Shape Constrained Nonparametric Baseline Estimators in the Cox Model

Joint work with Rik Lopuhaä
(TU Delft)

Tina Nane, Center for Science and Technology Studies.
Basic concepts in survival analysis

- Events of interest - death, onset (relapse) of a disease, etc
- Let $X \sim F$ denote the survival time, with density $f$
- Functions that characterize the distribution of $X$
  - The survival function $S(x) = \mathbb{P}(X > x)$
  - The hazard function
    \[
    \lambda(x) = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(x \leq X < x + \Delta x | X \geq x)}{\Delta x} = \frac{f(x)}{S(x)}
    \]
  - The cumulative hazard function $\Lambda(x) = \int_0^x \lambda(u) \, du$
- Let $C \sim G$ denote the censoring time
- Let $Z$ denote the covariate (age, weight, treatment)
The Cox proportional hazards model

- Right-censored data \((T_i, \Delta_i, Z_i)\), for \(i = 1, \ldots, n\)
  - \(T = \min(X, C)\) denotes the follow-up time
  - \(\Delta = \{X \leq C\}\) is the censoring indicator
  - The covariate vector \(Z \in \mathbb{R}^p\) is time invariant
  - \(X|Z \perp C|Z\)

- The Cox model
  \[
  \lambda(x|z) = \lambda_0(x)e^{\beta_0'z},
  \]
  where
  - \(\lambda_0\) is the underlying baseline hazard function
  - \(\beta_0 \in \mathbb{R}^p\) is the vector of the underlying regression coefficients
Assumptions

- $X \sim F$, $C \sim G$, $T \sim H$

- $F$, $G$ are assumed absolutely continuous.

- (A.1) Let $\tau_F$, $\tau_G$ and $\tau_H$ be the end points of the support of $F$, $G$, $H$. Then

\[ \tau_H = \tau_G < \tau_F \]

- (A.2) There exists $\varepsilon > 0$ such that

\[ \sup_{|\beta - \beta_0| \leq \varepsilon} \mathbb{E} \left[ |Z|^2 e^{2\beta'Z} \right] < \infty, \]

where $|\cdot|$ denotes the Euclidean norm
Estimating monotone baseline hazards in the Cox model

- The NPMLE $\hat{\lambda}_n$ of a nondecreasing baseline hazard
- Let $T(1) \leq \cdots \leq T(n)$ denote the ordered follow-up times
- For $\beta$ fixed, maximize the (log)likelihood function over all nondecreasing baseline hazards and obtain $\hat{\lambda}_n(x; \beta)$
  - zero, for $x < T(1)$
  - constant on $[T(i), T(i+1))$, for $i = 1, 2, \ldots, n-1$
  - $\infty$, for $x \geq T(n)$
- Replace $\beta$ in $\hat{\lambda}_n(x; \beta)$ by $\hat{\beta}_n$, the maximum partial likelihood estimator
- We propose $\hat{\lambda}_n(x) = \hat{\lambda}_n(x; \hat{\beta}_n)$ as our estimator of $\lambda_0$
Estimating monotone baseline hazards in the Cox model

- Grenander-type estimator
- 1. Start from the Breslow estimator $\Lambda_n$ of the baseline cumulative hazard $\Lambda_0$
Estimating monotone baseline hazards in the Cox model

- 2. Take its Greatest Convex Minorant (GCM) $\tilde{\Lambda}_n$
Estimating monotone baseline hazards in the Cox model

• 3. The Grenander-type estimator $\tilde{\lambda}_n$ is defined as the left-hand slope of $\tilde{\Lambda}_n$
Estimating monotone baseline hazards in the Cox model

- Another estimator of a nondecreasing baseline hazard was proposed by Chung and Chang (1994)
- Consistency: $\hat{\lambda}_n(x) \rightarrow \lambda_0(x)$ a.s.
- No limiting distribution available
Estimating monotone baseline hazards in the Cox model

- Comparison between the three baseline hazard estimators
Estimating monotone baseline densities in the Cox model

- Grenander-type estimator of a monotone baseline density $f_0$
- Since
  \[ F_0(x) = 1 - e^{-\Lambda_0(x)} \]
- We propose
  \[ F_n(x) = 1 - e^{-\Lambda_n(x)}, \]
  where $\Lambda_n$ is the Breslow estimator.

- Define the nonincreasing Grenander-type estimator $\tilde{f}_n$ as the left derivative of the Least Concave Majorant (LCM) of $F_n$
Pointwise consistency

- **Theorem 1** (Lopuhaä & Nane, 2013a)

Assume that (A.1) and (A.2) hold and that $\lambda_0$ is nondecreasing on $[0, \infty)$ and $f_0$ is nonincreasing on $[0, \infty)$. Then, for any $x_0 \in (0, \tau_H)$,

\[
\begin{align*}
\lambda_0(x_0-) & \leq \liminf_{n \to \infty} \hat{\lambda}_n(x_0) \leq \limsup_{n \to \infty} \hat{\lambda}_n(x_0) \leq \lambda_0(x_0+), \\
\lambda_0(x_0-) & \leq \liminf_{n \to \infty} \tilde{\lambda}_n(x_0) \leq \limsup_{n \to \infty} \tilde{\lambda}_n(x_0) \leq \lambda_0(x_0+), \\
f_0(x_0+) & \leq \liminf_{n \to \infty} \tilde{f}_n(x_0) \leq \limsup_{n \to \infty} \tilde{f}_n(x_0) \leq f_0(x_0-),
\end{align*}
\]

with probability one. The values $\lambda_0(x_0-)$, $f_0(x_0-)$ and $\lambda_0(x_0+)$, $f_0(x_0+)$ denote the left (right) limit of the baseline hazard and density function at $x_0$. 
Asymptotic distribution

- Typical features for isotonic estimators
  - $n^{1/3}$ rate of convergence
  - non-normal limiting distribution

- Groeneboom (1985) recipe
  1. Define an inverse process
  2. Use the switching relationship
  3. Use the Hungarian embedding (KMT construction) to derive the limiting distribution of the inverse process
  4. Obtain the limiting distribution of the monotone estimator
Asymptotic distribution

• For the Grenander-type estimator $\tilde{\lambda}_n$
  1. Inverse process

$$U_n(a) = \arg\min_{x \in [0, T_n]} \{ \Lambda_n(x) - ax \},$$

for $a > 0$, where $\arg\min$ denotes the largest location of the minimum

2. For any $a > 0$, the following switching relationship holds

$$U_n(a) \geq x \iff \tilde{\lambda}_n(x) \leq a,$$

with probability one
Asymptotic distribution

• For a fixed $x_0$,

$$\mathbb{P} \left( \frac{n^{1/3}}{3} \left[ \tilde{\lambda}_n(x_0) - \lambda_0(x_0) \right] > a \right) = \mathbb{P} \left( \frac{n^{1/3}}{3} \left[ U_n(\lambda_0(x_0) + n^{-1/3} a) - x_0 \right] < 0 \right)$$

• Moreover

$$\frac{n^{1/3}}{3} \left[ U_n(\lambda_0(x_0) + n^{-1/3} a) - x_0 \right] = \arg\min_{x \in I_n(x_0)} \{ \mathbb{Z}_n(x) - ax \}$$

where $I_n(x_0) = [-n^{1/3} x_0, n^{1/3} (T(n) - x_0)]$ and for $x \in I_n(x_0)$

$$\mathbb{Z}_n(x) = n^{2/3} \left\{ \left[ \Lambda_n(x_0 + n^{-1/3} x) - \Lambda_0(x_0 + n^{-1/3} x) \right] - \left[ \Lambda_n(x_0) - \Lambda_0(x_0) \right] + \Lambda_0(x_0 + n^{-1/3} x) - \Lambda_0(x_0) - n^{-1/3} \lambda_0(x_0) x \right\}$$
Asymptotic distribution

- 3. No embedding available for the Breslow estimator
- 3’. Linearization result of the Breslow estimator (Lopuhaä & Nane, 2013b)

Let \( \Phi(\beta_0, x) = \mathbb{E}[\{ T \geq x \} e^{\beta_0'Z}] \)

**Theorem 2** (Lopuhaä & Nane, 2013a)

Assume \((A.1)\) and \((A.2)\) and let \(x_0 \in (0, \tau_H)\). Suppose that \(\lambda_0\) is nondecreasing on \([0, \infty)\) and continuously differentiable in a neighborhood of \(x_0\), with \(\lambda_0(x_0) \neq 0\) and \(\lambda_0'(x_0) > 0\). Then,

\[
 n^{1/3} \left( \frac{\Phi(\beta_0, x_0)}{4\lambda_0(x_0)\lambda_0'(x_0)} \right)^{1/3} \left[ \tilde{\lambda}_n(x_0) - \lambda_0(x_0) \right] \rightarrow_d \arg\min_{t \in \mathbb{R}} \{ W(t) + t^2 \},
\]

where \(W\) is a standard two-sided Brownian motion originating from zero.
Asymptotic distribution

- **Theorem 3** (Lopuhaä & Nane, 2013a)

  Assume (A.1) and (A.2) and let \( x_0 \in (0, \tau_H) \). Suppose that \( \lambda_0 \) is nondecreasing on \([0, \infty)\) and continuously differentiable in a neighborhood of \( x_0 \), with \( \lambda_0(x_0) \neq 0 \) and \( \lambda'_0(x_0) > 0 \). Then,

  \[
  n^{1/3} \left( \frac{\Phi(\beta_0, x_0)}{4\lambda_0(x_0)\lambda'_0(x_0)} \right)^{1/3} \left[ \hat{\lambda}_n(x_0) - \lambda_0(x_0) \right] \rightarrow_d \underset{t \in \mathbb{R}}{\text{argmin}} \{ W(t) + t^2 \},
  \]

  where \( W \) is a standard two-sided Brownian motion originating from zero.
Asymptotic distribution

- **Theorem 4** (Lopuhaä & Nane, 2013a)

Assume (A.1) and (A.2) and let $x_0 \in (0, \tau_H)$. Suppose that $f_0$ is nonincreasing on $[0, \infty)$ and continuously differentiable in a neighborhood of $x_0$, with $f_0(x_0) \neq 0$ and $f_0'(x_0) < 0$. Let $F_0$ be the baseline distribution function. Then,

$$n^{1/3} \left( \frac{\Phi(\beta_0, x_0)}{4f_0(x_0)f_0'(x_0)[1 - F_0(x_0)]} \right)^{1/3} \left[ \tilde{f}_n(x_0) - f_0(x_0) \right] \rightarrow_d \argmin_{t \in \mathbb{R}} \{W(t) + t^2\},$$

where $W$ is a standard two-sided Brownian motion originating from zero.
Hypothesis testing

- Likelihood ratio test of $H_0 : \lambda_0(x_0) = \theta_0$ versus $H_1 : \lambda_0(x_0) \neq \theta_0$

- Let $L_{\beta}(\lambda_0)$ the (log)likelihood function

- For fixed $\beta \in \mathbb{R}^p$, $x_0 \in (0, \tau_H)$ and $\theta_0 \in (0, \infty)$ fixed
  
  maximize $L_{\beta}(\lambda_0)$ under $H_0$

- Propose $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(x; \hat{\beta}_n)$ as the constrained NPMLE
Hypothesis testing

- Recall $\hat{\lambda}_n(x) = \hat{\lambda}_n(x; \hat{\beta}_n)$, the unconstrained NPMLE estimator of a nonincreasing $\lambda_0$
- By Theorem 3,

$$n^{1/3} \left[ \hat{\lambda}_n(x_0) - \lambda_0(x_0) \right] \xrightarrow{d} \left( \frac{4\lambda_0(x_0)\lambda_0'(x_0)}{\Phi(\beta_0, x_0)} \right)^{1/3} \arg\min_{t \in \mathbb{R}} \{ W(t) + t^2 \}$$

$$\equiv C(x_0) \arg\min_{t \in \mathbb{R}} \{ W(t) + t^2 \}$$

$$\equiv \frac{C(x_0)}{2} g(0),$$

where $g(x)$ is the slope at $x$ of the GCM of $\{ W(t) + t^2 \}$
Hypothesis testing

Similarly, it can be shown that

\[ n^{1/3} \left[ \hat{\lambda}_n(x_0) - \lambda_0(x_0) \right] \rightarrow_d \frac{C(x_0)}{2} g^0(0), \]

where \( g^0 \) is the constrained slope process of the GCM of \( \{W(t) + t^2\} \).
Hypothesis testing

- Banerjee & Wellner (2001)

Fig. 3. The one-sided convex minorants $\tilde{G}_L$ and $\tilde{G}_R$ and $W(t) + t^2$. 
Hypothesis testing

- Banerjee & Wellner (2001)

**Fig. 4.** Close-up view of $G_{1,1}$, $\tilde{G}_{L,R}$, $G^0_{1,1}$ and $W(t) + t^2$. 
Hypothesis testing

- Replacing $\beta$ by $\hat{\beta}_n$ in $L_\beta(\lambda_0)$ gives

\[ 2 \log \xi_n(\theta_0) = 2L_{\hat{\beta}_n}(\hat{\lambda}_n) - 2L_{\hat{\beta}_n}(\hat{\lambda}_n^0) \]

- **Theorem 5** (Nane, 2013)

Suppose that (A.1) and (A.2) hold and let $x_0 \in (0, \tau_H)$. Assume that $\lambda_0$ is nondecreasing on $[0, \infty)$ and continuously differentiable in a neighborhood of $x_0$, with $\lambda_0(x_0) \neq 0$ and $\lambda_0'(x_0) > 0$. Then, under the null hypothesis,

\[ 2 \log \xi_n(\theta_0) \to_d \mathbb{D}, \]

where $\mathbb{D} = \int [(g(u))^2 - (g^0(u))^2]du$. 


Interval estimation

- Pointwise confidence intervals for $\lambda_0(x_0)$
  - Likelihood ratio method
    \[
    \{ \theta : 2 \log \xi_n(\theta) \leq q(D, 1 - \alpha) \}
    \]
    where $q(D, 1 - \alpha)$ is the $(1 - \alpha)^{th}$ quantile of $D$
    (Banerjee & Wellner, 2005)
- Asymptotic distribution
  
  $[\hat{\lambda}_n(x_0) - n^{-1/3} \hat{C}_n(x_0) q(Z, 1 - \alpha/2), \hat{\lambda}_n(x_0) + n^{-1/3} \hat{C}_n(x_0) q(Z, 1 - \alpha/2)]$,

  where $Z = \text{argmin}\{W(t) + t^2\}$ and $q(Z, 1 - \alpha/2)$ is the
  $(1 - \alpha/2)^{th}$ quantile of the distribution $Z$
  (Groeneboom & Wellner, 2001)
Current research

- Citation analysis
  - Event of interest - time to first citation, $5^{th}$ citation, etc
  - Time frame - first five years after publication (field specific)
  - Censored data
  - Covariates - document type, collaboration type, number of authors, number of pages, etc
  - Nondecreasing baseline hazard
References


