Oracle inequalities for the Lasso for the conditional hazard rate in high-dimensional setting

Sarah Lemler
Thesis supervisors : Agathe Guilloux
Marie-Luce Taupin

Laboratoire "Statistique et Génome"

Séminaire de Statistiques de Strasbourg
Tuesday november 27th 2012
Outline

1. Framework and estimation procedure
   - Framework and models
   - Estimation procedure: the Lasso
   - Estimation criterion and loss function
   - Existing results for the Lasso in the Cox’s model

2. Non-asymptotic oracle inequalities for the Cox model when $\alpha_0$ is known
   - Empirical Bernstein’s inequality and choice of the weights
   - Slow non-asymptotic oracle inequality
   - Fast non-asymptotic oracle inequalities
   - Variable selection in the Cox model

3. Oracle inequalities for general conditional hazard rate
   - Empirical Bernstein’s inequalities and choice of the weights
   - Fast non-asymptotic oracle inequalities for the complete conditional hazard rate function
1 Framework and estimation procedure
   • Framework and models
   • Estimation procedure: the Lasso
   • Estimation criterion and loss function
   • Existing results for the Lasso in the Cox’s model

2 Non-asymptotic oracle inequalities for the Cox model when $\alpha_0$ is known

3 Oracle inequalities for general conditional hazard rate
Framework and estimation procedure

Context and notations

Context :

- **Problem**: to obtain a prognostic on the survival time adjusted on covariates in a high-dimensional setting
- **Example**:
  - $n = 191$ patients with follicular lymphoma
  - variable of interest: the survival time, that can be right-censored
  - covariates: clinical variables, 44929 levels of gene expression

**Goal**: to predict the survival from follicular lymphoma adjusted on covariates

Specific case of right censoring :

- For individual $i$, $i = 1, \ldots, n$
  - $T_i$ survival time,
  - $C_i$ censoring time,
  - $\delta_i = 1_{T_i \leq C_i}$ censoring indicator
- Observations: $X_i = \min(T_i, C_i)$, $\delta_i$ and $Z_i = (Z_{i,1}, \ldots, Z_{i,p})^T$
- $[0, \tau]$ time interval between the beginning and the end of the study
Framework and estimation procedure

Counting processes

Counting processes in the case of right censoring:

- \( Y_i(t) = \mathbb{1}_{\{X_i \geq t\}} \) the at-risk process
- \( N_i(t) = \mathbb{1}_{\{X_i \leq t, \delta_i = 1\}} \) counting process
- Observations: \((Z_i, N_i(t), Y_i(t), i = 1, ..., n, 0 \leq t \leq \tau)\)

Remark: all the results that follow are true for \( N_i \) a marked counting process and \( Y_i \) a predictable random process in \([0, 1]\)

Assumption 1. Let \( \Lambda_i(t) \) be the compensator of the process \( N_i(t) \). \( N_i \) satisfies the Aalen multiplicative intensity model: for all \( t \geq 0 \),

\[
\Lambda_i(t) = \int_0^t \lambda_0(s, Z_i)Y_i(s)ds,
\]

where \( \lambda_0 \) is an unknown nonnegative function called intensity.
Framework and estimation procedure

Models

- **Conditional hazard rate function of the survival time** $T_i$:
  \[
  \lambda_0(t, Z_i) = \lim_{dt \to 0} \frac{1}{dt} \mathbb{P}(t < T_i \leq t + dt | T_i > t, Z_i)
  \]
  characterizes the conditional distribution of $T_i$

- **Model 1**: The Cox model
  \[
  \lambda_0(t, Z_i) = \alpha_0(t) \exp(f_0(Z_i))
  \]
  $f_0$ the regression function and $\alpha_0$ the baseline hazard function assumed to be known

- **Model 2**: General case
  $\lambda_0(t, Z_i)$ does not rely on an underlying model

- **Goal**: estimation of the complete conditional hazard rate function by the best Cox model
Approximation of $\lambda_0$:

- Two dictionaries:
  \[ F_M = \{ f_1, \ldots, f_M \} \text{ where } f_j : \mathbb{R}^p \to \mathbb{R}, \quad ||f_j||_{n,\infty} = \max_{1 \leq i \leq n} |f_j(Z_i)| < \infty \]
  \[ G_N = \{ \theta_1, \ldots, \theta_N \} \text{ where } \theta_k : \mathbb{R}_+^* \to \mathbb{R}, \quad ||\theta_k||_{\infty} = \max_{t \in [0,\tau]} |\theta_k(t)| < \infty \]

- $\lambda_0$ assumed to be well approximated by a function of the form
  \[ \lambda_{\beta,\gamma}(t, Z_i) = \alpha_{\gamma}(t) e^{f_{\beta}(Z_i)}, \]
  where
  \[ \log \alpha_{\gamma} = \sum_{k=1}^{N} \gamma_k \theta_k \text{ and } f_{\beta} = \sum_{j=1}^{M} \beta_j f_j \]

Remark: particular case of the Cox model with $\alpha_0$ known
$\lambda_0$ well approximated by a function of the form $\lambda_{\beta}(t, Z_i) = \alpha_0(t) e^{f_{\beta}(Z_i)}$
Framework and estimation procedure
Choice of the penalization

High-dimensional setting :

$\text{→ risk of overfitting by minimizing a criterion } C_n(\lambda \beta) \text{ (least square criterion, opposite of the log-likelihood...)}$

$\text{→ necessity of a penalty}$

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \{C_n(\lambda \beta) + \text{pen}(\beta)\}$$

$\text{→ } \ell_0\text{-penalization :}$

$$\text{pen}(\beta) \propto ||\beta||_0 := \text{Card}\{\beta_j, \beta_j \neq 0\}$$

$\text{→ examples of } \ell_0\text{-penalized criteria : AIC and BIC}$

Problems associated with this penalty :

$\text{→ AIC and BIC not adapted to the case } p > n$

$\text{→ non-convex minimization problem}$
The Lasso procedure: minimization of an $\ell_1$-penalized criterion

Example: Lasso estimator of the regression function $f_0$ in the Cox model given $F_M = \{f_1, \ldots, f_M\}$

$$f_{\hat{\beta}_L} = \sum_{j=1}^{M} \hat{\beta}_{L,j} f_j$$

with

$$\hat{\beta}_L = \arg \min_{\beta \in \mathbb{R}^M} \{ C_n(\beta) + \Gamma \| \beta \|_1 \}$$

- $C_n$: criterion (least squares criterion, opposite of the log-likelihood, ...)
- $\Gamma$: tuning parameter

Advantages of this procedure:

- convex minimization problem $\Rightarrow$ computable in practice
- sparsity of the Lasso estimator $\Rightarrow$ results easily interpretable
**Framework and estimation procedure**

The weighted Lasso estimation procedure

**General case**: Estimation of $\beta$ and $\gamma$ simultaneously via a weighted Lasso procedure:

$$(\hat{\beta}_L, \hat{\gamma}_L) = \arg\min_{(\beta, \gamma) \in \mathbb{R}^M \times \mathbb{R}^N} \{C_n(\lambda_{\beta, \gamma}) + \text{pen}(\beta) + \text{pen}(\gamma)\},$$

with

$$\text{pen}(\beta) = \sum_{j=1}^{M} \omega_j |\beta_j| \quad \text{and} \quad \text{pen}(\gamma) = \sum_{k=1}^{N} \delta_k |\gamma_k|,$$

where $\omega_j$ and $\delta_k$ are positive data-driven weights defined via empirical Bernstein’s inequalities for martingales with jumps.

**Special case of the Cox model with $\alpha_0$ known**:

$$\hat{\beta}_L = \arg\min_{\beta \in \mathbb{R}^M} \{C_n(\lambda_\beta) + \text{pen}(\beta)\} \quad \text{with} \quad \text{pen}(\beta) = \sum_{j=1}^{M} \omega_j |\beta_j|.$$
Framework and estimation procedure
Choice of the estimation criterion and its associated loss function

1 Choice of the criterion

- **Log partial likelihood**: 
  - used to estimate $f_0$ without having to know $\alpha_0$
  - difficult to use
  - no linearity in $\beta$

- **Least squares criterion**: 
  - direct link with the empirical standard
  - two unknown parameters: $\alpha_0$ and $f_0$
  - no linearity in $\beta$

- **Log likelihood**: 
  - two unknown parameters: $\alpha_0$ and $f_0$

2 Choice of a loss function associated to the estimation criterion
Framework and estimation procedure

Estimation criterion and loss function

**Estimation criterion**: the total empirical log-likelihood

\[
C_n(\lambda_\beta, \gamma) = -\frac{1}{n} \sum_{i=1}^{n} \left\{ \int_0^\tau \log \lambda_{\beta, \gamma}(t, Z_i) dN_i(t) - \int_0^\tau \lambda_{\beta, \gamma}(t, Z_i) Y_i(t) dt \right\}
\]

**Loss function**: the empirical Kullback divergence

\[
\tilde{K}_n(\lambda_0, \lambda_\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left( \log \lambda_0(t, Z_i) - \log \lambda_{\beta, \gamma}(t, Z_i) \right) \lambda_0(t, Z_i) Y_i(t) dt \\
- \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left( \lambda_0(t, Z_i) - \lambda_{\beta, \gamma}(t, Z_i) \right) Y_i(t) dt
\]
Framework and estimation procedure
Relation between the empirical Kullback divergence and a weighted empirical norm

Weighted empirical norm: for all function $h$ on $[0, \tau] \times \mathbb{R}^p$

$$
\|h\|_{n,\Lambda} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \int_0^\tau (h(t, Z_i))^2 d\Lambda_i(t)}
$$

Assumption 2. There exists $\mu > 0$ such that for all $(\beta, \gamma) \in \mathbb{R}^M \times \mathbb{R}^N$

$$
\| \log \lambda_{\beta,\gamma} - \log \lambda_0 \|_{n,\infty} \leq \mu
$$

Proposition
Under Assumption 2, there exist $\mu', \mu'' > 0$ such that $\forall (\beta, \gamma) \in \mathbb{R}^M \times \mathbb{R}^N$

$$
\mu' \| \log \lambda_{\beta,\gamma} - \log \lambda_0 \|_{n,\Lambda}^2 \leq \tilde{K}_n(\lambda_0, \lambda_{\beta,\gamma}) \leq \mu'' \| \log \lambda_{\beta,\gamma} - \log \lambda_0 \|_{n,\Lambda}^2
$$
Definition of the oracle inequality in the case of the Cox model: inequality that compares the performances of an estimator \( f_{\hat{\beta}_L} \) obtained without a priori knowledge of the true function \( f_0 \), to those of the best approximation \( f_\beta \) of \( f_0 \) in the dictionary \( \mathbb{F}_M \)

\[
\|f_{\hat{\beta}_L} - f_0\|_{n,\Lambda}^2 \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M} \{\|f_\beta - f_0\|_{n,\Lambda}^2 + T_{\zeta,n,M}(\beta)\},
\]

where \( T_{\zeta,n,M}(\beta) \) is a variance term of order \( \sqrt{\log M/n} \) or \( \log M/n \) for a slow or fast rate of convergence respectively.
Existing results for the Lasso in the Cox’s model

Asymptotic results

Two type of results:

▶ Estimation results: \( ||\hat{\beta}_L - \beta_0|| \), (in this case, \( f_0(Z_i) = \beta_0^T Z_i \))
▶ Prediction results: \( ||f_{\hat{\beta}_L} - f_0|| \)

Asymptotic estimation results:

\[
\hat{\beta}_L = \arg \min_{\beta \in \mathbb{R}^p} \left\{ -l_n^*(\beta) + \lambda_n \sum_{j=1}^{p} |\beta_j| \right\},
\]

where \( l_n^*(\beta) \) is the log partial likelihood and \( \lambda_n \) a regularization parameter.

Theorem [Bradic et al. (2010)]

Let \( \hat{\beta}^1 \) be a subvector of \( \beta \) formed by all nonzero components. Under some assumptions, we have:

\[
||\hat{\beta}_L^1 - \beta_0^1||_2 = O_P(\sqrt{J(\beta_0)\lambda_n}),
\]

where \( \lambda_n > n^{-0.5+\alpha} \), \( \alpha > 0 \) and \( |J(\beta_0)| = \text{Card}\{j \in \{1, ..., p\}; \beta_{0j} \neq 0\} \).
Existing results for the Lasso in the Cox’s model

Non asymptotic results : oracle inequalities

- Non-asymptotic oracle inequality for the excess risk

**Theorem [Kong and Nan (2012)]**

Under some assumptions and with a certain probability, we have

$$\varepsilon(f_{\hat{\beta}_L}) \leq (1 + \zeta) \inf_{\beta} \{ \varepsilon(f_{\beta}) + T_{\zeta,n,M}(\beta) \},$$

where $\varepsilon(f)$ is the excess risk of $f$ and $T_{\zeta,n,M}(\beta)$ is a variance term.

- Non-asymptotic oracle inequality when $f_0 \in F_M$

**Theorem [Bradic and Song (2012)]**

Under some assumptions and with a certain probability, we have

$$\|f_{\hat{\beta}_L} - f_0\|_{n,b_{\hat{\beta}}}^2 \leq \min_{\beta \in \mathbb{R}^M} \{ \|f_{\beta} - f_0\|_{n,b^*}^2 + T_{n,M}(\beta) \},$$

for a certain empirical weighted norm.
Outline

1  Framework and estimation procedure

2  Non-asymptotic oracle inequalities for the Cox model when $\alpha_0$ is known
   - Empirical Bernstein’s inequality and choice of the weights
   - Slow non-asymptotic oracle inequality
   - Fast non-asymptotic oracle inequalities
   - Variable selection in the Cox model

3  Oracle inequalities for general conditional hazard rate
Oracle inequalities for the Cox model when $\alpha_0$ is known

Model and notations

**Model 1**: $\lambda_0(t, Z_i) = \alpha_0(t) e^{f_0(Z_i)}$, with $\alpha_0$ known

**Lasso estimator**:

$$\lambda_{\hat{\beta}_L}(t, Z_i) = \alpha_0(t) e^{\hat{f}_{\hat{\beta}_L}(Z_i)}$$

with

$$\hat{\beta}_L = \arg \min_{\beta \in \mathbb{R}^M} \{ C_n(\lambda_\beta) + \text{pen}(\beta) \} \text{ and } \text{pen}(\beta) = \sum_{j=1}^{M} \omega_j |\beta_j|$$

$$C_n(\lambda_\beta) = -\frac{1}{n} \sum_{i=1}^{n} \left\{ \int_0^\tau \log \lambda_\beta(t, Z_i) dN_i(t) - \int_0^\tau \lambda_\beta(t, Z_i) Y_i(t) dt \right\}$$
Oracle inequalities for the Cox model when $\alpha_0$ is known

Empirical Bernstein’s inequality

By definition, for all $\beta \in \mathbb{R}^M$,

$$C_n(\lambda_{\hat{\beta}_L}) + \text{pen}(\hat{\beta}_L) \leq C_n(\lambda_{\beta}) + \text{pen}(\beta),$$

$$C_n(\lambda_{\hat{\beta}_L}) - C_n(\lambda_{\beta}) = \tilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) - \tilde{K}_n(\lambda_0, \lambda_{\beta}) + \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \eta_{n,\tau}(f_j)$$

where

$$\eta_{n,\tau}(f_j) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} f_j(Z_i) dM_i(s)$$
Oracle inequalities for the Cox model when $\alpha_0$ is known

Empirical Bernstein’s inequality

By definition, for all $\beta \in \mathbb{R}^M$,

$$ C_n(\lambda_{\hat{\beta}_L}) + \text{pen}(\hat{\beta}_L) \leq C_n(\lambda_\beta) + \text{pen}(\beta), $$

$$ C_n(\lambda_{\hat{\beta}_L}) - C_n(\lambda_\beta) = \tilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) - \tilde{K}_n(\lambda_0, \lambda_\beta) + \sum_{j=1}^M (\hat{\beta}_L - \beta)_j \eta_{n,\tau}(f_j) $$

where

$$ \eta_{n,\tau}(f_j) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau f_j(Z_i) dM_i(s) $$

Comparison with the additive regression model: $Y_i = f_0(Z_i) + W_i$

$$ \|f_{\hat{\beta}_L} - f_0\|^2_n \leq \|f_\beta - f_0\|^2_n + \sum_{j=1}^M (\hat{\beta}_L - \beta)_j V_n(f_j) + \text{pen}(\beta) - \text{pen}(\hat{\beta}_L) $$

with

$$ V_n(f_j) = \frac{1}{n} \sum_{i=1}^n f_j(Z_i) W_i $$

Sarah Lemler (Statistique et Génome)
Oracle inequalities for the Cox model when $\alpha_0$ is known

Empirical Bernstein’s inequality

Standard Bernstein’s inequality for the additive regression case: If the $W_i$ are i.i.d, centered and bounded for $i = 1, \ldots, n$,

$$\mathbb{P}(|V_n(f_j)| \geq \sqrt{2vx + cx}) \leq 2e^{-x}$$

with $v = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[(f_j(Z_i)W_i)^2]$ and $c = \frac{1}{3n} \max_{1 \leq i \leq n} |f_j(Z_i)W_i|$
Oracle inequalities for the Cox model when $\alpha_0$ is known

Empirical Bernstein’s inequality

Standard Bernstein’s inequality for the additive regression case: If the $W_i$ are i.i.d, centered and bounded for $i = 1, \ldots, n$,

$$\mathbb{P}(|V_n(f_j)| \geq \sqrt{2vx + cx}) \leq 2e^{-x}$$

with $v = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[(f_j(Z_i)) W_i]^2$ and $c = \frac{1}{3n} \max_{1 \leq i \leq n} |f_j(Z_i) W_i|

Problem in our case: martingale with jumps and with a predictable variation not observable:

- Predictable variation of $\eta_{n,t}(f_j)$:

$$V_{n,t}(f_j) = n < \eta_n(f_j) >_t = \frac{1}{n} \sum_{i=1}^{n} \int_0^t (f_j(Z_i))^2 \alpha_0(s) e^{f_0(Z_i)} Y_i(s) ds$$

- Optional variation of $\eta_{n,t}(f_j)$:

$$\hat{V}_{n,t}(f_j) = n[\eta_n(f_j)]_t = \frac{1}{n} \sum_{i=1}^{n} \int_0^t (f_j(Z_i))^2 dN_i(s)$$
Oracle inequalities for the Cox model when $\alpha_0$ is known

Empirical Bernstein’s inequality

**Theorem**

For any $x > 0$ and $c_1$, $c_2$, $c_3$ some positive constants

\[
P\left[ |\eta_{n,t}(f_j)| \geq \left( c_1 \sqrt{\frac{x + \hat{\ell}_{n,x}(f_j)}{n}} \hat{V}_{n,t}(f_j) + c_2 \frac{x + 1 + \hat{\ell}_{n,x}(f_j)}{n} \right) \|f_j\|_{n,\infty} \right] \leq c_3 e^{-x}
\]

where

\[
\hat{\ell}_{n,x}(f_j) = 2 \log \log \left( \frac{6en\hat{V}_{n,t}(f_j) + 56ex\|f_j\|_{n,\infty}^2}{24\|f_j\|_{n,\infty}^2} \lor e \right)
\]
Oracle inequalities for the Cox model when $\alpha_0$ is known

Empirical Bernstein’s inequality

**Theorem**

For any $x > 0$ and $c_1, c_2, c_3$ some positive constants

$$\mathbb{P}\left[|\eta_{n,t}(f_j)| \geq \left(c_1 \sqrt{\frac{x + \hat{\ell}_{n,x}(f_j)}{n}} \hat{V}_{n,t}(f_j) + c_2 \frac{x + 1 + \hat{\ell}_{n,x}(f_j)}{n} ||f_j||_{n,\infty}\right) ||f_j||_{n,\infty} \right] \leq c_3 e^{-x}$$

where

$$\hat{\ell}_{n,x}(f_j) = 2 \log \log \left( \frac{6en\hat{V}_{n,t}(f_j) + 56ex||f_j||_{n,\infty}^2}{24||f_j||_{n,\infty}^2} \vee e \right)$$

**Choice of the weights** : data-driven weights for $j = 1, \ldots, M$

$$\omega_j = \left(c_1, \varepsilon \sqrt{\frac{x + \log M + \hat{\ell}_{n,x}(f_j)}{n}} \hat{V}_{n,\tau}(f_j) + c_2, \varepsilon \frac{x + 1 + \log M + \hat{\ell}_{n,x}(f_j)}{n} \right) ||f_j||_{n,\infty}$$

$$\Rightarrow \omega_j \propto \sqrt{\frac{\log M}{n}} \hat{V}_{n,\tau}(f_j)$$
Oracle inequalities for the Cox model when $\alpha_0$ is known

Slow non-asymptotic oracle inequality

$$\tilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) \leq \tilde{K}_n(\lambda_0, \lambda_{\beta}) + \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \eta_{n,\tau}(f_j) + \sum_{j=1}^{M} \omega_j |\beta_j| - \sum_{j=1}^{M} \omega_j |\hat{\beta}_{L,j}|$$

On $A = \bigcap_{j=1}^{M} \{ |\eta_{n,\tau}(f_j)| \leq \omega_j \}$, we have

$$\left| \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \eta_{n,\tau}(f_j) \right| \leq \sum_{j=1}^{M} \omega_j |(\hat{\beta}_L - \beta)_j|,$$

and

$$\mathbb{P}(A^c) \leq \sum_{j=1}^{M} \mathbb{P}(|\eta_{n,\tau}(f_j)| > \omega_j) \leq c_3 e^{-x}$$

Theorem: Slow non-asymptotic oracle inequality for the Cox model

With probability larger than $1 - c_3 e^{-x}$, we have

$$\tilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) \leq \inf_{\beta \in \mathbb{R}^M} \left( \tilde{K}_n(\lambda_0, \lambda_{\beta}) + 2 \text{pen}(\beta) \right) \text{ with } \text{pen}(\beta) \propto \|\beta\|_1 \sqrt{\frac{\log M}{n}}$$
Oracle inequalities for the Cox model when $\alpha_0$ is known

The restricted eigenvalue condition $\text{RE}(s, c_0)$

Notations:

$$X = (f_j(Z_i))_{1 \leq i \leq n, 1 \leq j \leq M}$$

$$G_n = \frac{1}{n} X^T C X \text{ with } C = (\text{diag}(\Lambda_i(\tau)))_{1 \leq i \leq n}$$

$$J(\beta) = \{j \in \{1, ..., M\} : \beta_j \neq 0\} \text{ and } |J(\beta)| = \text{Card}\{J(\beta)\}$$

Restricted eigenvalue condition $\text{RE}(s, c_0)$: For some integer $s \in \{1, ..., M\}$ and a constant $c_0 > 0$, we assume that $G_n$ satisfies:

$$0 < \kappa(s, c_0) = \min_{J \subset \{1, ..., M\}, |J| \leq s} \min_{b \in \mathbb{R}^M \setminus \{0\}, \|b_Jc\|_1 \leq c_0 \|b_J\|_1} \frac{(b^T G_n b)^{1/2}}{\|b_J\|_2}$$
Oracle inequalities for the Cox model when $\alpha_0$ is known

Fast non-asymptotic oracle inequalities

**Assumptions:**

1) $\mathbb{E}\left(s, \left(3 + \frac{4}{\zeta}\right) \max_{1 \leq j \leq M} \frac{\omega_j}{\min_{1 \leq j \leq M} \omega_j}\right)$

2) $\|f_j\|_{n,\infty} = \max_{1 \leq i \leq n} |f_j(Z_i)| < \infty, \forall j \in \{1, ..., M\}$

3) $\exists \mu > 0, \|f_\beta - f_0\|_{n,\Lambda} \leq \mu, \forall \beta \in \mathbb{R}^M$

**Theorem:** Non-asymptotic oracle inequalities for the Cox model

Let $A > 0$ and $x > 0$ fixed. Under these assumptions, we have with a probability larger than $1 - Ae^{-x}$,

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, |J(\beta)| \leq s} \left\{ \widetilde{K}_n(\lambda_0, \lambda_\beta) + C(\zeta, \mu') \frac{|J(\beta)|}{\kappa^2} \left( \max_{1 \leq j \leq M} \omega_j \right)^2 \right\}$$

$$\|f_{\hat{\beta}_L} - f_0\|_{n,\Lambda}^2 \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, |J(\beta)| \leq s} \left\{ \|f_\beta - f_0\|_{n,\Lambda}^2 + C'(\zeta, \mu') \frac{|J(\beta)|}{\kappa^2} \left( \max_{1 \leq j \leq M} \omega_j \right)^2 \right\}$$
Oracle inequalities for the Cox model when \( \alpha_0 \) is known

Fast non-asymptotic oracle inequalities

Assumptions:
1) \( \mathbb{E}(s, (3 + \frac{4}{\zeta}) \frac{\max_{1 \leq j \leq M} \omega_j}{\min_{1 \leq j \leq M} \omega_j}) \)

2) \( \|f_j\|_{n, \infty} = \max_{1 \leq i \leq n} |f_j(Z_i)| < \infty, \forall j \in \{1, \ldots, M\} \)

3) \( \exists \mu > 0, \|f_\beta - f_0\|_{n, \Lambda} \leq \mu, \forall \beta \in \mathbb{R}^M \)

Theorem: Non-asymptotic oracle inequalities for the Cox model

Let \( A > 0 \) and \( x > 0 \) fixed. Under these assumptions, we have with a probability larger than \( 1 - Ae^{-x} \),

\[
\tilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}) \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, |J(\beta)| \leq s} \left\{ \tilde{K}_n(\lambda_0, \lambda_\beta) + C(\zeta, \mu') \frac{|J(\beta)|}{\kappa^2} \left( \max_{1 \leq j \leq M} \omega_j \right)^2 \right\}
\]

\[
\|f_{\hat{\beta}_L} - f_0\|^2_{n, \Lambda} \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, |J(\beta)| \leq s} \left\{ \|f_\beta - f_0\|^2_{n, \Lambda} + C'(\zeta, \mu') \frac{|J(\beta)|}{\kappa^2} \left( \max_{1 \leq j \leq M} \omega_j \right)^2 \right\}
\]

\( \hookrightarrow \) Fast non-asymptotic oracle inequalities of order \( \log M/n \)
Oracle inequalities for the Cox model when $\alpha_0$ is known

 Variable selection in the Cox model

Context of the variable selection :

- $f_0(Z_i) = \beta_0^T Z_i$
- $M = p$
- $f_j(Z_i) = Z_{i,j}$ for all $j \in \{1, ..., p\}$
- $X = (Z_{i,j})_{1 \leq i \leq n; 1 \leq j \leq p}$ design matrix

Notation :

$$b_0 = 4 \frac{\max \omega_j}{\min \omega_j} - 1$$

Theorem

Under $RE(s, b_0)$ and the two other previous assumptions, with a probability larger than $1 - Ae^{-x}$, we have

$$\|X(\hat{\beta}_L - \beta_0)\|_{n,\Lambda}^2 \leq \frac{4}{\mu' \kappa^2} \left|J(\beta_0)\right| \left(\max_{1 \leq j \leq p} \omega_j\right)^2$$

$$\|\hat{\beta}_L - \beta_0\|_1 \leq \frac{1 + b_0}{\mu'} \frac{|J(\beta_0)|}{\kappa^2} \max_{1 \leq j \leq p} \omega_j$$
Outline

1. Framework and estimation procedure
2. Non-asymptotic oracle inequalities for the Cox model when $\alpha_0$ is known
3. Oracle inequalities for general conditional hazard rate
   - Empirical Bernstein’s inequalities and choice of the weights
   - Fast non-asymptotic oracle inequalities for the complete conditional hazard rate function
Oracle inequalities for general conditional hazard rate

Empirical Bernstein’s inequalities

Model 2: \( \lambda_0(t, Z_i) \) estimated by \( \lambda_{\hat{\beta}_L, \hat{\gamma}_L}(t, Z_i) = \alpha \hat{\gamma}_L(t) e^{f \hat{\beta}_L(Z_i)} \)

By definition, for all \((\beta, \gamma) \in \mathbb{R}^M \times \mathbb{R}^N\),

\[
C_n(\lambda_{\hat{\beta}_L, \hat{\gamma}_L}) + \text{pen}(\hat{\beta}_L) + \text{pen}(\hat{\gamma}_L) \leq C_n(\lambda_\beta, \gamma) + \text{pen}(\beta) + \text{pen}(\gamma),
\]

\[
C_n(\lambda_{\hat{\beta}_L, \hat{\gamma}_L}) - C_n(\lambda_\beta, \gamma) = \tilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L, \hat{\gamma}_L}) - \tilde{K}_n(\lambda_0, \lambda_\beta, \gamma)
\]

\[
+ \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \eta_{n, \tau}(f_j) + \sum_{k=1}^{N} (\hat{\gamma}_L - \gamma)_k \nu_{n, \tau}(\theta_k)
\]

where

\[
\eta_{n, \tau}(f_j) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} f_j(Z_i) dM_i(s),
\]

\[
\nu_{n, \tau}(\theta_k) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \theta_k(s) dM_i(s), \quad \hat{R}_{n, t}(\theta_k) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} (\theta_k(s))^2 dN_i(s)
\]
Oracle inequalities for general conditional hazard rate

Empirical Bernstein’s inequalities

**Theorem**

For any \( x > 0, \) \( y > 0 \) and \( c_1, c_2, c_3, c'_1, c'_2, c'_3 \) some positive constants

\[
P \left[ |\eta_{n,t}(f_j)| \geq \left( c_1 \sqrt{\frac{x + \hat{\ell}_{n,x}(f_j)}{n} \hat{V}_{n,t}(f_j) + c_2 \frac{x + 1 + \hat{\ell}_{n,x}(f_j)}{n}} \right) \|f_j\|_{n,\infty} \right] \leq c_3 e^{-x}
\]

\[
P \left[ |\nu_{n,t}(\theta_k)| \geq \left( c'_1 \sqrt{\frac{y + \hat{\ell}'_{n,y}(\theta_k)}{n} \hat{R}_{n,t}(\theta_k) + c'_2 \frac{y + 1 + \hat{\ell}'_{n,y}(\theta_k)}{n}} \right) \|\theta_k\|_{\infty} \right] \leq c'_3 e^{-y}
\]

where

\[
\hat{\ell}_{n,x}(f_j) = 2 \log \log \left( \frac{6en\hat{V}_{n,t}(f_j) + 56ex\|f_j\|_{n,\infty}^2}{24\|f_j\|_{n,\infty}^2} \vee e \right)
\]

\[
\hat{\ell}'_{n,y}(\theta_k) = 2 \log \log \left( \frac{6en\hat{R}_{n,t}(\theta_k) + 56ey\|\theta_k\|_{\infty}^2}{24\|\theta_k\|_{\infty}^2} \vee e \right)
\]
Oracle inequalities for general conditional hazard rate

Choice of the weights

Data-driven weights: for $j = 1, \ldots, M$ and $k = 1, \ldots, N$

$$
\omega_j = \left( c_{1,\epsilon} \sqrt{\frac{x + \log M + \hat{\ell}_{n,x}(f_j)}{n}} \hat{V}_{n,\tau}(f_j) + c_{2,\epsilon} \frac{x + 1 + \log M + \hat{\ell}_{n,x}(f_j)}{n} \right) \| f_j \|_{n,\infty}
$$

$$
\Rightarrow \omega_j \propto \sqrt{\frac{\log M}{n}} \hat{V}_{n,\tau}(f_j)
$$

and

$$
\delta_k = \left( c'_{1,\epsilon'} \sqrt{\frac{y + \log N + \hat{\ell}'_{n,x}(\theta_k)}{n}} \hat{R}_{n,\tau}(\theta_k) + c'_{2,\epsilon'} \frac{y + 1 + \log N + \hat{\ell}'_{n,x}(\theta_k)}{n} \right) \| \theta_k \|_{\infty}
$$

$$
\Rightarrow \delta_k \propto \sqrt{\frac{\log N}{n}} \hat{R}_{n,\tau}(\theta_k)
$$
Notations:

\[ \tilde{X}(t) = \begin{bmatrix} X \\ \theta_1(t) \ldots \theta_N(t) \end{bmatrix} \in \mathbb{R}^{n \times (M+N)}, \quad X = (f_j(Z_i))_{1 \leq i \leq n, 1 \leq j \leq M} \]

\[ \tilde{G}_n = \frac{1}{n} \int_0^\tau \tilde{X}(t)^T \tilde{C}(t) \tilde{X}(t) dt, \quad \tilde{C} = (\text{diag}(\lambda_0(t, Z_i) Y_i(t)))_{1 \leq i \leq n} \]

Restricted eigenvalue condition \( \text{RE}(s, c_0) \) for the matrix \( \tilde{G}_n \): For some integer \( s \in \{1, \ldots, M + N\} \) and a constant \( c_0 > 0 \), we assume that \( \tilde{G}_n \) satisfies

\[ 0 < \tilde{\kappa}(s, c_0) = \min_{J \subseteq \{1, \ldots, M+N\}, |J| \leq s} \min_{b \in \mathbb{R}^{M+N} \setminus \{0\}, \|b_Jc\|_1 \leq c_0} \frac{(b^T \tilde{G}_n b)^{1/2}}{\|b_J\|_2}. \]
Assumptions:

1) $\mathbb{R} E(s, r_0)$ with $r_0 = \left(3 + \frac{8}{\zeta} \max \left(\sqrt{|J(\beta)|}, \sqrt{|J(\gamma)|}\right)\right) \frac{\max}{\min} \{\omega_j, \delta_k\}_{1 \leq j \leq M, 1 \leq k \leq N}$

2) $\|f_j\|_{n,\infty} < \infty, \forall j \in \{1, \ldots, M\}$ and $\|\theta_k\|_{\infty} < \infty, \forall k \in \{1, \ldots, N\}$

3) $\exists \mu > 0, \|\log \lambda_{\beta,\gamma} - \log \lambda_0\|_{n,\Lambda} \leq \mu, \forall (\beta, \gamma) \in \mathbb{R}^M \times \mathbb{R}^N$
Theorem : Non-asymptotic oracle inequalities for the hazard rate function

Let $B > 0$ and $z > 0$ fixed. Under Assumptions 1,2 and 3, with probability larger than $1 - Be^{-z}$, we have

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}, \hat{\gamma}_L) \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, \gamma \in \mathbb{R}^N} \left\{ \widetilde{K}_n(\lambda_0, \lambda_{\beta, \gamma}) + \widetilde{C}(\zeta, \mu') \frac{\max(|J(\beta)|, |J(\gamma)|)}{\tilde{\kappa}^2} \max \left\{ \omega_j^2, \delta_k^2 \right\} \right\}$$

$$\| \log \lambda_0 - \log \lambda_{\hat{\beta}_L, \hat{\gamma}_L} \|^2_{n, \Lambda} \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, \gamma \in \mathbb{R}^N} \left\{ \| \log \lambda_0 - \log \lambda_{\beta, \gamma} \|^2_{n, \Lambda} + \widetilde{C}'(\zeta, \mu') \frac{\max(|J(\beta)|, |J(\gamma)|)}{\tilde{\kappa}^2} \max \left\{ \omega_j^2, \delta_k^2 \right\} \right\}$$
Oracle inequalities for general conditional hazard rate

Fast non-asymptotic oracle inequalities for the complete conditional hazard rate function

Theorem: Non-asymptotic oracle inequalities for the hazard rate function

Let $B > 0$ and $z > 0$ fixed. Under Assumptions 1, 2 and 3, with probability larger than $1 - B e^{-z}$, we have

$$
\tilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L}, \hat{\gamma}_L) \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, \gamma \in \mathbb{R}^N} \left\{ \tilde{K}_n(\lambda_0, \lambda_{\beta, \gamma}) + \tilde{C}(\zeta, \mu') \max(|J(\beta)|, |J(\gamma)|) \frac{\max\{\omega_j^2, \delta_k^2\}}{\tilde{\kappa}^2} \right\}
$$

$$
\left\| \log \lambda_0 - \log \lambda_{\hat{\beta}_L, \hat{\gamma}_L} \right\|_{n, \Lambda}^2 \leq (1 + \zeta) \inf_{\beta \in \mathbb{R}^M, \gamma \in \mathbb{R}^N} \left\{ \left\| \log \lambda_0 - \log \lambda_{\beta, \gamma} \right\|_{n, \Lambda}^2 + \tilde{C}'(\zeta, \mu') \frac{\max(|J(\beta)|, |J(\gamma)|)}{\tilde{\kappa}^2} \max\{\omega_j^2, \delta_k^2\} \right\}
$$

with

$$
\left( \max_{1 \leq j \leq M, 1 \leq k \leq N} \{\omega_j, \delta_k\} \right)^2 \propto \max \left\{ \frac{\log M}{n}, \frac{\log N}{n} \right\},
$$
Conclusion and perspectives

Conclusion: We obtain

▶ a fast non-asymptotic oracle inequality for the Cox model when $\alpha_0$ is assumed to be known

▶ a fast non-asymptotic oracle inequality for a general conditional hazard rate function that does not rely on an underlying model

→ allows to predict the survival time throughout the conditional hazard rate function in a high dimensional setting

▶ an empirical Bernstein’s inequality that hold true for martingales with jumps, when the predictable variation is not observable

Perspectives:

▶ to obtain a fast non-asymptotic oracle inequality for a weighted Lasso estimator in the Cox model without any assumption on $\alpha_0$

▶ to compare the existing algorithms for the Lasso in the Cox model by introducing the obtained weights
References


