derived from $\ln f(Y, \Delta \mid X)$,

$$q(Y, \Delta, X; \lambda, \beta) = \Delta \cdot (\ln \lambda(Y) + \beta^T X) - e^{\beta^T X} \int_0^Y \lambda(u) du.$$

In the presence of measurement errors, due to Augustin (2004),

$$\mathrm{E}[q^{cor}(Y, \Delta, W; \lambda, \beta) | Y, \Delta, X] = q(Y, \Delta, X; \lambda, \beta).$$

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

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

In vast amount of literature, e.g., in Cox (1972) or Augustin (2004), the baseline hazard function is assumed belonging to a parametric set of piece-wise constant functions, while in Kukush et al. (2011) and Chimisov & Kukush (2014)

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

The consistency and asymptotic normality of a simultaneous estimator

$$(\hat{\lambda}_n, \hat{\beta}_n) = \underset{(\lambda,\ \beta) \in \Theta_\lambda \times \Theta_\beta}{\arg\max} \ Q_n^{cor}(\lambda, \beta)$$

were studied.

We consider an unbounded parameter set of Lipschitz functions

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

Assumptions

1. $\Theta_\lambda = \{f : [0,\tau] \to \mathbb{R} \mid f(t) \geq 0,\ |f(t) - f(s)| \leq L|t-s|, \forall t,s \in [0,\tau]\}$, where L is a fixed positive constant.
2. $\Theta_\beta \subset \mathbb{R}^k$ is a compact set.
3. $\mathrm{E}\, U = 0$ and for some constant $\epsilon > 0$,

$$\mathrm{E}\, e^{2D\|U\|} < \infty \text{ and } \mathrm{E}\, e^{2D\|X\|} < \infty, \text{ where } D := \max_{\beta \in \Theta_\beta} \|\beta\| + \epsilon.$$

4. τ is the right endpoint of the distribution of C.
5. Matrix $\mathrm{E}[XX^\top]$ is positive definite.

Additionally:

6. True β is an interior point of Θ_β.
7. True λ belongs to Θ_λ^ϵ for some $\epsilon > 0$, where

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

8. $\mathrm{P}(C > 0) = 1$.

Estimation

Definition

Fix a sequence $\{\varepsilon_n\}$ *of positive numbers, with* $\varepsilon_n \downarrow 0$*, as* $n \to \infty$*. Any Borel function* $(\hat{\lambda}_n, \hat{\beta}_n)$ *of observations* (Y_i, Δ_i, W_i)*,* $i = 1, ..., n$*, with values in* $\Theta := \Theta_\lambda \times \Theta_\beta$ *and such that*

$$Q_n^{cor}(\hat{\lambda}_n, \hat{\beta}_n) \geq \sup_{(\lambda,\beta)\in\Theta} Q_n^{cor}(\lambda, \beta) - \varepsilon_n,$$

is called the estimator of (λ, β)*.*

Theorem (Kukush & Chernova, 2018)

Under conditions (i) to (v), $(\hat{\lambda}_n, \hat{\beta}_n)$ *is a strongly consistent estimator of the true parameters* (λ, β)*, that is*

$$\max_{t\in[0,\tau]} |\hat{\lambda}_n(t) - \lambda(t)| \to 0 \quad \text{and} \quad \hat{\beta}_n \to \beta \text{ a.s. as } n \to \infty.$$

Estimation

Definition

Fix a sequence $\{\varepsilon_n\}$ *of positive numbers, with* $\varepsilon_n \downarrow 0$*, as* $n \to \infty$*. Any Borel function* $(\hat{\lambda}_n, \hat{\beta}_n)$ *of observations* (Y_i, Δ_i, W_i)*,* $i = 1, \ldots, n$*, with values in* $\Theta := \Theta_\lambda \times \Theta_\beta$ *and such that*

$$Q_n^{cor}(\hat{\lambda}_n, \hat{\beta}_n) \geq \sup_{(\lambda,\beta)\in\Theta} Q_n^{cor}(\lambda, \beta) - \varepsilon_n,$$

is called the estimator of (λ, β)*.*

Theorem (Kukush & Chernova, 2018)

Under conditions (i) to (v), $(\hat{\lambda}_n, \hat{\beta}_n)$ *is a strongly consistent estimator of the true parameters* (λ, β)*, that is*

$$\max_{t\in[0,\tau]} |\hat{\lambda}_n(t) - \lambda(t)| \to 0 \quad and \quad \hat{\beta}_n \to \beta \text{ a.s. as } n \to \infty.$$

Theorem (Kukush & Chernova, 2018)

Assume conditions (i) – (viii). Then

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N_k(0, \Sigma).$$

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

$$\sqrt{n} \int_0^\tau (\hat{\lambda}_n - \lambda)(u) f(u) du \xrightarrow{d} N(0, \sigma^2(f)),$$

with a nonsingular matrix Σ *and a function* $\sigma^2(f)$ *which is positive if f is not identical zero.*

There exist explicit formulae for matrices and functions from the latter Theorem.

The following statistical procedures were developed:

- construction of asymptotic confidence ellipsoid for the regression parameter and of asymptotic confidence interval for an integral functional of the baseline hazard function,
- hypothesis testing for the regression parameter and for an integral functional of the baseline hazard function,
- goodness-of-fit test for the model.

Heteroscedastic measurement errors

Now, consider non-i.i.d. measurement errors. Let

$$W_i = X_i + U_i$$

be observed, where measurement errors U_i are independent but not necessarily identically distributed, while distributions of U_i are known.

The couple (T_i, X_i), censor C, and U_i are independent. We collect observations (Y_i, Δ_i, W_i), $i = 1, \ldots, n$.

Assumptions

1. $\Theta_\lambda := \{ f : [0,\tau] \to \mathbb{R} \mid f(t) \geq c, \forall t \in [0,\tau],\ f(0) \leq A, \text{ and } |f(t) - f(s)| \leq L|t-s|, \forall t,s \in [0,\tau] \}$, where $L > 0$ is a fixed constant.
2. $\Theta_\beta \subset \mathbb{R}^k$ is a compact set.
3. For all $i \geq 1$, $\mathsf{E}\, \mathsf{U_i} = 0$ and for some positive constants K and ϵ,

$$\mathsf{E}\ e^{2D\|\mathsf{U_i}\|} \leq K,\ \text{ with } D := \max_{\beta\in\Theta_\beta} \|\beta\| + \epsilon.$$

4. $\mathsf{E}\ e^{2D\|\mathsf{X}\|} < \infty$, with D defined in assumption 3.
5. τ is the right endpoint of censor's distribution.
6. Matrix $\mathsf{E}\ \mathsf{X}\mathsf{X}^\top$ is positive definite.

Definition

Any Borel function $(\hat{\lambda}_n, \hat{\beta}_n)$ *of observations* (Y_i, Δ_i, W_i), $i = 1, ..., n$, *with values in* $\Theta := \Theta_\lambda \times \Theta_\beta$ *and such that*

$$(\hat{\lambda}_n, \hat{\beta}_n) = \underset{(\lambda,\beta)\in\Theta_\lambda\times\Theta_\beta}{\arg\max}\ Q_n^{cor}(\lambda, \beta).$$

is called the estimator of (λ, β), *where* Q_n^{cor} *is the objective function corrected for heteroscedastic measurement errors.*

Theorem

Under conditions (1) to (6), $(\hat{\lambda}_n, \hat{\beta}_n)$ *is a strongly consistent estimator of the true parameters* (λ, β), *that is*

$$\max_{t\in[0,\tau]} |\hat{\lambda}_n(t) - \lambda(t)| \to 0 \quad and \quad \hat{\beta}_n \to \beta \text{ a.s. as } n \to \infty.$$

References

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.

Cox, D.R. *Regression models and life-tables*, J. R. Stat. Soc., Ser. B **34** (1972), 187–220.

O. Chernova, A. Kukush, *Confidence regions in Cox proportional hazards model with measurement errors and unbounded parameter set*, Mod. Stoch. Theory Appl. **5** (2018), no. 1, 37–52.

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

- 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.
- A. Kukush, O. Chernova, *Consistent estimation in Cox proportional hazards model with measurement errors and unbounded parameter set*, Theor. Probability and Math. Statist. **96** (2018), 101–110.
- A. Kukush, O. Chernova, *Goodness-of-fit test in the Cox proportional hazards model with measurement errors*, Theor. Probability and Math. Statist. **99** (2019), 125–135.