# High-dimensional estimation of counting process intensities

Sarah Lemler
Thesis supervisors : Agathe Guilloux
Marie-Luce Taupin

Laboratoire "Statistique et Génome"

Séminaire TEST

# Outline

1. Oracle Inequalities for the Lasso in the high-dimensional Aalen multiplicative intensity model
   - Framework and model
   - Estimation procedure in the case of an additive regression model
   - Estimation procedure
   - Slow non-asymptotic oracle inequality on the intensity
   - Fast non-asymptotic oracle inequalities on the intensity
   - Comparison with the existing results for the Cox model

# Outline

1. Oracle Inequalities for the Lasso in the high-dimensional Aalen multiplicative intensity model
   - Framework and model
   - Estimation procedure in the case of an additive regression model
   - Estimation procedure
   - Slow non-asymptotic oracle inequality on the intensity
   - Fast non-asymptotic oracle inequalities on the intensity
   - Comparison with the existing results for the Cox model

**Context :**

- Example :
  - $n = 191$ patients with follicular lymphoma
  - variable of interest : the survival time, that can be right-censored
  - covariates : clinical variables, $\mathbf{44929}$ levels of gene expression

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

**Specific case of right censoring :**

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

**Counting processes in the case of right censoring :**

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

Let $\Lambda_i(t)$ be the compensator of $N_i(t)$, so that

$$M_i(t) = N_i(t) - \Lambda_i(t) \in \mathcal{M}^2_{loc}.$$

**Assumption 1.** *$N_i$ satisfies the Aalen multiplicative intensity model : for all $t \geq 0$,*

$$\Lambda_i(t) = \int_0^t \lambda_0(s, \boldsymbol{Z_i}) Y_i(s) ds,$$

*where $\lambda_0$ is an unknown nonnegative function called intensity*

- **Conditional hazard rate function of the survival time $T_i$ :**

$$\lambda_0(t, \boldsymbol{Z_i}) = \lim_{dt \to 0} \frac{1}{dt} \mathbb{P}(t < T_i \leq t + dt | T_i > t, \boldsymbol{Z_i})$$

$\hookrightarrow$ characterizes the conditional distribution of $T_i$

Example : The Cox model

$$\lambda_0(t, \boldsymbol{Z_i}) = \alpha_0(t) \exp(f_0(\boldsymbol{Z_i}))$$

$f_0$ the regression function and $\alpha_0$ the baseline hazard function

- General case
  $\lambda_0(t, \boldsymbol{Z_i})$ does not rely on an underlying model

$\hookrightarrow$ **Goal :** estimation of the complete conditional hazard rate function by the "best" Cox model

# Estimation procedure for the additive regression model I

**The additive regression model [see Bickel, Ritov and Tsybakov (2009)] :**

$$Y_i = f_0(\boldsymbol{Z_i}) + W_i, \quad i = 1, ..., n$$

- $f_0$ unknown function to be estimated,
- $\boldsymbol{Z_i}$ fixed elements of $\mathbb{R}^p$,
- $W_i$ independent $\mathcal{N}(0, \sigma^2)$ random variables

**Approximation of $f_0$ in the additive regression model**

- Dictionary : $\mathbb{F}_M = \{f_1, ..., f_M\}$, $f_j : \mathbb{R}^p \to \mathbb{R}$
- Candidates for the estimation of $f_0$ : $f_{\boldsymbol{\beta}} = \sum_{j=1}^{M} \beta_j f_j$, for $\boldsymbol{\beta} \in \mathbb{R}^M$
- Least squares criterion : $C_n(f_{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - f_{\boldsymbol{\beta}}(\boldsymbol{Z_i}))^2$
- Empirical norm : $||g||_n^2 = \frac{1}{n} \sum_{i=1}^{n} g^2(\boldsymbol{Z_i})$

# Estimation procedure for the additive regression model II

The Lasso procedure : minimization of an $\ell_1$-penalized criterion

Lasso estimator of the regression function $f_0$ in the additive regression model given $\mathbb{F}_M = \{f_1, ..., f_M\}$ :

$$f_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}} = \sum_{j=1}^{M} \hat{\beta}_{L,j} f_j, \quad \text{with} \quad \hat{\boldsymbol{\beta}}_{\boldsymbol{L}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^M} \{C_n(f_{\boldsymbol{\beta}}) + \mathrm{pen}(\boldsymbol{\beta})\},$$

$\mathrm{pen}(\boldsymbol{\beta}) = 2r \sum_{j=1}^{M} ||f_j||_n |\beta_j|$ and $r > 0$ some tuning constant.

**Theorem : Oracle inequality [Bickel, Ritov and Tsybakov (2009)]**

For $r = A\sigma\sqrt{\log M/n}$, $A > 2\sqrt{2}$, with probability at least $1 - M^{1-A^2/8}$,

$$||f_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}} - f_0||_n^2 \leq (1+\zeta) \inf_{\boldsymbol{\beta} \in \mathbb{R}^M} \{||f_{\boldsymbol{\beta}} - f_0||_n^2 + T_{\zeta,n,M}\},$$

where $T_{\zeta,nM}$ is a variance term of order $\sqrt{\log M/n}$ or $\log M/n$.

**The multiplicative Aalen intensity model :**

$$\mathrm{d}N_i(t) = \lambda_0(t, \boldsymbol{Z_i})Y_i(t)\mathrm{d}t + \mathrm{d}M_i(t), \quad i = 1, ..., n$$

where $M_i = N_i - \Lambda_i \in \mathcal{M}^2_{loc}$.

**Approximation of $\lambda_0$ in the multiplicative Aalen intensity model :**

- Two dictionaries :
  $\mathbb{F}_M = \{f_1, ..., f_M\}$ where $f_j : \mathbb{R}^p \to \mathbb{R}$, $||f_j||_{n,\infty} = \max_{1 \le i \le n} |f_j(\boldsymbol{Z_i})| < \infty$

  $\mathbb{G}_N = \{\theta_1, ..., \theta_N\}$ where $\theta_k : \mathbb{R}^*_+ \to \mathbb{R}$, $||\theta_k||_\infty = \max_{t \in [0,\tau]} |\theta_k(t)| < \infty$
- Candidates for the estimation of $\lambda_0$ : $\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t, \boldsymbol{Z_i}) = \alpha_{\boldsymbol{\gamma}}(t)\mathrm{e}^{f_{\boldsymbol{\beta}}(\boldsymbol{Z_i})}$,

$$\text{where} \quad \log \alpha_{\boldsymbol{\gamma}} = \sum_{k=1}^{N} \gamma_k \theta_k \quad \text{and} \quad f_{\boldsymbol{\beta}} = \sum_{j=1}^{M} \beta_j f_j$$

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

$$C_n(\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\int_0^\tau \log \lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t,\boldsymbol{Z_i})\mathrm{d}N_i(t) - \int_0^\tau \lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t,\boldsymbol{Z_i})Y_i(t)\mathrm{d}t\right\}$$

- **Loss function** : the empirical Kullback divergence

$$\widetilde{K}_n(\lambda_0,\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau (\log \lambda_0(t,\boldsymbol{Z_i}) - \log \lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t,\boldsymbol{Z_i}))\lambda_0(t,\boldsymbol{Z_i})Y_i(t)\mathrm{d}t$$
$$- \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau (\lambda_0(t,\boldsymbol{Z_i}) - \lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t,\boldsymbol{Z_i}))Y_i(t)\mathrm{d}t$$

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

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

Lasso procedure : minimization of an $\ell_1$-penalized criterion

Estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ simultaneously via a weighted Lasso procedure :

$$(\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}) = \underset{(\boldsymbol{\beta},\boldsymbol{\gamma})\in\mathbb{R}^M\times\mathbb{R}^N}{\arg\min} \{C_n(\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}) + \mathrm{pen}(\boldsymbol{\beta}) + \mathrm{pen}(\boldsymbol{\gamma})\},$$

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

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

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

$$C_n(\lambda_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}}) + \mathrm{pen}(\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}) + \mathrm{pen}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}) \leq C_n(\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}) + \mathrm{pen}(\boldsymbol{\beta}) + \mathrm{pen}(\boldsymbol{\gamma}).$$

Using the Doob-Meyer decomposition $N_i = M_i + \Lambda_i$, we obtain

$$\begin{aligned}\widetilde{K}_n(\lambda_0, \lambda_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}}) \leq\ & \widetilde{K}_n(\lambda_0, \lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}) \\ & + \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \ \eta_{n,\tau}(f_j) + \mathrm{pen}(\boldsymbol{\beta}) - \mathrm{pen}(\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}) \\ & + \sum_{k=1}^{N} (\hat{\gamma}_L - \gamma)_k \ \nu_{n,\tau}(\theta_k) + \mathrm{pen}(\boldsymbol{\gamma}) - \mathrm{pen}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{L}})\end{aligned}$$

where $\eta_{n,\tau}(f_j) = \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau} f_j(\boldsymbol{Z_i}) \mathrm{d}M_i(s)$,

$$\nu_{n,\tau}(\theta_k) = \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau} \theta_k(s) \mathrm{d}M_i(s).$$

Standard Bernstein's inequality for martingales [van de Geer (1995)] :

$$\mathbb{P}\Big[\eta_{n,\tau}(f_j) \geq \Big(\sqrt{\frac{2\omega x}{n}} + \frac{x}{3n}\Big)||f_j||_{n,\infty}, V_{n,\tau}(f_j) \leq \omega\Big] \leq \mathrm{e}^{-x}$$

Standard Bernstein's inequality for martingales [van de Geer (1995)] :

$$\mathbb{P}\Big[\eta_{n,\tau}(f_j) \geq \Big(\sqrt{\frac{2\omega x}{n}} + \frac{x}{3n}\Big)||f_j||_{n,\infty}, V_{n,\tau}(f_j) \leq \omega\Big] \leq \mathrm{e}^{-x}$$

Problem in our case : martingale with a predictable variation $V_{n,\tau}(f_j)$ not observable :

$$V_{n,\tau}(f_j) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau (f_j(\boldsymbol{Z_i}))^2 \lambda_0(t, \boldsymbol{Z_i}) Y_i(s) ds$$

$\Rightarrow$ We will replace the predictable variation by the optional variation of $\eta_{n,\tau}(f_j)$ (see Hansen et al. (2012) and Gaïffas and Guilloux (2011)) :

$$\hat{V}_{n,\tau}(f_j) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau (f_j(\boldsymbol{Z_i}))^2 dN_i(s)$$

## Theorem : Empirical Bernstein's inequality

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

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

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

## Theorem : Empirical Bernstein's inequality

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

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

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

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

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

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

$$C_n(\lambda_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}}) + \mathrm{pen}(\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}) + \mathrm{pen}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}) \leq C_n(\lambda_{\boldsymbol{\beta}, \boldsymbol{\gamma}}) + \mathrm{pen}(\boldsymbol{\beta}) + \mathrm{pen}(\boldsymbol{\gamma}).$$

Using the Doob-Meyer decomposition $N_i = M_i + \Lambda_i$, we obtain

$$\begin{aligned} \widetilde{K}_n(\lambda_0, \lambda_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}}) \leq \widetilde{K}_n(\lambda_0, \lambda_{\boldsymbol{\beta}, \boldsymbol{\gamma}}) & \\ + \sum_{j=1}^{M} (\hat{\beta}_L - \beta)_j \; \eta_{n,\tau}(f_j) + \mathrm{pen}(\boldsymbol{\beta}) - \mathrm{pen}(\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}) & \\ + \sum_{k=1}^{N} (\hat{\gamma}_L - \gamma)_k \; \nu_{n,\tau}(\theta_k) + \mathrm{pen}(\boldsymbol{\gamma}) - \mathrm{pen}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}) & \end{aligned}$$

where $\eta_{n,\tau}(f_j) = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \int_0^\tau f_j(\boldsymbol{Z_i}) \mathrm{d}M_i(s)$,

$$\nu_{n,\tau}(\theta_k) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \theta_k(s) \mathrm{d}M_i(s).$$

From two empirical Bernstein's inequalities on $\eta_{n,\tau}(f_j)$ and $\nu_{n,\tau}(\theta_k)$, we deduce some data-weights : for $j = 1, ..., M$ and $k = 1, ..., N$

$$\omega_j \approx \sqrt{\frac{\log M}{n} \hat{V}_{n,\tau}(f_j)} \text{ and } \delta_k \approx \sqrt{\frac{\log N}{n} \hat{R}_{n,\tau}(\theta_k)},$$

where $\hat{R}_{n,\tau}(\theta_k) = \frac{1}{n}\sum_{i=1}^{n} \int_0^\tau (\theta_k(s))^2 \mathrm{d}N_i(s)$

**Theorem : Slow non-asymptotic oracle inequality for $\lambda_0$**

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

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}}) \leq \inf_{(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \mathbb{R}^M \times \mathbb{R}^N} \Big( \widetilde{K}_n(\lambda_0, \lambda_{\boldsymbol{\beta}}) + 2\,\mathrm{pen}(\boldsymbol{\beta}) + 2\,\mathrm{pen}(\boldsymbol{\gamma}) \Big),$$

with $\mathrm{pen}(\boldsymbol{\beta}) + \mathrm{pen}(\boldsymbol{\gamma}) \approx \ ||\boldsymbol{\beta}||_1 \sqrt{\log M/n} + ||\boldsymbol{\gamma}||_1 \sqrt{\log N/n}$.

**Theorem : Slow non-asymptotic oracle inequality for $\lambda_0$**

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

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{L}}}) \leq \inf_{(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \mathbb{R}^M \times \mathbb{R}^N} \Big( \widetilde{K}_n(\lambda_0, \lambda_\beta) + 2\,\mathrm{pen}(\boldsymbol{\beta}) + 2\,\mathrm{pen}(\boldsymbol{\gamma}) \Big),$$

with $\mathrm{pen}(\boldsymbol{\beta}) + \mathrm{pen}(\boldsymbol{\gamma}) \approx \ ||\boldsymbol{\beta}||_1 \sqrt{\log M/n} + ||\boldsymbol{\gamma}||_1 \sqrt{\log N/n}$.

- Non-asymptotic oracle inequality obtained without any assumptions
- First non-asymptotic oracle inequality on the intensity
- $f_{\boldsymbol{\beta}} : \mathbb{R}^p \to \mathbb{R}$ and $\alpha_{\boldsymbol{\gamma}} : \mathbb{R} \to \mathbb{R}$ are estimated at once
  - $\Rightarrow$ the resulting rate of convergence is the sum of the two expected rates in both situation separately : $\sim \sqrt{\log M/n} + \sqrt{\log N/n}$
  - Choice of $N$ of order $n$ to estimate $\alpha_{\boldsymbol{\gamma}}$ (see Bertin et al. (2011)) $\Rightarrow$ in a very high-dimensional setting, leading error term of order $\sqrt{\log M/n}$

$\hookrightarrow$ To obtain a fast non-asymptotic oracle inequality, need of further notations and assumptions

**Notations :**

- Design matrix : $\boldsymbol{X} = (f_j(\boldsymbol{Z_i}))_{\substack{1\leq i\leq n \\ 1\leq j\leq M}}$,

$$\tilde{\boldsymbol{X}}(t) = \left[ \begin{array}{c|ccc} & \theta_1(t) & \dots & \theta_N(t) \\ \boldsymbol{X} & \vdots & & \vdots \\ & \theta_1(t) & \dots & \theta_N(t) \end{array} \right] \in \mathbb{R}^{n\times(M+N)}$$

- Sparsity set and index :

$$J(\boldsymbol{b}) = \{j \in \{1, ..., M\} : b_j \neq 0\} \text{ and } |J(\boldsymbol{b})| = \mathrm{Card}\{J(\boldsymbol{b})\}$$

- Gram matrix : $\boldsymbol{\Psi_n} = \frac{1}{n}\boldsymbol{X^T X}$
- Our **weighted** Gram matrix :

$$\tilde{\boldsymbol{G}}_{\boldsymbol{n}} = \frac{1}{n}\int_0^{\tau} \tilde{\boldsymbol{X}}(t)^T \tilde{\boldsymbol{C}}(t)\tilde{\boldsymbol{X}}(t)dt,\ \ \tilde{\boldsymbol{C}} = (\mathrm{diag}(\lambda_0(t, \boldsymbol{Z_i})Y_i(t)))_{1\leq i\leq n}$$

$\hookrightarrow$ Our weighted Gram matrix is random

- Classical **Restricted Eigenvalue condition** $\mathbf{RE}(s, c_0)$ for the additive regression model (see Bickel, Ritov and Tsybakov (2009)) :

For some integer $s \in \{1, ..., M\}$ and a constant $c_0 > 0$,

$$0 < \boldsymbol{\kappa_0}(s, c_0) = \min_{\substack{J \subset \{1,...,M\}, \\ |J| \leq s}} \min_{\substack{\boldsymbol{b} \in \mathbb{R}^M \setminus \{0\}, \\ ||\boldsymbol{b}_{J^c}||_1 \leq c_0 ||\boldsymbol{b}_J||_1}} \frac{(\boldsymbol{b^T \Psi_n b})^{1/2}}{||\boldsymbol{b}_J||_2}.$$

- Classical **Restricted Eigenvalue condition** $\mathbf{RE}(s, c_0)$ for the additive regression model (see Bickel, Ritov and Tsybakov (2009)) :

  For some integer $s \in \{1, ..., M\}$ and a constant $c_0 > 0$,

$$0 < \boldsymbol{\kappa_0}(s, c_0) = \min_{\substack{J \subset \{1,...,M\}, \\ |J| \leq s}} \min_{\substack{\boldsymbol{b} \in \mathbb{R}^M \setminus \{0\}, \\ ||\boldsymbol{b}_{J^c}||_1 \leq c_0 ||\boldsymbol{b}_J||_1}} \frac{(\boldsymbol{b^T \Psi_n b})^{1/2}}{||\boldsymbol{b}_J||_2}.$$

- **Restricted Eigenvalue condition** $\widetilde{\mathbf{RE}}(s, r_0)$ **on** $\mathbb{E}(\tilde{\boldsymbol{G}}_{\boldsymbol{n}})$ **:**

  For some integer $s \in \{1, ..., M + N\}$ and a constant $r_0 > 0$,

$$0 < \tilde{\boldsymbol{\kappa}}_{\boldsymbol{0}}(s, r_0) = \min_{\substack{J \subset \{1,...,M+N\}, \\ |J| \leq s}} \min_{\substack{\boldsymbol{b} \in \mathbb{R}^{M+N} \setminus \{0\}, \\ ||\boldsymbol{b}_{J^c}||_1 \leq r_0 ||\boldsymbol{b}_J||_1}} \frac{(\boldsymbol{b^T} \mathbb{E}(\tilde{\boldsymbol{G}}_{\boldsymbol{n}}) \boldsymbol{b})^{1/2}}{||\boldsymbol{b}_J||_2}.$$

$\hookrightarrow$ Link with a Restricted Eigenvalue condition on $\tilde{\boldsymbol{G}}_{\boldsymbol{n}}$

## Lemma : Link between the $\widetilde{\mathbf{RE}}$ condition on $\mathbb{E}(\tilde{G}_n)$ and on $\tilde{G}_n$

Under Assumption $\widetilde{\mathbf{RE}}(s, r_0)$, with probability larger than $1 - \tilde{\pi}_n$, with $\tilde{\pi}_n = \exp\left(-\frac{n\tilde{\boldsymbol{\kappa}}^4}{2L^2(1+r_0)^2 s(L^2(1+r_0)^2 s + \tilde{\boldsymbol{\kappa}}^2)}\right)$, we have

$$0 < \tilde{\boldsymbol{\kappa}} = \min_{\substack{J \subset \{1,\dots,M\}, \\ |J| \leq s}} \min_{\substack{\boldsymbol{b} \in \mathbb{R}^M \setminus \{0\}, \\ ||\boldsymbol{b}_{J^c}||_1 \leq r_0 ||\boldsymbol{b}_J||_1}} \frac{(\boldsymbol{b}^T \tilde{G}_n \boldsymbol{b})^{1/2}}{||\boldsymbol{b}_J||_2} \text{ and } \tilde{\boldsymbol{\kappa}} = (1/\sqrt{2})\tilde{\boldsymbol{\kappa}}_{\mathbf{0}}(s, r_0).$$

**Remark :** $\begin{pmatrix} \hat{\boldsymbol{\beta}}_L - \boldsymbol{\beta} \\ \hat{\gamma}_L - \gamma \end{pmatrix}^T \tilde{G}_n \begin{pmatrix} \hat{\boldsymbol{\beta}}_L - \boldsymbol{\beta} \\ \hat{\gamma}_L - \gamma \end{pmatrix} = || \log \lambda_{\hat{\boldsymbol{\beta}}_L, \hat{\gamma}_L} - \log \lambda_{\boldsymbol{\beta}, \gamma} ||^2_{n,\Lambda}$

$\hookrightarrow$ We need a connection between $||.||_{n,\Lambda}$ and $\widetilde{K}_n$

**Assumption 3.** *There exists* $\rho > 0$*, such that*

$$\widetilde{\Gamma}(\rho) = \{(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathbb{R}^M \times \mathbb{R}^N : || \log \lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}} - \log \lambda_0 ||_{n,\infty} \leq \rho\}$$

*contains a non-empty open set of* $\mathbb{R}^M \times \mathbb{R}^N$*.*

**Proposition : Connection between $||.||_{n,\Lambda}$ and $\widetilde{K}_n$**

Under Assumption 3, for all $(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \widetilde{\Gamma}(\rho)$,

$$\rho' || \log \lambda_\beta - \log \lambda_0 ||_{n,\Lambda}^2 \leq \widetilde{K}_n(\lambda_0, \lambda_\beta) \leq \rho'' || \log \lambda_\beta - \log \lambda_0 ||_{n,\Lambda}^2.$$

Assumptions :

1) $\widetilde{\mathbf{RE}}(s, r_0)$ with $r_0 = \left(3 + \frac{8}{\zeta} \max\left(\sqrt{|J(\boldsymbol{\beta})|}, \sqrt{|J(\boldsymbol{\gamma})|}\right)\right)$

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

3) Assumption 3

**Theorem : Fast non-asymptotic oracle inequalities on the intensity**

Under Assumptions 1,2 and 3, we have with probability larger than $1 - c_3 \mathrm{e}^{-x} - c_3' \mathrm{e}^{-y} - \tilde{\pi}_n$,

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L, \hat{\gamma}_L}) \leq (1+\zeta) \inf_{\substack{(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \widetilde{\Gamma}(\rho) \\ \max(|J(\boldsymbol{\beta})|,|J(\boldsymbol{\gamma})|) \leq s}} \left\{ \widetilde{K}_n(\lambda_0, \lambda_{\beta,\gamma}) + \widetilde{C}_{(\zeta,\rho)} \frac{\max(|J(\boldsymbol{\beta})|, |J(\boldsymbol{\gamma})|)}{\tilde{\boldsymbol{\kappa}}^2} \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \{\omega_j^2, \delta_k^2\} \right\}$$

$$|| \log \lambda_0 - \log \lambda_{\hat{\beta}_L, \hat{\gamma}_L} ||_{n,\Lambda}^2 \leq (1+\zeta) \inf_{\substack{(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \widetilde{\Gamma}(\rho) \\ \max(|J(\boldsymbol{\beta})|,|J(\boldsymbol{\gamma})|) \leq s}} \left\{ \|\log \lambda_0 - \log \lambda_{\beta,\gamma}\|_{n,\Lambda}^2 + \widetilde{C}'_{(\zeta,\rho)} \frac{\max(|J(\boldsymbol{\beta})|,|J(\boldsymbol{\gamma})|)}{\tilde{\boldsymbol{\kappa}}^2} \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \{\omega_j^2, \delta_k^2\} \right\}$$

# Fast non-asymptotic oracle inequalities on the intensity

**Theorem : Fast non-asymptotic oracle inequalities on the intensity**

Under Assumptions 1,2 and 3, we have with probability larger than $1 - c_3 \mathrm{e}^{-x} - c_3' \mathrm{e}^{-y} - \tilde{\pi}_n$,

$$\widetilde{K}_n(\lambda_0, \lambda_{\hat{\beta}_L, \hat{\gamma}_L}) \leq (1+\zeta) \inf_{\substack{(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \widetilde{\Gamma}(\rho) \\ \max(|J(\boldsymbol{\beta})|,|J(\boldsymbol{\gamma})|) \leq s}} \left\{ \widetilde{K}_n(\lambda_0, \lambda_{\beta,\gamma}) + \widetilde{C}_{(\zeta,\rho)} \frac{\max(|J(\boldsymbol{\beta})|, |J(\boldsymbol{\gamma})|)}{\tilde{\boldsymbol{\kappa}}^2} \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \{\omega_j^2, \delta_k^2\} \right\}$$

$$|| \log \lambda_0 - \log \lambda_{\hat{\beta}_L, \hat{\gamma}_L} ||_{n,\Lambda}^2 \leq (1+\zeta) \inf_{\substack{(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \widetilde{\Gamma}(\rho) \\ \max(|J(\boldsymbol{\beta})|,|J(\boldsymbol{\gamma})|) \leq s}} \left\{ \|\log \lambda_0 - \log \lambda_{\beta,\gamma}\|_{n,\Lambda}^2 + \widetilde{C}'_{(\zeta,\rho)} \frac{\max(|J(\boldsymbol{\beta})|,|J(\boldsymbol{\gamma})|)}{\tilde{\boldsymbol{\kappa}}^2} \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \{\omega_j^2, \delta_k^2\} \right\}$$

with
$$\left( \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \{\omega_j, \delta_k\} \right)^2 \approx \max \left\{ \frac{\log M}{n}, \frac{\log N}{n} \right\}.$$

**Model** : $\lambda_0(t, \boldsymbol{Z_i}) = \alpha_0(t)\mathrm{e}^{f_0(\boldsymbol{Z_i})}$

- Kong and Nan (2012) : results on $f_0$, lower rate of convergence, confidence that depends on $n$ and $M$
- Bradic and Song (2012) : results on $f_0$, $f_0$ taken in the dictionary
- Huang et al. (2013) : results on $f_0(\boldsymbol{Z_i}(t)) = \boldsymbol{\beta_0^T Z_i(t)}$

All results are based on the Cox partial log-likelihood :

$$C_n(\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}) = -\frac{1}{n}\sum_{i=1}^{n}\Big\{\int_0^{\tau}\log\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t,\boldsymbol{Z_i})\mathrm{d}N_i(t) - \int_0^{\tau}\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t,\boldsymbol{Z_i})Y_i(t)\mathrm{d}t\Big\}$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\Big\{\int_0^{\tau}\log\big(\alpha_{\boldsymbol{\gamma}}(t)S_n(f_{\boldsymbol{\beta}},t)\big)\mathrm{d}N_i(t)\Big\} - \int_0^{\tau}\alpha_{\boldsymbol{\gamma}}(t)S_n(f_{\boldsymbol{\beta}},t)\mathrm{d}t$$

$$\underbrace{-\frac{1}{n}\sum_{i=1}^{n}\Big\{\int_0^{\tau}\log\frac{\mathrm{e}^{f_{\boldsymbol{\beta}}(\boldsymbol{Z_i})}}{S_n(f_{\boldsymbol{\beta}},t)}\mathrm{d}N_i(t)\Big\}}_{l_n^*(f_{\boldsymbol{\beta}})\ \text{Cox partial log-likelihood}} \quad \text{with } S_n(\beta,t) = \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\mathrm{e}^{f_{\boldsymbol{\beta}}(\boldsymbol{Z_i})}$$

**Conclusion :**

- We obtain a fast non-asymptotic oracle inequality for a general intensity

  ↪ allows to predict the survival time throughout the conditional intensity in a high dimensional setting

**Perspectives :**

- to compare the predictive accuracy of the Lasso in the Cox model with and without our new data-driven weights
- to obtain a fast non-asymptotic oracle inequality on the intensity in the Cox model with the usual two-step procedure (estimation of $f_0$ using the Cox partial log-likelihood and then estimation of $\alpha_0$)

# References I


- Bach F. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, (2010).
- Bertin, K., Le Pennec, E. and Rivoirard, V. Adaptive Dantzig density estimation. *Annales de l'IHP, Probabilités et Statistiques* (2011).
- Bickel, P.J., Ritov, Y. and Tsybakov, A.B. Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics* (2009).
- Bradic, J. and Song, R. Gaussian Oracle Inequalities for Structured Selection in Non-Parametric Cox Model.arXiv preprint arXiv :1207.4510 (2012).
- Gaïffas S. and Guilloux A. High-dimensional additive hazard models and the Lasso. *Electronic Journal of Statistics* (2011).

# References II


- Hansen, N. R., Reynaud-Bouret, P. and Rivoirard, V. Lasso and probabilistic inequalities for multivariate point processes. arXiv preprint arXiv1208.0570 (2012).
- Huang, J., Sun, T., Ying, Z., Yu, Y. and Zhang, C.H. Oracle inequalities for the lasso in the Cox model. *The Annals of Statistics* (2013).
- Kong S. and Nan B. Non-asymptotic oracle inequalities for the high-dimensional Cox regression via Lasso. arXiv :1204.1992 (2012).
- Lemler S. Oracle inequalities for the Lasso for the conditional hazard rate in a high-dimensional setting. arXiv :1206.5628v3 (2013).
- van de Geer, S. Exponential Inequalities for Martingales, with Application to Maximum Likelihood Estimation for Counting Processes. *The Annals of Statistics* (1995).