High-dimensional estimation of counting process intensities

Sarah Lemler

Thesis supervisors : Agathe Guilloux
Marie-Luce Taupin

Laboratoire "Statistique et Génome", LaMME

Onzième Colloque Jeunes Probabilistes et Statisticiens

Forges-Les-Eaux, du 6 au 11 avril 2014

Outline

1. Framework and model
   - Context
   - Counting processes and multiplicative Aalen intensity model
   - Model
2. Oracle Inequalities for the Lasso in the high-dimensional Aalen multiplicative intensity model
   - Estimation procedure
   - Slow non-asymptotic oracle inequality on the intensity
   - Comparison with the existing results for the Cox model
3. Estimation of the intensity in the Cox model using a two-step procedure
   - Model and framework
   - Estimation of the regression parameter
   - Estimation of the baseline function
   - Non-asymptotic oracle inequality for the baseline function
   - Simulations

Outline

Context :

- Example :
  - $n = 144$ lymph node positive breast cancer patients
  - Variable of interest : Metastasis-free follow-up time, that can be right-censored
  - Covariates : 5 clinical variables, $\mathbf{70}$ levels of gene expression

Goal : to predict the survival from metastasis for the breast cancer 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 : $X_i = \min(T_i, C_i)$, δ_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 λ_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

- The Cox model

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

f_0 the regression function and α_0 the baseline hazard function

- To predict the survival :

$$\mathbb{E}(T_i) = \int_0^{+\infty} \mathrm{e}^{-\int_0^t \alpha_0(s) \mathrm{e}^{f_0(\boldsymbol{Z_i})} \mathrm{d}s} \mathrm{d}t$$

- **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

- The Cox model

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

f_0 the regression function and α_0 the baseline hazard function

- To predict the survival :

$$\mathbb{E}(T_i) = \int_0^{+\infty} \mathrm{e}^{-\int_0^t \alpha_0(s) \mathrm{e}^{f_0(\boldsymbol{Z_i})} \mathrm{d}s} \mathrm{d}t$$

$\Rightarrow$ Estimation of both parameters

Outline

1. Framework and model
2. Oracle Inequalities for the Lasso in the high-dimensional Aalen multiplicative intensity model
   - Estimation procedure
   - Slow non-asymptotic oracle inequality on the intensity
   - Comparison with the existing results for the Cox model
3. Estimation of the intensity in the Cox model using a two-step procedure

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 λ_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 λ_0 : $\lambda_{\boldsymbol{\beta},\boldsymbol{\gamma}}(t, \boldsymbol{Z_i}) = \alpha_{\boldsymbol{\gamma}}(t)\mathrm{e}^{f_{\boldsymbol{\beta}}(\boldsymbol{Z_i})}$,

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

- Lasso procedure : minimization of an ℓ_1-penalized criterion
  Estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ simultaneously via a weighted Lasso procedure :

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

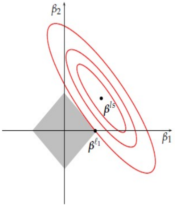
$$\text{with} \quad \mathrm{pen}(\boldsymbol{\beta}) = \sum_{j=1}^{M} \omega_j |\beta_j| \text{ and } \mathrm{pen}(\boldsymbol{\gamma}) = \sum_{k=1}^{N} \delta_k |\gamma_k|,$$

  ω_j and δ_k positive data-driven weights to be defined

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

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

- Advantages of the Lasso procedure :
  - convex minimization problem
    $\Rightarrow$ computable in practice
  - sparsity of the Lasso estimator
    $\Rightarrow$ results easily interpretable



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

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

Theorem : Slow non-asymptotic oracle inequality for λ_0

With large probability, we have

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

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

- First non-asymptotic oracle inequality on the intensity
- With further assumptions, fast non-asymptotic oracle inequality of order $\max\{\log M/n, \log N/n\}$
- Choice of N of order n to estimate $\alpha_{\boldsymbol{\gamma}}$ (see Bertin et al. (2011))

$\Rightarrow$ in a high-dimensional setting, leading error term of order $\sqrt{\log M/n}$ or $\log M/n$

Comparison with the existing results for the Cox model

Preprints on non-asymptotic oracle inequalities for the Lasso in the Cox model

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 :

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

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

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

Outline

The Cox model :

$$\lambda_0(t, \boldsymbol{Z_i}) = \alpha_0(t)\mathrm{e}^{\boldsymbol{\beta_0^T Z_i}}$$

Goal : Estimation of α_0 :

- non-parametrically
- adaptively
- using model selection

Usually :

- Estimation of $\boldsymbol{\beta_0}$ using the Cox partial log-likelihood
- Breslow estimator $\hat{\Lambda}_{Br}^{\hat{\boldsymbol{\beta}}}$ of the cumulative baseline function
- Kernel estimator $\hat{\alpha}_{Kern}^{\hat{\boldsymbol{\beta}}}$ (cf. Andersen et al., 1993) :

$$\hat{\alpha}_{Kern}^{\hat{\boldsymbol{\beta}}}(t) = \sum_{i=1}^{n} \int_0^{\tau} \frac{1}{b} K\Big(\frac{t-u}{b}\Big) \frac{1}{\sum_{j=1}^{n} \mathrm{e}^{\hat{\boldsymbol{\beta}}^{\boldsymbol{T}} \boldsymbol{Z_j}} Y_j(u)} \mathrm{d}N_i(u),$$

for some bandwidths $b > 0$.

$\hookrightarrow$ any theoretical adaptation result on the estimation of α_0

We propose :

- the estimation of $\boldsymbol{\beta_0}$ with the Cox partial log-likelihood and the use of existing results in high-dimension
- the construction of an adaptive estimator of α_0 depending on the estimation of $\boldsymbol{\beta_0}$
- an estimation in small-dimension and in high-dimension
- theoretical results on the estimation of α_0
- comparison between the two procedures (the Kernel procedure and the model selection procedure) through simulations

Estimation of the regression parameter $\boldsymbol{\beta_0}$

For the estimation of $\boldsymbol{\beta_0}$:

- Cox partial log-likelihood $l_n^*(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \log \frac{\mathrm{e}^{\boldsymbol{\beta^T Z_i}}}{S_n(\boldsymbol{\beta}, t)} \mathrm{d}N_i(t)$
- In high-dimension, from Huang et al. (2013),

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{L}} = \arg\min_{\boldsymbol{\beta}} \left\{ l_n^*(\boldsymbol{\beta}) + \sqrt{\frac{\log p}{n}} ||\boldsymbol{\beta}||_1 \right\}.$$

Under some classical assumptions, they obtain

$$||\hat{\boldsymbol{\beta}}_{\boldsymbol{L}} - \boldsymbol{\beta_0}||_1 \leq C(s)\sqrt{\frac{\log p}{n}}$$

Estimation of the regression parameter $\boldsymbol{\beta_0}$

Assumption 4. *There exists a positive constant* B *such that for all* $i \in \{1, ..., n\}$ *and* $j \in \{1, ..., p\}$, $|Z_{i,j}| \leq B$.

Proposition

Under Assumption 4,

$$||\hat{\boldsymbol{\beta}}_{\boldsymbol{L}}^{\boldsymbol{T}} \boldsymbol{Z} - \boldsymbol{\beta}_{\boldsymbol{0}}^{\boldsymbol{T}} \boldsymbol{Z}||_1 \leq C'(s) \sqrt{\frac{\log p}{n}} \tag{1}$$

Remark : The estimation of α_0 hardly depends on the estimation of $\boldsymbol{\beta_0}$

$\hookrightarrow$ We assume that we have a $\hat{\boldsymbol{\beta}}$ that verifies Equation (1).

- Criterion of estimation :

$$C_n(\alpha, \boldsymbol{\beta}) = -\frac{2}{n}\sum_{i=1}^{n}\int_0^{\tau}\alpha(t)\mathrm{d}N_i(t) + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\alpha^2(t)\mathrm{e}^{\boldsymbol{\beta^T Z_i}}Y_i(t)\mathrm{d}t$$

- Norms associated to the criterion of estimation :
  - Random norm : $||\alpha||^2_{rand} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\alpha^2(t)\mathrm{e}^{\boldsymbol{\beta_0^T Z_i}}Y_i(t)\mathrm{d}t$
  - Deterministic norm : $||\alpha||^2_{det} = \int_0^{\tau}\alpha^2(t)\mathbb{E}[\mathrm{e}^{\boldsymbol{\beta_0^T Z_1}}Y_1(t)]\mathrm{d}t$

If we minimize $C_n(\alpha, \boldsymbol{\beta_0})$ over all α, we find ∞.

$\Rightarrow$ restriction to approximated parametric models

Models collection :

- $\{S_m, m \in \mathcal{M}_n\}$ a collection of models such that

$$S_m = \{\alpha : \alpha(t) = \sum_{j \in J_m} a_j^m \varphi_j^m(t), a_j^m \in \mathbb{R}\},$$

  where $(\varphi_j^m)_{j \in J_m}$ is an orthonormal basis of $(L^2 \cap L^\infty)([0, \tau])$
- $D_m = |J_m|$, the cardinality of S_m

Let define :

- Candidates for the estimation of α_0 : for $m \in \mathcal{M}_n$,

$$\hat{\alpha}_m^{\hat{\boldsymbol{\beta}}} = \arg\min_{\alpha \in S_m} \{C_n(\alpha, \hat{\boldsymbol{\beta}})\}$$

- Projection of α_0 on S_m : α_m
- Selection of the relevant space :

$$\hat{m}^{\hat{\boldsymbol{\beta}}} = \arg\min_{m \in \mathcal{M}_n} \{C_n(\hat{\alpha}_m^{\hat{\boldsymbol{\beta}}}, \hat{\boldsymbol{\beta}}) + \mathrm{pen}(m)\}$$

- Final estimator : $\tilde{\alpha}(t) = \hat{\alpha}_{\hat{m}^{\hat{\boldsymbol{\beta}}}}^{\hat{\boldsymbol{\beta}}}(t)$

Non-asymptotic oracle inequalities for the baseline function

Theorem (Non-asymptotic oracle inequality for the baseline function)

Let assume that $\max_{m\in\mathcal{M}_n} D_m \leq \sqrt{n}/\log n$ *and let define the penalty as follows :*

$$\mathrm{pen}(m) := K_0(1+||\alpha_0||_\infty)\frac{D_m}{n},$$

where K_0 *is a numerical constant. Under classical assumptions, we have*

$$\mathbb{E}[||\tilde{\alpha}-\alpha_0||_{det}^2] \leq \kappa_0 \inf_{m\in\mathcal{M}_n}\{||\alpha_0-\alpha_m||_{det}^2 + 2\,\mathrm{pen}(m)\} + \frac{C}{n} + C'(s)\frac{\log p}{n},$$

where $\kappa_0 = C(\tau, \phi, ||\alpha_0||_\infty, f_0, \mathrm{e}^{B||\boldsymbol{\beta_0}||_1})$.

Non-asymptotic oracle inequalities for the baseline function

Theorem (Non-asymptotic oracle inequality for the baseline function)

Let assume that $\max_{m \in \mathcal{M}_n} D_m \leq \sqrt{n}/\log n$ *and let define the penalty as follows :*

$$\text{pen}(m) := K_0(1 + ||\alpha_0||_\infty)\frac{D_m}{n},$$

where K_0 *is a numerical constant. Under classical assumptions, we have*

$$\mathbb{E}[||\tilde{\alpha} - \alpha_0||_{det}^2] \leq \kappa_0 \inf_{m \in \mathcal{M}_n} \{||\alpha_0 - \alpha_m||_{det}^2 + 2\,\text{pen}(m)\} + \frac{C}{n} + C'(s)\frac{\log p}{n},$$

where $\kappa_0 = C(\tau, \phi, ||\alpha_0||_\infty, f_0, \mathrm{e}^{B||\boldsymbol{\beta_0}||_1})$.

Non-asymptotic oracle inequalities for the baseline function

Theorem (Non-asymptotic oracle inequality for the baseline function)

Let assume that $\max_{m \in \mathcal{M}_n} D_m \leq \sqrt{n}/\log n$ *and let define the penalty as follows :*

$$\text{pen}(m) := K_0(1 + ||\alpha_0||_\infty)\frac{D_m}{n},$$

where K_0 *is a numerical constant. Under classical assumptions, we have*

$$\mathbb{E}[||\tilde{\alpha} - \alpha_0||_{det}^2] \leq \kappa_0 \inf_{m \in \mathcal{M}_n} \{||\alpha_0 - \alpha_m||_{det}^2 + 2\,\text{pen}(m)\} + \frac{C}{n} + C'(s)\frac{\log p}{n},$$

where $\kappa_0 = C(\tau, \phi, ||\alpha_0||_\infty, f_0, \mathrm{e}^{B||\boldsymbol{\beta_0}||_1})$.

Simulations :

- n=1000
- Design matrix : $Z \sim \mathcal{U}([-1, 1])$, $Z \in \mathbb{R}^{n \times p}$
- Rate of censoring $\approx 20\%$
- $\alpha_0(t) = a\lambda t^{a-1}$ (Weibull distribution)

1. "small-dimension" : $p = 7$, $\boldsymbol{\beta_0} = (0.1, 0.3, 0.5, 0, 0, 0, 0)$
2. "high-dimension" : $p = 30$, $\boldsymbol{\beta_0} = (0.1, 0.3, 0.5, 0, ..., 0) \in \mathbb{R}^{30}$

1. **Estimation of β_0 :**
   - function coxph on R (whithout penalization)
   - function glmnet on R (ℓ_1-penalization)
   - function glmnet+coxph on R (first variable selection, then estimation)
2. **Estimation of α_0 :**
   - Kernel estimator :
     - Epanechnikov kernel

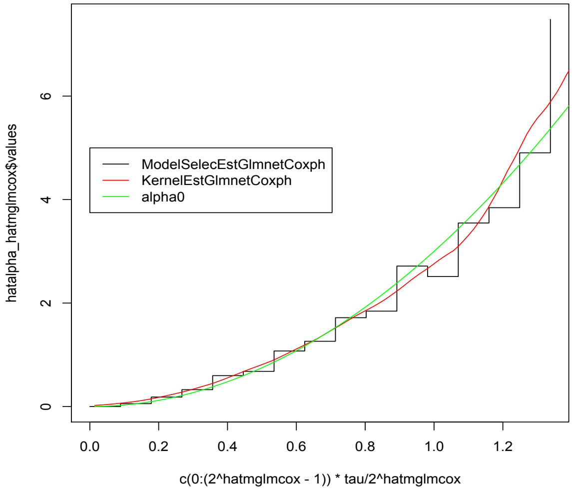
$$K(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{\{|u| \leq 1\}}$$

     - Choice of the bandwidth by cross-validation
   - Model selection
     - Histogram basis : $\varphi_j(t) = \frac{1}{\sqrt{\tau}} 2^{m/2} \mathbb{1}_{[(j-1)\tau/2^m, j\tau/2^m[}(t)$, for $j = 1, ..., 2^m$, $D_m = 2^m$.

1. Estimations of $\boldsymbol{\beta_0} = (0.1, 0.3, 0.5, 0, 0, 0, 0)$:

   - function coxph on R (whithout penalization) :
     $\hat{\boldsymbol{\beta}}_{\boldsymbol{coxph}} = (0.14, 0.32, 0.47, -0.03, -0.01, 0.04, -0.001)$
   - function glmnet for the Cox model on R (ℓ_1-penalization) :
     $\hat{\boldsymbol{\beta}}_{\boldsymbol{glmnet}} = (0.09, 0.27, 0.42, 0, 0, 0, 0)$
   - function glmnet+coxph on R (first variable selection, then estimation)
     $\hat{\boldsymbol{\beta}}_{\boldsymbol{glmnet+coxph}} = (0.14, 0.32, 0.47)$

2. Estimation of α_0 :



Estimation of the baseline function α_0 I

Simulations : p of order $\sqrt{n}$

1. Estimations of $\boldsymbol{\beta_0} = (0.1, 0.3, 0.5, 0, ..., 0) \in \mathbb{R}^{30}$:

    - glmnet for variable selection : selection of 11 variables

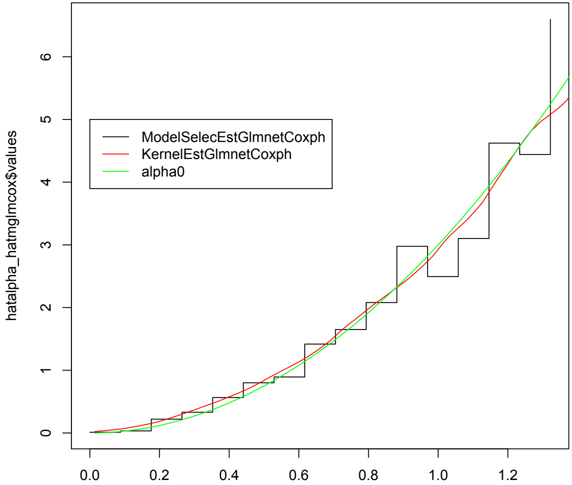
    - coxph on these 11 variables :

      $$\hat{\boldsymbol{\beta}} = (0.15, 0.27, 0.45, 0.09, -0.06, 0.08, -0.1, -0.11, 0.07, 0.17, 0.09)$$

Estimation of the baseline function α_0 II

Simulations : p of order $\sqrt{n}$

2. Estimation of α_0 :



Perspectives

- Improvement of the simulations (increase in p, variation of the censoring...)

- Comparison between the two methods to estimate the whole intensity :
  - the one-step procedure : ℓ_1-procedure on α_0 and on $\boldsymbol{\beta_0}$
  - the two-step procedure : ℓ_1-procedure to estimate $\boldsymbol{\beta_0}$ and then ℓ_0-procedure to estimate α_0

References I

- Andersen, P.K., Borgan, ∅., Gill, R.D. and Keiding, N. Statistical models based on counting processes. *Springer Series in Statistics*. Springer (1993).
- 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).

References II

- Brunel, E., Comte, F. and Lacour, C. Minimax estimation of the conditional cumulative distribution function. *Sankhya A* (2010).
- Comte, F., Gaïffas, S. and Guilloux, A. Adaptive estimation of the conditional intensity of marker-dependent counting processess. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* (2011).
- Gaïffas S. and Guilloux A. High-dimensional additive hazard models and the Lasso. *Electronic Journal of Statistics* (2011).
- 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).

References III

- 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).
- Reynaud-Bouret, P. Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli* (2006).
- van de Geer, S. Exponential Inequalities for Martingales, with Application to Maximum Likelihood Estimation for Counting Processes. *The Annals of Statistics* (1995).