Estimation dans un modèle de Cox stratifié partiellement observé

Detais A., Dupuy J.-F.
Institut de Mathématiques de Toulouse
Université Toulouse III

Journées MAS de la SMAI, Rennes, 27-29 août 2008

Let T_i^0 and $\widetilde{X}_i$ be the random **failure time** and **covariate** for individual i. Let C_i be a positive random variable.

For each of n iid $(T_i^0, C_i, \widetilde{X}_i)$, we observe $(T_i, \Delta_i, \widetilde{X}_i)$, where

(Observations)

$$T_i := \min(T_i^0, C_i) \text{ and } \Delta_i := 1(T_i^0 \leq C_i),$$

under the assumption: T_i^0 and C_i are independent $|\widetilde{X}_i$.

(Conditional hazard rate)

Based on the observations $(T_i, \Delta_i, \widetilde{X}_i), i = 1, \ldots, n$, we wish to ***estimate*** *the* ***conditional hazard rate*** *of $T^0|\widetilde{X}$:*

$$\lambda_{T^0|\widetilde{X}}(t, x) = \frac{f_{T^0|\widetilde{X}}(t, x)}{1 - F_{T^0|\widetilde{X}}(t, x)}, \quad t \geq 0, \ x \in \mathbb{R}^p.$$

Define the

(Failure counting and "at-risk" processes)

$$N_i(t) := 1(T_i \le t, \Delta_i = 1) \quad \textit{and} \quad Y_i(t) := 1(T_i \ge t)$$

$N_i(t)$ admits the

(Doob-Meyer decomposition)

$$N_i(t) = \int_0^t \lambda_{T^0|\widetilde{X}}(s, \widetilde{X}_i) Y_i(s)\, ds + M_i(t)$$

for the filtration $\mathcal{F}_t := \sigma\{N_i(s), Y_i(s), \widetilde{X}_i : 0 \le s \le t, 1 \le i \le n\}$ and $M_i(t)$ is an $(\mathcal{F}_t)$**-martingale**.

One popular model for the **intensity** $\lambda_{T^0|\widetilde{X}}(s, \widetilde{X}_i)Y_i(s)$ of $N_i(s)$ is the **Cox stratified proportional intensity** model.

$\hookrightarrow$ **mixture-model** extension of the “usual” Cox model: the population is divided into **K sub-populations** (called strata) such that:

(Intensity of $N_i(t)$ in stratum k ($k = 1, \ldots, K$))

$$\lambda_{T^0|\widetilde{X}}(t, \widetilde{X}_i) = \lambda_{0,k}(t) \exp(\beta_0' X_i) \quad \textit{when } S_i = k$$

where

$$\widetilde{X}_i = (X_i, S_i),$$

the $\lambda_{0,k}(\cdot)$ ($k = 1, \ldots, K$) are **unknown nonnegative functions**, and $\beta_0 \in \mathbb{R}^p$.

$\hookrightarrow$ **semiparametric** model. The estimation of β_0 from n copies

$$(T_i, \Delta_i, X_i, S_i)$$

proceeds by **maximum partial likelihood estimation**. The resulting estimator possesses ”nice” **asymptotic properties** (ABGK, 1993).

Problem: suppose **it is unknown**, for some (but not all) subjects, to **which sub-population they belong**.

(Observation for individual i)

Instead of $(T_i, \Delta_i, X_i, S_i)$*, one observes:*

$$\mathcal{O}_i = (T_i, \Delta_i, X_i, R_i, R_i S_i, W_i),$$

where

- $S_i \in \mathcal{K} = \{1, \ldots, K\}$ *is the random stratum indicator,*
- $W_i \in \mathbb{R}^m$ *is a random variable of covariate,*
- $R_i = 1(S_i$ *is observed*$)$, R_i *is independent of* $S_i | X_i, W_i$.

Statistical problem: estimate the semiparametric model

$$Y(t)\lambda_{0,k}(t)\exp(\beta_0' X), \quad k = 1, \ldots, K$$

from the independent **incomplete** observations $\mathcal{O}_i$, $i = 1, \ldots, n$.

Basic idea: **weighting the likelihood contribution of every incomplete observation** by replacing the actual sub-population indicator by an

(Approximated version)

$$R_i 1(S_i = k) + (1 - R_i)\mathbb{E}[1(S_i = k)|T_i, \Delta_i, X_i, W_i]$$

- simpler weights yield asymptotically biased estimators,
- need for an **estimator of baseline intensities**,
- **rules out martingale** arguments (integrands are no more predictable processes).

The likelihood function based on **observing** $\mathcal{O}_i, i = 1, \ldots, n$ is:

$$L_n(\theta) = \prod_{i=1}^{n} \left[\prod_{k=1}^{K} \left\{ \lambda_k(T_i)^{\Delta_i} \exp\left(\Delta_i \beta' X_i - e^{\beta' X_i} \Lambda_k(T_i) \right) \pi_{k,\gamma}(W_i) \right\}^{1(S_i = k)} \right]^{R_i}$$
$$\times \left[\sum_{k=1}^{K} \lambda_k(T_i)^{\Delta_i} \exp\left(\Delta_i \beta' X_i - e^{\beta' X_i} \Lambda_k(T_i) \right) \pi_{k,\gamma}(W_i) \right]^{1 - R_i}.$$

where $\pi_{k,\gamma}(W) := \mathbb{P}(S = k|W)$, $\Lambda_k = \int \lambda_k$, and $\theta = (\beta, \gamma, \Lambda_k(\cdot); k \in \mathcal{K})$.

- The maximum of $L_n(\theta)$ **is infinity** when the $\Lambda_k(\cdot)$ range across the space of absolutely continuous cumulative hazards.

  $\hookrightarrow$ consider functions $\Lambda_k(\cdot)$ with fixed values at the T_i, and let $\lambda_k(T_i)$ go to infinity for some T_i with $\Delta_i R_i 1(S_i = k) = 1$ or $\Delta_i(1 - R_i) = 1$.

- **Sieve estimation (NPMLE)**: a sequence of **embedded approximating spaces**

  $\hookrightarrow$ allow each $\Lambda_k(\cdot)$ to be **increasing right-continuous**.

Theorem

The NPMLE $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{\gamma}_n, \widehat{\Lambda}_{k,n}(\cdot); k \in \mathcal{K})$ *exists and satisfies*

$$\widehat{\Lambda}_{k,n}(t) = \int_0^t \sum_{i=1}^n \frac{Q(\mathcal{O}_i, k, \widehat{\theta}_n)}{\sum_{j=1}^n Q(\mathcal{O}_j, k, \widehat{\theta}_n) \exp(\widehat{\beta}_n' X_j) Y_j(s)} \, dN_i(s), \;\; k \in \mathcal{K},$$

where $Q(\mathcal{O}_i, k, \theta)$ *denotes the conditional expectation of* $1(S_i = k)$ *given* $\mathcal{O}_i$ *and the parameter value* θ*.*

Consistency

Corollary

For every $k \in \mathcal{K}$, the sequence $\widehat{\Lambda}_{k,n}(\tau)$ is almost surely bounded as n goes to infinity.

Theorem

Under some regularity conditions, $\|\widehat{\beta}_n - \beta_0\|$, $\|\widehat{\gamma}_n - \gamma_0\|$, $\sup_t |\widehat{\Lambda}_{k,n}(t) - \Lambda_{0,k}(t)|$ (for every $k \in \mathcal{K}$) converge to 0 almost surely.

Outline of the proof: Show that

1. every subsequence of n contains a further subsequence along which the NPMLE $\widehat{\theta}_n$ converges,
2. the limit of every convergent subsequence of $\widehat{\theta}_n$ is θ_0 (by identifiability).

Asymptotic normality

To **encompass** the discrete $\widehat{\Lambda}_{k,n}(\cdot)$ and their absolutely continuous limit $\Lambda_{0,k}(\cdot)$ within a same space, we use a suitable

(Re-parameterization of the model)

$$\theta : \mathbf{h} \mapsto \theta(\mathbf{h}) := h'_\beta \beta + h'_\gamma \gamma + \sum_{k=1}^{K} \int_0^\tau h_{\Lambda_k}(s)\, d\Lambda_k(s)$$

where $\mathbf{h} = (h_\beta, h_\gamma, h_{\Lambda_k}(\cdot); k \in \mathcal{K})$ belongs to a suitable space of directions $\mathbf{H}$ (typically, a subset of $\mathbb{R}^p \times \mathbb{R}^q \times (\text{BV}[0,\tau])^{\otimes K}$)

$\hookrightarrow$ consider θ as a **linear functional** on $\mathbf{H}$,

$\hookrightarrow$ consider the **parameter space** Θ as a **subset of** $l^\infty(\mathbf{H})$,

$\hookrightarrow$ treat $(\widehat{\beta}_n - \beta_0, \widehat{\gamma}_n - \gamma_0, \widehat{\Lambda}_{k,n} - \Lambda_{0,k}; k \in \mathcal{K})$ as **a random element in** $l^\infty(\mathbf{H})$.

- Function analytic approach (Murphy, 1995, Fang et al., 2005; Lu, 2008).
- Consider **one-dimensional submodels** $\widehat{\theta}_{n,\eta}$ passing through $\widehat{\theta}_n$, **differentiate** the log-likelihood wrt η, **evaluate** at $\eta = 0$:

(One-dimensional submodels)

$$\eta \mapsto \widehat{\theta}_{n,\eta} := (\widehat{\beta}_n + \eta h_\beta, \widehat{\gamma}_n + \eta h_\gamma, \int_0^{\cdot} \{1 + \eta h_{\Lambda_k}(s)\} d\widehat{\Lambda}_{k,n}(s); k \in \mathcal{K})$$

where $\mathbf{h} = (h_\beta, h_\gamma, h_{\Lambda_k}(\cdot); k \in \mathcal{K}) \in \mathbf{H}$. This yields the

(Score equation)

$$S_n(\widehat{\theta}_n)(\mathbf{h}) = 0$$

where $S_n : \Theta \subset l^\infty(\mathbf{H}) \longrightarrow l^\infty(\mathbf{H})$ is defined by:

(Score equation)

$$S_n(\widehat{\theta}_n)(\mathbf{h}) = \mathbb{P}_n \left[h'_\beta S_\beta(\widehat{\theta}_n) + h'_\gamma S_\gamma(\widehat{\theta}_n) + \sum_{k=1}^{K} S_{\Lambda_k}(\widehat{\theta}_n)(h_{\Lambda_k}) \right]$$

with

$$S_\beta(\theta) = \Delta X - \sum_{k=1}^{K} Q(\mathcal{O}, k, \theta) X \exp(\beta' X) \Lambda_k(T),$$

$$S_\gamma(\theta) = (S_{\gamma_1}(\theta)', \ldots, S_{\gamma_{K-1}}(\theta)')' \text{ with } S_{\gamma_k}(\theta) = W\left[Q(\mathcal{O}, k, \theta) - \pi_{k,\gamma}(W)\right],$$

$$S_{\Lambda_k}(\theta)(h_{\Lambda_k}) = Q(\mathcal{O}, k, \theta) \left[h_{\Lambda_k}(T)\Delta - \exp(\beta' X) \int_0^T h_{\Lambda_k}(s)\, d\Lambda_k(s) \right].$$

$\hookrightarrow$ **Z-estimation**

$\hookrightarrow$ check conditions for **asymptotic normality of Z-estimators**.

Theorem (Th. 3.3.1, Van der Vaart and Wellner, 1996)

Assume

1. *(Asymptotic distribution of the score)* $\sqrt{n}(S_n - S)(\theta)$ *converges weakly to a Gaussian process* Z *on* $l^\infty(\mathbf{H})$.
2. *(Differentiablity of the asymptotic score)*

$$\sqrt{n}(S(\widehat{\theta}_n) - S(\theta)) = -\sqrt{n}\dot{S}(\theta)(\widehat{\theta}_n - \theta) + o_p(1 + \sqrt{n}\|\widehat{\theta}_n - \theta\|)$$

   where $\dot{S}(\theta)$ *is a continuous linear operator.*

3. *(Invertibility of the information operator)* $\dot{S}(\theta)$ *is continuously invertible on its range.*

Then $\sqrt{n}(\widehat{\theta}_n - \theta)$ *converges weakly to* $\dot{S}(\theta)^{-1}Z$.

- The information operator has the form

$$h'_\beta \sigma_\beta(h) + h'_\gamma \sigma_\gamma(h) + \sum_{k=1}^{K} \int_0^\tau \sigma_{\Lambda_k}(h)(u) h_{\Lambda_k}(u)\, d\Lambda_{0,k}(u)$$

where $\sigma = (\sigma_\beta, \sigma_\gamma, \sigma_{\Lambda_k}; k \in \mathcal{K}) : \mathbf{H} \to \mathbf{H}$ has continuous inverse $\sigma^{-1} = (\sigma_\beta^{-1}, \sigma_\gamma^{-1}, \sigma_{\Lambda_k}^{-1}; k \in \mathcal{K})$.

Theorem (Asymptotic normality of $\widehat{\beta}_n$)

Let $(e_1, \ldots, e_p)$ be the canonical basis of $\mathbb{R}^p$. Under some regularity conditions,

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma_\beta)$$

where $\Sigma_\beta = (\sigma_\beta^{-1}(e_1, 0, 0; k \in \mathcal{K}), \ldots, \sigma_\beta^{-1}(e_p, 0, 0; k \in \mathcal{K}))$.

A consistent estimator of Σ_β is obtained by essentially replacing σ^{-1} by its empirical version.