Contributions théoriques et appliquées à l'analyse de survie

Olivier Bouaziz

MAP5 (CNRS 8145), Université Paris Descartes, Sorbonne Paris Cité

Soutenance d'Habilitation à Diriger des Recherches

Outline

The Steno memorial hospital dataset

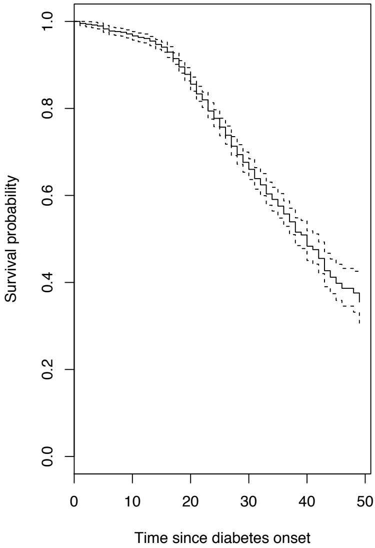
- Cohort dataset of 2 709 Danish diabetic patients collected between 1933 and 1981 from *Andersen et al., 1993.*
- The variable of interest is the time from diabetes onset until death (in years).
- 74% of right censoring due to emigration or end of study (December, 31st 1984).
- Left truncation due to delayed entry into the study.
- Gender and calendar year of diabetes onset (range : 1933 – 1972) were also collected for each patient.

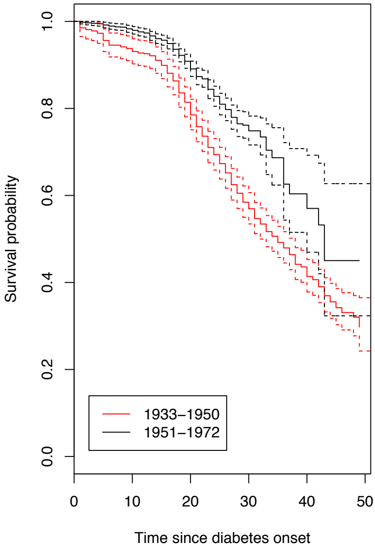
The Steno memorial hospital dataset

- Cohort dataset of 2 709 Danish diabetic patients collected between 1933 and 1981 from *Andersen et al., 1993.*
- The variable of interest is the time from diabetes onset until death (in years).
- 74% of right censoring due to emigration or end of study (December, 31st 1984).
- Left truncation due to delayed entry into the study.
- Gender and calendar year of diabetes onset (range : 1933 – 1972) were also collected for each patient.

- Classical survival analysis except that we want to take into account a possible cohort effect due to the wide range of year of diabetes onset.

Illustration of the cohort effect





A breakpoint model

- Suppose there are K segments and let $R_1, \ldots, R_n$ be the segment indexes of each individual. For example, $n = 10$ and $R_{1:10} = 1112222333$ means 2 breakpoints occur in positions 3 and 7.
- The model is :

$$\lambda(t|Z_i, R_i = k) = \lim_{\triangle t \to 0} \frac{\mathbb{P}(t \leq T_i < t + \triangle t | T_i \geq t, Z_i, R_i = k)}{\triangle t}$$
$$= \lambda_k(t) \exp(\beta_k Z_i)$$

The goal is :

- Estimate the *a posteriori* probability of a breakpoint, $\mathbb{P}(R_i = k, R_{i+1} = k + 1|\text{data})$.
- Estimate the λ_ks and β_ks.

The EM algorithm

Introduce data $= (T^{\text{obs}}_{1:n}, \Delta_{1:n}, Z_{1:n})$ and $\boldsymbol{\theta} = (\lambda_1, \ldots, \lambda_K, \beta_1, \ldots, \beta_K)$.

- (E-step) Compute :

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}}) = \int_{R_{1:n}} \mathbb{P}(R_{1:n}|\text{data}; \boldsymbol{\theta}_{\text{old}}) \log \mathbb{P}(\text{data}|R_{1:n}; \boldsymbol{\theta}) dR_{1:n}$$
$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{P}(R_i = k|\text{data}; \boldsymbol{\theta}_{\text{old}}) \log \mathbb{P}(\text{data}_i|R_i = k; \boldsymbol{\theta}),$$

  where $\boldsymbol{\theta}_{\text{old}}$ represents the previous update of the parameter.

- (M-step) Maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}})$ with respect to $\boldsymbol{\theta}$.

Computation of the emission probability and the posterior segment distribution

- The contribution of the ith individual to the likelihood conditionally to its segment index is :

$$\log \mathbb{P}(T_i^{\text{obs}}, \Delta_i, Z_i | R_i = k; \boldsymbol{\theta})$$
$$= \Delta_i \left\{ \log \left(\lambda_k(T_i^{\text{obs}}) \right) + \beta_k Z_i \right\} - \int_0^{T_i^{\text{obs}}} \lambda_k(t) \exp(\beta_k Z_i) dt,$$

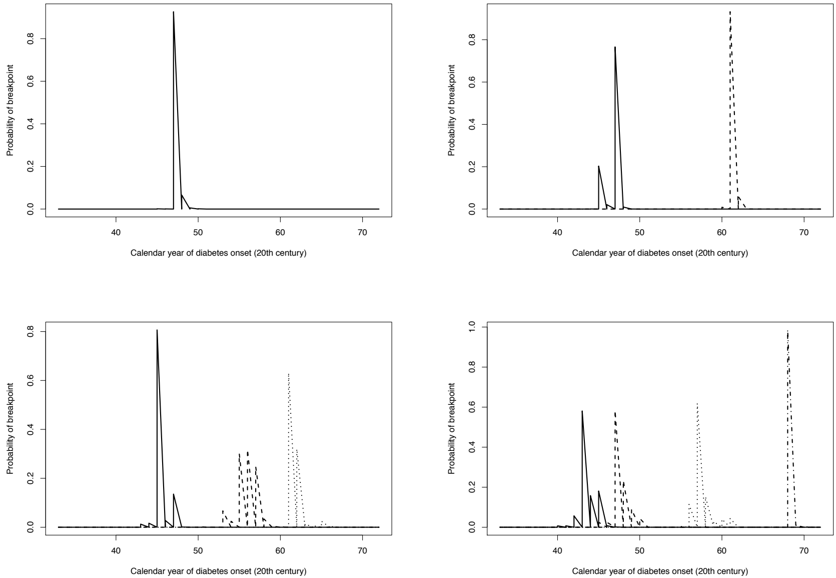
with a parametric or non-parametric baseline hazard

- Taking a prior distribution $\eta_i(k) = \mathbb{P}(R_i = k + 1 | R_{i-1} = k)$, the posterior segment distributions is computed from constrained Hidden Markov Model theory with the forward and backward quantities.

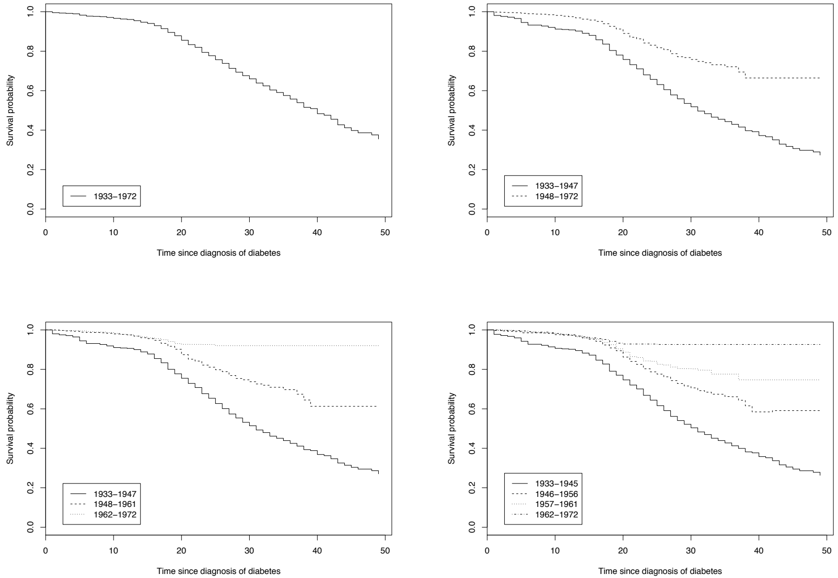
The Steno memorial hospital dataset (exp. baseline, one covariate : gender)

	No bp	One bp 1948	Two bp 1948, 62	Three bp 1946, 57, 62	Four bp 1944, 48, 58, 69
$\hat{\lambda}_1$	0.012	0.022	0.023	0.023	0.024
$\hat{\lambda}_2$		0.006	0.008	0.011	0.015
$\hat{\lambda}_3$			0.003	0.006	0.009
$\hat{\lambda}_4$				0.003	0.004
$\hat{\lambda}_5$					0.001
$e^{\hat{\beta}_1}$	1.32	1.29	1.29	1.29	1.25
$e^{\hat{\beta}_2}$		1.61	1.60	1.41	1.43
$e^{\hat{\beta}_3}$			1.44	1.80	1.50
$e^{\hat{\beta}_4}$				1.46	1.66
$e^{\hat{\beta}_5}$					0.90
BIC	7426.405	7214.413	7179.012	7187.442	7194.631

Marginal distributions of the breakpoints



Weighted Kaplan-Meier estimators



Confidence intervals

A bootstrap procedure is implemented to obtain 95% confidence intervals.
In the two breakpoints model (with covariate gender) :

- 1933 – 1947

$$\hat{\lambda} = 0.023[0.020; 0.027] \quad \exp(\hat{\beta}) = 1.29[1.06; 1.55]$$

- 1948 – 1961

$$\hat{\lambda} = 0.008[0.007; 0.012] \quad \exp(\hat{\beta}) = 1.60[1.12; 2.09]$$

- 1962 – 1972

$$\hat{\lambda} = 0.003[0.001; 0.005] \quad \exp(\hat{\beta}) = 1.44[0.90; 3.40]$$

Confidence intervals

A bootstrap procedure is implemented to obtain 95% confidence intervals.
In the two breakpoints model (with covariate gender) :

- 1933 – 1947

$$\hat{\lambda} = 0.023[0.020; 0.027] \quad \exp(\hat{\beta}) = 1.29[1.06; 1.55]$$

- 1948 – 1961

$$\hat{\lambda} = 0.008[0.007; 0.012] \quad \exp(\hat{\beta}) = 1.60[1.12; 2.09]$$

- 1962 – 1972

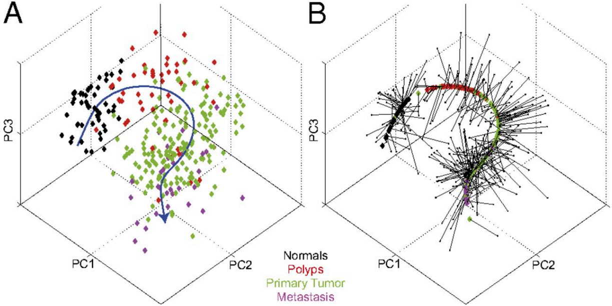
$$\hat{\lambda} = 0.003[0.001; 0.005] \quad \exp(\hat{\beta}) = 1.44[0.90; 3.40]$$

A change-point model for detecting heterogeneity in ordered survival responses. O. Bouaziz and G. Nuel. **Statistical Methods in Medical Research** (2017)

Perspective : detection of multi-dimensional heterogeneity

Detection of G$\times$E interaction for cancer data (European Prospective Investigation into Cancer and Nutrition (EPIC) data).

- Implement a Cox model with SNPs as a covariate
- Order data with respect to environmental data
  - Each individual is represented as a multidimensional point (environmental data)
  - Fit a principal curve in this multi-dimensional space
  - Project the data on the curve.



Outline

The piecewise constant hazard model

- The model :

$$\lambda(t) = \sum_{k=1}^{K} \lambda_k \mathbb{1}_{c_{k-1} < t \leq c_k}$$

- Goal : estimate the λ_ks.

The log-likelihood is equal to :

$$\ell_n(\boldsymbol{\lambda}) = \sum_{k=1}^{K} \left\{ \bar{O}_k \log\left(\lambda_k\right) - \lambda_k \bar{R}_k \right\},$$

where

- $\bar{O}_k = \sum_i \Delta_i \mathbb{1}_{c_{k-1} < T_i^{\text{obs}} \leq c_k}$: number of observed events in $(c_{k-1}, c_k]$
- $\bar{R}_k = \sum_i (T_i^{\text{obs}} \wedge c_k - c_{k-1})$: total time at risk in $(c_{k-1}, c_k]$

The piecewise constant hazard model

- The model :

$$\lambda(t) = \sum_{k=1}^{K} \lambda_k \mathbb{1}_{c_{k-1} < t \leq c_k}$$

- Goal : estimate the λ_ks.

The log-likelihood is equal to :

$$\ell_n(\boldsymbol{\lambda}) = \sum_{k=1}^{K} \left\{ \bar{O}_k \log(\lambda_k) - \lambda_k \bar{R}_k \right\},$$

where

- $\bar{O}_k = \sum_i \Delta_i \mathbb{1}_{c_{k-1} < T_i^{\text{obs}} \leq c_k}$: number of observed events in $(c_{k-1}, c_k]$
- $\bar{R}_k = \sum_i (T_i^{\text{obs}} \wedge c_k - c_{k-1})$: total time at risk in $(c_{k-1}, c_k]$

The maximum likelihood estimator is explicit :

$$\hat{\lambda}_k^{\text{mle}} = \frac{\bar{O}_k}{\bar{R}_k}$$

Penalizing the maximum likelihood estimator

- We want to choose the number and location of the cuts from the data
- We start from a large grid of cuts ($K = 100, 1\,000, \ldots$)

Penalizing the maximum likelihood estimator

- We want to choose the number and location of the cuts from the data
- We start from a large grid of cuts ($K = 100, 1\,000, \ldots$)

Set $\log \lambda_k = a_k$. Estimation of $\boldsymbol{a}$ is achieved through penalized

log-likelihood :

$$\ell_n^{\text{pen}}(\boldsymbol{a}) = \underbrace{\ell_n(\boldsymbol{a})}_{\text{log-likelihood}} - \underbrace{\frac{\text{pen}}{2}\left\{\sum_{k=1}^{K-1} w_k \left(a_{k+1} - a_k\right)^2\right\}}_{\text{regularization term}},$$

Penalizing the maximum likelihood estimator

- We want to choose the number and location of the cuts from the data
- We start from a large grid of cuts ($K = 100, 1\,000, \dots$)

Set $\log \lambda_k = a_k$. Estimation of $\boldsymbol{a}$ is achieved through penalized

log-likelihood :

$$\ell_n^{\text{pen}}(\boldsymbol{a}) = \underbrace{\ell_n(\boldsymbol{a})}_{\text{log-likelihood}} - \underbrace{\frac{\text{pen}}{2}\left\{\sum_{k=1}^{K-1} w_k \left(a_{k+1} - a_k\right)^2\right\}}_{\text{regularization term}},$$

- $\boldsymbol{w}$ represents a weight
- pen is a penalty term

Two types of regularization

1. L_2 regularization (Ridge) with $\boldsymbol{w} = \mathbf{1}$
2. L_0 regularization with the adaptive ridge procedure.
   Iterative updates of the weights :

$$w_k = \left(\left(a_{k+1} - a_k \right)^2 + \varepsilon^2 \right)^{-1},$$

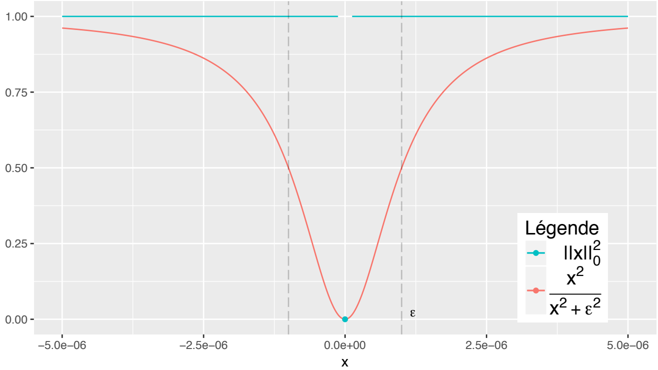
   with $\varepsilon \ll 1$.

F. Frommlet and G. Nuel, *An Adaptive Ridge Procedure for L0 Regularization*. **PlosOne** (2016).

L_0 norm approximation

When $\varepsilon \ll 1$:

$$w_k\left(a_{k+1}-a_k\right)^2 \simeq \|a_{k+1}-a_k\|_0^2 = \begin{cases} 0 & \text{if } a_{k+1}=a_k \\ 1 & \text{if } a_{k+1}\neq a_k \end{cases}$$



The *Adaptive Ridge* procedure

- Maximization of the penalized log-likelihood is performed from the Newton-Raphson algorithm.
- The Hessian matrix is tri-diagonal
- $\implies$ computation time for the inversion of the Hessian is $\mathcal{O}(K)$

The *Adaptive Ridge* procedure

- Maximization of the penalized log-likelihood is performed from the Newton-Raphson algorithm.
- The Hessian matrix is tri-diagonal
- $\Longrightarrow$ computation time for the inversion of the Hessian is $\mathcal{O}(K)$

procedure $\textsc{Adaptive-Ridge}(\boldsymbol{O}, \boldsymbol{R}, \text{pen})$
  $(\boldsymbol{a},\boldsymbol{w}) \leftarrow (\mathbf{0},\mathbf{1})$
  while not converge **do**
    $\boldsymbol{a}^{\text{new}} \leftarrow \textsc{Newton-Raphson}(\boldsymbol{O}, \boldsymbol{R}, \text{pen}, \boldsymbol{w})$
    $w_k^{\text{new}} \leftarrow \left(\left(a_{k+1}^{\text{new}} - a_k^{\text{new}}\right)^2 + \varepsilon^2\right)^{-1}$
    $(\boldsymbol{a},\boldsymbol{w}) \leftarrow (\boldsymbol{a}^{\text{new}},\boldsymbol{w}^{\text{new}})$
  end while

end procedure

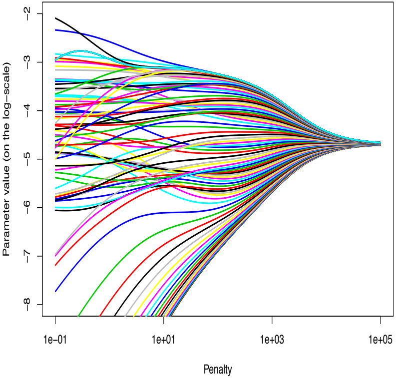
The *Adaptive Ridge* procedure

- Maximization of the penalized log-likelihood is performed from the Newton-Raphson algorithm.
- The Hessian matrix is tri-diagonal
- $\Longrightarrow$ computation time for the inversion of the Hessian is $\mathcal{O}(K)$

procedure $\textsc{Adaptive-Ridge}(\boldsymbol{O}, \boldsymbol{R}, \text{pen})$
  $(\boldsymbol{a},\boldsymbol{w}) \leftarrow (\mathbf{0},\mathbf{1})$
  while not converge **do**
    $\boldsymbol{a}^{\text{new}} \leftarrow \textsc{Newton-Raphson}(\boldsymbol{O}, \boldsymbol{R}, \text{pen}, \boldsymbol{w})$
    $w_k^{\text{new}} \leftarrow \left(\left(a_{k+1}^{\text{new}} - a_k^{\text{new}}\right)^2 + \varepsilon^2\right)^{-1}$
    $(\boldsymbol{a},\boldsymbol{w}) \leftarrow (\boldsymbol{a}^{\text{new}},\boldsymbol{w}^{\text{new}})$
  end while
  Compute $(\boldsymbol{O}^{\text{sel}}, \boldsymbol{R}^{\text{sel}})$ **from** $(\boldsymbol{a}^{\text{new}}, \boldsymbol{w}^{\text{new}})$
  $\exp(\hat{\boldsymbol{a}}^{\text{mle}}) \leftarrow \boldsymbol{O}^{\text{sel}}/\boldsymbol{R}^{\text{sel}}$
  return $\hat{\boldsymbol{a}}^{\text{mle}}$
end procedure

Comparison of the two regularization methods

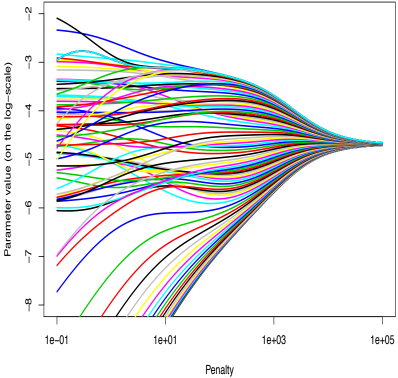
$$\text{pen} = 0 \implies \widehat{\boldsymbol{a}} = \widehat{\boldsymbol{a}}^{\text{mle}}$$
$$\text{pen} = \infty \implies \widehat{\boldsymbol{a}} = \text{ constant}$$



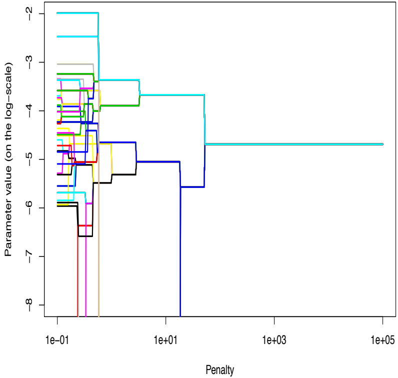
L_2 regularization

Comparison of the two regularization methods

$$\text{pen} = 0 \quad \Longrightarrow \quad \widehat{\boldsymbol{a}} = \widehat{\boldsymbol{a}}^{\text{mle}}$$
$$\text{pen} = \infty \quad \Longrightarrow \quad \widehat{\boldsymbol{a}} = \text{ constant}$$

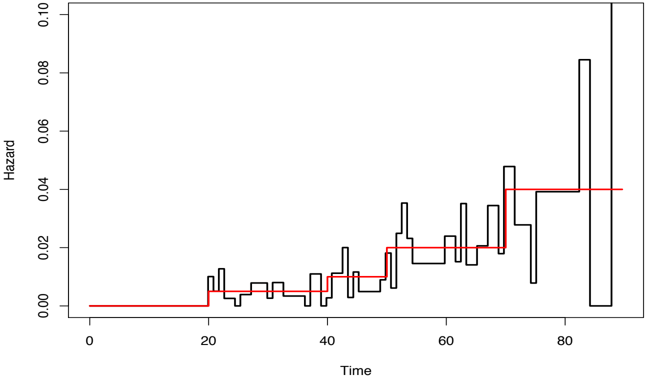


L_2 regularization



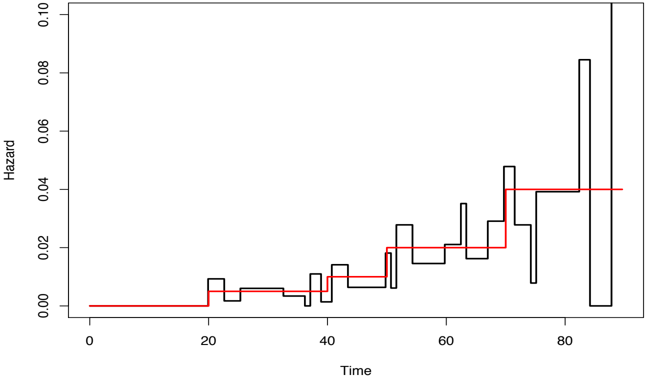
L_0 regularization

Model selection for the *Adaptive Ridge* estimator ($n = 400$)



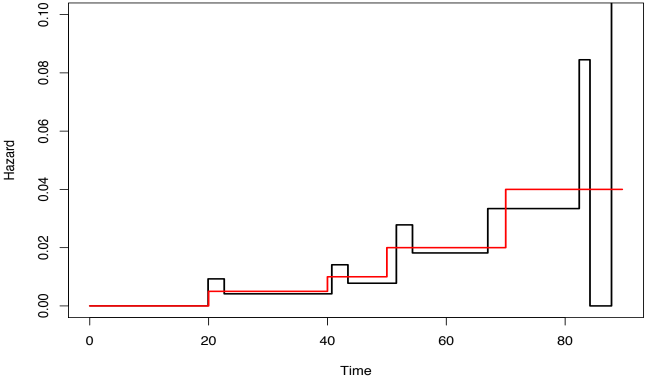
- In red the true hazard function
- In black the hazard estimator for $\mathrm{pen} = 0.1$

Model selection for the *Adaptive Ridge* estimator ($n = 400$)



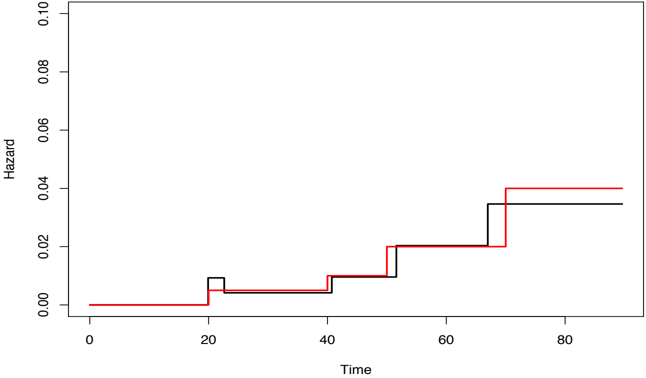
- In red the true hazard function
- In black the hazard estimator for $\mathrm{pen} = 0.27$

Model selection for the *Adaptive Ridge* estimator ($n = 400$)



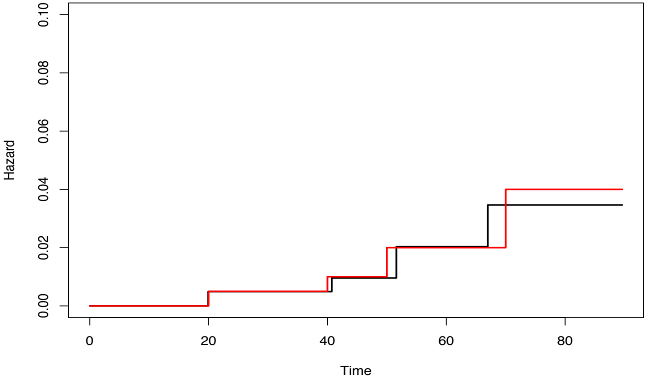
- In red the true hazard function
- In black the hazard estimator for $\mathrm{pen} = 0.55$

Model selection for the *Adaptive Ridge* estimator ($n = 400$)



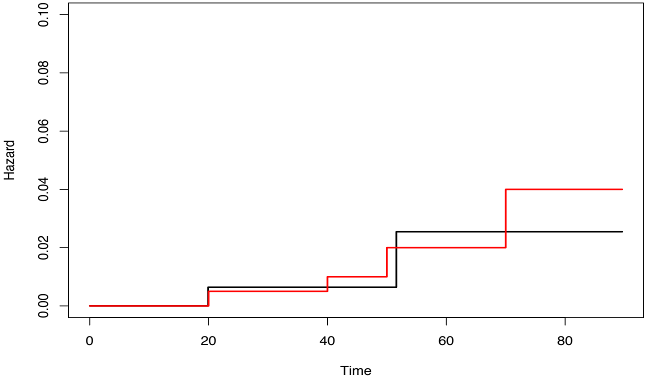
- In red the true hazard function
- In black the hazard estimator for $\mathrm{pen} = 0.77$

Model selection for the *Adaptive Ridge* estimator ($n = 400$)



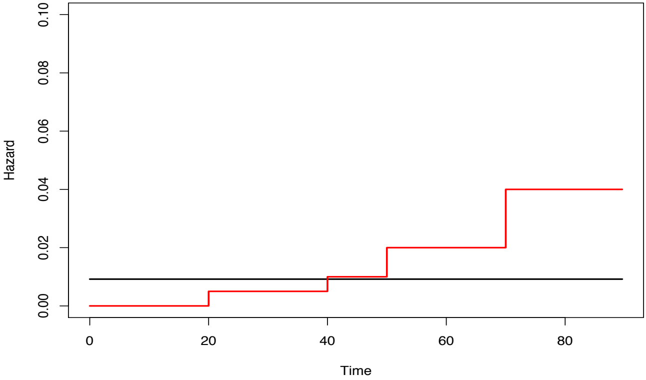
- In red the true hazard function
- In black the hazard estimator for $\mathrm{pen} = 1.54$

Model selection for the *Adaptive Ridge* estimator ($n = 400$)



- In red the true hazard function
- In black the hazard estimator for $\mathrm{pen} = 6.16$

Model selection for the *Adaptive Ridge* estimator ($n = 400$)



- In red the true hazard function
- In black the hazard estimator for $\mathrm{pen} = 52.70$

Model selection for the *Adaptive Ridge* estimator

Three different methods to perform model selection :

1. $\mathrm{BIC}(m) = -2\ell_n(\widehat{\boldsymbol{a}}_m^{\mathrm{mle}}) + m \log n$
2. $\mathrm{AIC}(m) = -2\ell_n(\widehat{\boldsymbol{a}}_m^{\mathrm{mle}}) + 2m$
3. K-fold Cross Validation (CV),

with m the dimension of the model :

$$m = \sum_{k=0}^{K-1} \mathbb{1}\{\hat{a}_{k+1,\mathrm{pen}} - \hat{a}_{k,\mathrm{pen}} \neq 0\}.$$

Model selection for the *Adaptive Ridge* estimator

Three different methods to perform model selection :

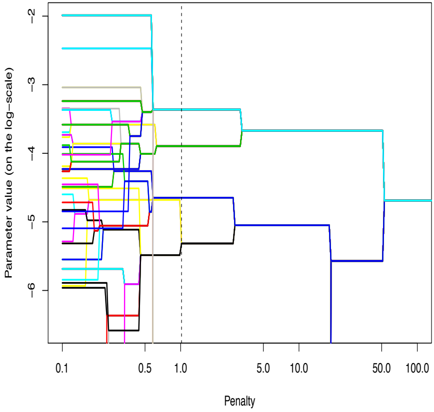
1. $\text{BIC}(m) = -2\ell_n(\widehat{\boldsymbol{a}}_m^{\text{mle}}) + m \log n$
2. $\text{AIC}(m) = -2\ell_n(\widehat{\boldsymbol{a}}_m^{\text{mle}}) + 2m$
3. K-fold Cross Validation (CV),

with m the dimension of the model :

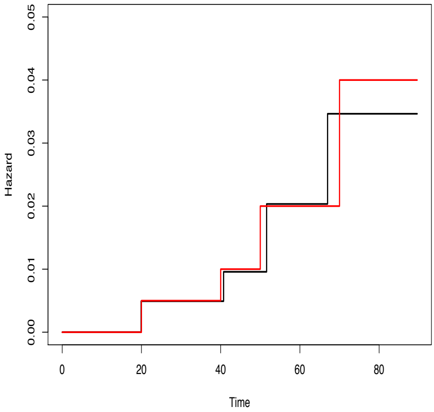
$$m = \sum_{k=0}^{K-1} \mathbb{1}\{\hat{a}_{k+1,\text{pen}} - \hat{a}_{k,\text{pen}} \neq 0\}.$$

L0 regularization for the estimation of piecewise constant hazard rates in survival analysis. O. Bouaziz and G. Nuel. **Applied Statistics** (2017)

Model selection for the *Adaptive Ridge* estimator using the BIC ($n = 400$)



Regularization path



Hazard estimator (in black)

The SEER data

- Huge american registry dataset of breast cancer
  https ://seer.cancer.gov
- Primary, unilateral, malignant and invasive cancers
- 1.2 million of patients, 60% of censoring
- The cancer diagnostics range from 1973 to 2014
- The time from cancer diagnostics to death or censoring ranges from 0 to 41 years.
- The variable of interest is the time from cancer diagnosis until death.

Aim : estimate the hazard of death as a function of both date of cancer diagnosis and time since diagnosis.

- We use the adaptive ridge procedure
- Penalization over the two directions.

The SEER data

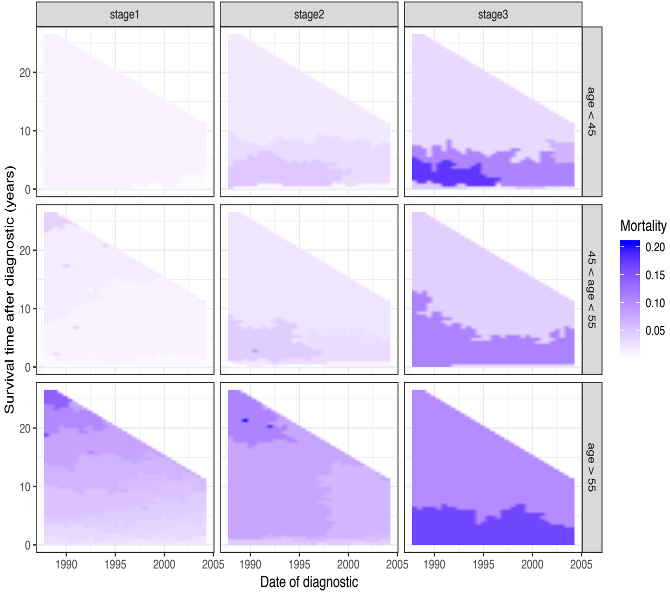
- Huge american registry dataset of breast cancer **https ://seer.cancer.gov**
- Primary, unilateral, malignant and invasive cancers
- 1.2 million of patients, 60% of censoring
- The cancer diagnostics range from 1973 to 2014
- The time from cancer diagnostics to death or censoring ranges from 0 to 41 years.
- The variable of interest is the time from cancer diagnosis until death.

Aim : estimate the hazard of death as a function of both date of cancer diagnosis and time since diagnosis.

- We use the adaptive ridge procedure
- Penalization over the two directions.

V. Goepp, J-C. Thalabard, G. Nuel and O. Bouaziz. *Regularized Bidimensional Estimation of the Hazard Rate*. **Submitted**.

The SEER data



The dental dataset

Data collected from Eva Lauridsen at the hospital Rigshospitalet (Denmark).

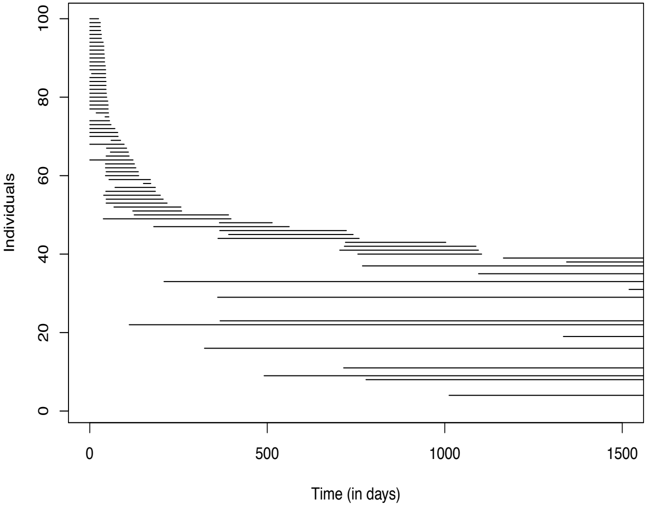
- Study of 322 patients with 400 avulsed and replanted permanent teeth from 1965 to 1988.
- The variable of interest is time from replantation until the ankylosis complication.
- Patients are examined at intermittent visits to the dentist.
  - Left-censoring (28%) if ankylosis occurred before the first visit.
  - Interval-censoring (35.75%) if ankylosis occurred between two visits.
  - Right-censoring (36.25%) if ankylosis did not occur yet after the last visit.

The dental dataset

Data collected from Eva Lauridsen at the hospital Rigshospitalet (Denmark).

- Study of 322 patients with 400 avulsed and replanted permanent teeth from 1965 to 1988.
- The variable of interest is time from replantation until the ankylosis complication.
- Patients are examined at intermittent visits to the dentist.
  - Left-censoring (28%) if ankylosis occurred before the first visit.
  - Interval-censoring (35.75%) if ankylosis occurred between two visits.
  - Right-censoring (36.25%) if ankylosis did not occur yet after the last visit.
- Covariates :
  - stage of root formation : 72.5% mature teeth, 27.5% immature teeth
  - length of extra-alveolar storage : mean time is 30.9 minutes
  - type of storage media : 85.25% physiologic, 14.75% non physiologic
  - age of the patient : mean age for mature teeth is 16.81 years

The raw data on a subsample of size 100



The observed likelihood

The observations are L_i, R_i, $i = 1, \ldots, n$.

- $0 = L_i < R_i < +\infty$ for left-censored observation ($\delta_i = 1$)
- $0 < L_i < R_i < +\infty$ for interval-censored observation ($\delta_i = 1$)
- $0 < L_i < R_i = +\infty$ for right-censored observation ($\delta_i = 0$)

$$\mathrm{L}^{\mathsf{obs}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \{S(L_i|Z_i,\boldsymbol{\theta}) - S(R_i|Z_i,\boldsymbol{\theta})\}^{\delta_i} \times \Big\{S(L_i|Z_i,\boldsymbol{\theta})\Big)\Big\}^{1-\delta_i}.$$

The observed likelihood

The observations are L_i, R_i, $i = 1, \dots, n$.

- $0 = L_i < R_i < +\infty$ for left-censored observation ($\delta_i = 1$)
- $0 < L_i < R_i < +\infty$ for interval-censored observation ($\delta_i = 1$)
- $0 < L_i < R_i = +\infty$ for right-censored observation ($\delta_i = 0$)

$$\mathrm{L}^{\mathsf{obs}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \exp\left(- \int_0^{L_i} \lambda_0(t) dt e^{\beta Z_i} \right) \left(1 - \exp\left(- \int_{L_i}^{R_i} \lambda_0(t) dt e^{\beta Z_i} \right) \right) \right\}^{\delta_i} \times \left\{ \exp\left(- \int_0^{L_i} \lambda_0(t) dt e^{\beta Z_i} \right) \right\}^{1-\delta_i},$$

for the Cox model $\lambda(t|Z_i) = \lambda_0(t) \exp(\beta Z_i)$.

The observed likelihood

The observations are L_i, R_i, $i = 1, \ldots, n$.

- $0 = L_i < R_i < +\infty$ for left-censored observation ($\delta_i = 1$)
- $0 < L_i < R_i < +\infty$ for interval-censored observation ($\delta_i = 1$)
- $0 < L_i < R_i = +\infty$ for right-censored observation ($\delta_i = 0$)

$$\mathrm{L}^{\text{obs}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \exp\left(- \int_0^{L_i} \lambda_0(t) dt e^{\beta Z_i} \right) \left(1 - \exp\left(- \int_{L_i}^{R_i} \lambda_0(t) dt e^{\beta Z_i} \right) \right) \right\}^{\delta_i} \times \left\{ \exp\left(- \int_0^{L_i} \lambda_0(t) dt e^{\beta Z_i} \right) \right\}^{1-\delta_i},$$

for the Cox model $\lambda(t|Z_i) = \lambda_0(t) \exp(\beta Z_i)$.

With a piecewise-constant baseline λ_0, maximization is performed using the Newton-Raphson algorithm.

- The Hessian is of full rank !
- Intractable solution if K is large !

The EM algorithm

The complete likelihood is defined as

$$\mathrm{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(T_i | Z_i, \boldsymbol{\theta}).$$

Introduce data $= (L_i, R_i, Z_i)$.

- E-step :

$$\mathbb{E}[\log(f(T_i|Z_i, \boldsymbol{\theta}))|\text{data}, \boldsymbol{\theta}_{\text{old}}] = \int f(t|\text{data}, \boldsymbol{\theta}_{\text{old}}) \log f(t|Z_i, \boldsymbol{\theta}) dt$$

- Under the assumption $\mathbb{P}(T \in [L, R]) = 1$,

$$f(t|\text{data}, \boldsymbol{\theta}_{\text{old}}) = \frac{f(t|Z_i, \boldsymbol{\theta}_{\text{old}}) \mathbb{1}(L_i < t < R_i)}{S(L_i|Z_i, \boldsymbol{\theta}_{\text{old}}) - S(R_i|Z_i, \boldsymbol{\theta}_{\text{old}})}.$$

The EM algorithm

- The M-step corresponds of maximizing, with respect to $\boldsymbol{\theta} = (a_1, \ldots, a_K, \beta)$,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}}) := \mathbb{E}_{T_{1:n}|\text{data},\boldsymbol{\theta}_{\text{old}}}[\log(\mathrm{L}(\boldsymbol{\theta}))]$$

The EM algorithm

- The M-step corresponds of maximizing, with respect to $\boldsymbol{\theta} = (a_1, \ldots, a_K, \beta)$,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \left(a_{i,k} - \sum_{j=1}^{k-1} (c_j - c_{j-1}) e^{a_{i,j}} \right) A_{k,i}^{\text{old}} - e^{a_{i,k}} B_{k,i}^{\text{old}} \right\},$$

  with explicit expressions of $A_{k,i}^{\text{old}}$ and $B_{k,i}^{\text{old}}$ and $a_{i,k} = a_k + \beta Z_i$.

- $A_{k,i}^{\text{old}}$ and $B_{k,i}^{\text{old}}$ depends only on $\boldsymbol{\theta}_{\text{old}}, L_i, R_i, Z_i$.

The EM algorithm

- The M-step corresponds of maximizing, with respect to $\boldsymbol{\theta} = (a_1, \ldots, a_K, \beta)$,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \left(a_{i,k} - \sum_{j=1}^{k-1} (c_j - c_{j-1}) e^{a_{i,j}} \right) A_{k,i}^{\text{old}} - e^{a_{i,k}} B_{k,i}^{\text{old}} \right\},$$

  with explicit expressions of $A_{k,i}^{\text{old}}$, $B_{k,i}^{\text{old}}$ and $a_{i,k} = a_k + \beta Z_i$.

- $A_{k,i}^{\text{old}}$ and $B_{k,i}^{\text{old}}$ depends only on $\boldsymbol{\theta}_{\text{old}}, L_i, R_i, Z_i$.
- In the absence of covariates ($Z_i = 0$, $a_{i,k} = a_k$, $\boldsymbol{\theta} = (a_1, \ldots, a_K)$) : the M-step is explicit.
- In the general regression framework : The M-step is solved using the Newton-Raphson procedure.
  - The block matrix of the Hessian for the a_ks is diagonal !
  - Using the Schurr complement, inversion of the Hessian is of order $\mathcal{O}(K)$ in the case $K >> d$.

Implementation

- The adaptive ridge procedure consists in maximizing

$$\ell(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}}) - \frac{\text{pen}}{2} \sum_{k=1}^{K-1} w_k (a_{k+1} - a_k)^2.$$

- We use the BIC to choose the correct model :

$$\text{BIC(m)} = -2\log(\mathrm{L}^{\text{obs}}(\hat{\boldsymbol{\theta}}_m^{\text{mle}})) + m\log(n),$$

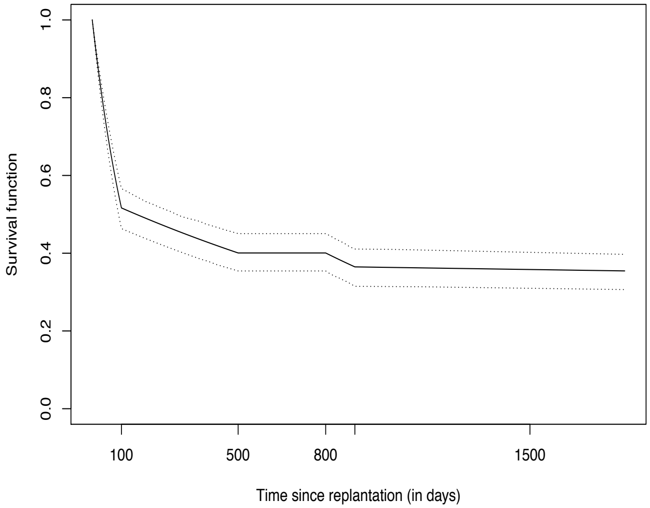
where

$$m = \sum_{k=0}^{K-1} \mathbb{1}\{\hat{a}_{k+1,\text{pen}} - \hat{a}_{k,\text{pen}} \neq 0\}.$$

- The MLE is re-implemented for each selection.
- The whole procedure combines
  - Newton-Raphson for each EM step : we stop the EM step as soon as Q increases $\Longrightarrow$ GEM.
  - EM : we use `turboEM` R package with the `squareEM` option.
  - the adaptive ridge procedure.

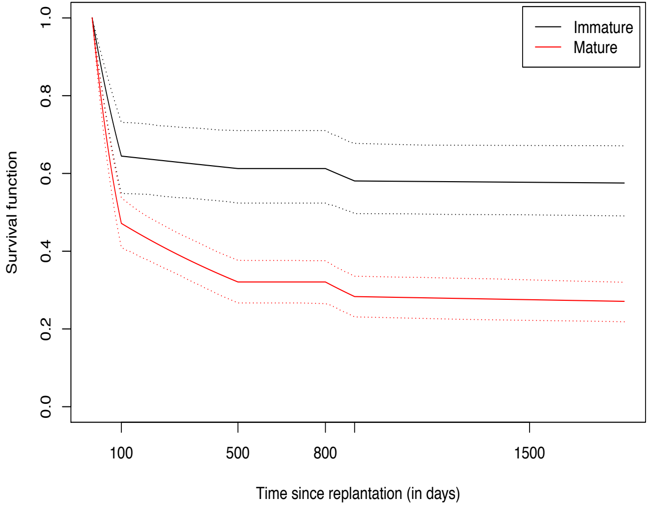
Dental dataset - non-parametric study

- The adaptive ridge method finds four cuts : $100, 500, 800, 900$.
- 95% confidence intervals computed using the bootstrap.



Dental dataset - non-parametric study

- The adaptive ridge method finds four cuts : $100, 500, 800, 900$.
- 95% confidence intervals computed using the bootstrap.



Dental dataset - Cox model

Covariates	HR	95% CI	p-value
Mature	2.00	$[1.74; 2.29]$	1.89×10^{-5}
Storage time (hours)	1.23	$[1.11; 1.34]$	0.0017
Physiologic storage	0.93	$[0.81; 1.06]$	0.6980
Age$>$20 (mature teeth)	1.27	$[0.99; 1.61]$	0.1272

- p-values are computed from likelihood ratio tests.
- CIs are also computed using a likelihood ratio test approach.

Dental dataset - Cox model

Covariates	HR	95% CI	p-value
Mature	2.00	[1.74; 2.29]	1.89×10^{-5}
Storage time (hours)	1.23	[1.11; 1.34]	0.0017
Physiologic storage	0.93	[0.81; 1.06]	0.6980
Age>20 (mature teeth)	1.27	[0.99; 1.61]	0.1272

- p-values are computed from likelihood ratio tests.
- CIs are also computed using a likelihood ratio test approach.

Regression modelling of interval censored data based on the adaptive ridge procedure. O. Bouaziz, Eva Lauridsen and G. Nuel.

Extensions : exact observations and the cure model

- With exact observations, Q can be decomposed as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{\text{old}}) = \sum_{i \text{ not exact}} \sum_{k=1}^{K} \left\{ \left(a_{i,k} - \sum_{j=1}^{k-1} (c_j - c_{j-1}) e^{a_{i,j}} \right) A_{k,i}^{\text{old}} - e^{a_{i,k}} B_{k,i}^{\text{old}} \right\}$$

$$+ \sum_{i \text{ exact}} \sum_{k=1}^{K} \left\{ O_{i,k} a_{i,k} - \exp(a_{i,k}) R_{i,k} \right\}.$$

- The Cox cure model (with $Y \in \{0,1\}$ a latent variable)

$$\lambda(t|Y,Z) = Y\lambda(t|Y=1,Z) = Y\lambda_0(t)\exp(\beta Z).$$

The complete likelihood is defined as

$$\mathrm{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} p_i^{Y_i}(1-p_i)^{1-Y_i} \prod_{i=1}^{n} \{f(T_i|Y_i=1, Z_i; \boldsymbol{\theta})\}^{Y_i}.$$

Perspectives

- Mixture model :

$$\lambda(t|Z, Y = k) = \lambda_k(t|Z) \exp(\beta_k Z).$$

This model is not identifiable when using the non-parametric baseline. We can combine piecewise constant hazard models and adaptive ridge !

- Frailty models :

$$\lambda(t|Z_i, W_i) = \lambda_0(t|Z_i) \exp(W_i + \beta Z_i),$$

with W_i a random variable following a log-normal/log-gamma distribution.

Outline

Atrial fibrillation (AF) database

- Patients with atrial fibrillation occasionally experience episodes of rapid and irregular heart rate which may lead to hospitalization to the cardiology ward.
- Data collected from January 2009 until March 2014 of 175 patients who were enrolled, having either paroxysmal or persistent AF.
- The event of interest is the time since study entry until hospitalization due to AF attack.
- Censoring : patients are censored at the end of the study.
- Terminal events : patients can move to permanent AF status or they can die.

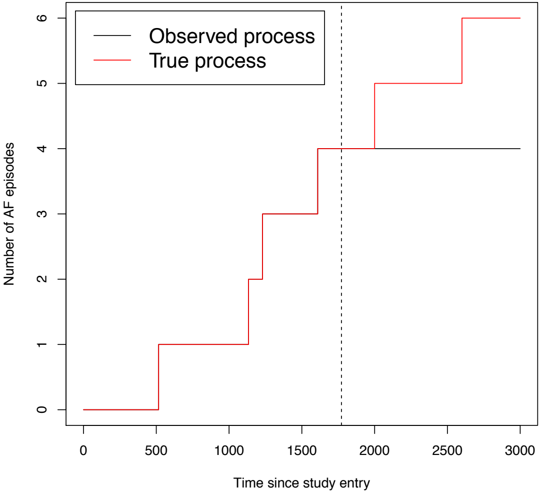
Atrial fibrillation (AF) database

- Patients with atrial fibrillation occasionally experience episodes of rapid and irregular heart rate which may lead to hospitalization to the cardiology ward.
- Data collected from January 2009 until March 2014 of 175 patients who were enrolled, having either paroxysmal or persistent AF.
- The event of interest is the time since study entry until hospitalization due to AF attack.
- Censoring : patients are censored at the end of the study.
- Terminal events : patients can move to permanent AF status or they can die.
- Covariates : type of AF (persistent, paroxysmal), gender, age at inclusion, alcohol consomption, tobacco use, hypertension, heart failure, heart valve disease, ischemic heart disease, hyperthyroidism, diabetes mellitus, COPD, kidney disease.

Counting process of interest v.s. observed counting process

start	end	event
0	516	TRUE
516	1134	TRUE
1134	1230	TRUE
1230	1609	TRUE
1609	1772	FALSE

- $N^*(\cdot)$ is the true process
- $N(\cdot)$ is the observed process.



A model for recurrent events with a terminal event

The time $t \geq 0$ represents time since study entry. Define for $i = 1, \ldots, n$:

- The process of interest : $N_i^*(t)$, $t \geq 0$. It represents the number of hospital admissions due to AF attacks since study entry.
- The time to terminal event : T_i and the non-observed at-risk process $Y_i^*(t) = \mathbb{1}\{T_i \geq t\}$
- The external covariate vector : $Z_i(t) = (Z_i^1(t), \ldots, Z_i^p(t))^T$.

A model for recurrent events with a terminal event

The time $t \geq 0$ represents time since study entry. Define for $i = 1, \ldots, n$:

- The process of interest : $N_i^*(t)$, $t \geq 0$. It represents the number of hospital admissions due to AF attacks since study entry.
- The time to terminal event : T_i and the non-observed at-risk process $Y_i^*(t) = \mathbb{1}\{T_i \geq t\}$
- The external covariate vector : $Z_i(t) = (Z_i^1(t), \ldots, Z_i^p(t))^T$.

The model for recurrent events with terminal event is :

$$\mathbb{E}[dN^*(t)|Z(t), Y^*(t)] = Y^*(t)\lambda(t|Z(t))dt,$$

λ is called the rate function. See Lin, Wei, Yang, Ying, **JRSSB** (2000).

A model for recurrent events with a terminal event

- We observe

$$\begin{cases} Z_i(t) \\ T_i^{obs} = T_i \wedge C_i \\ N_i(t) = N_i^*(t \wedge T_i^{obs}), i = 1, \dots, n, \end{cases}$$

  with C_i the time to censoring.

- Let $Y(t) = \mathbb{1}\{T^{obs} \geq t\}$ be the observed at-risk process.

A model for recurrent events with a terminal event

- We observe

$$\begin{cases} Z_i(t) \\ T_i^{obs} = T_i \wedge C_i \\ N_i(t) = N_i^*(t \wedge T_i^{obs}), i = 1, \ldots, n, \end{cases}$$

  with C_i the time to censoring.

- Let $Y(t) = \mathbb{1}\{T^{obs} \geq t\}$ be the observed at-risk process.

- Under independent censoring, we have :

$$\mathbb{E}[dN(t)|Z(t), Y(t)] = Y(t)\lambda(t|Z(t))dt,$$

  where $N(t)$ and $Y(t)$ are observed processes !

Estimation of the rate function

We assume independent censoring.

- In a non-parametric context, the cumulative mean function is estimated by :

$$\widehat{\mathbb{E}[N^*(t)]} = \int_0^t \frac{\sum_{i=1}^n \hat{S}(u) dN_i(u)}{\sum_{j=1}^n Y_j(u)},$$

with $\hat{S}$ the Kaplan-Meier estimator of T.

Estimation of the rate function

We assume independent censoring.

- In a non-parametric context, the cumulative mean function is estimated by :

$$\widehat{\mathbb{E}[N^*(t)]} = \int_0^t \frac{\sum_{i=1}^n \hat{S}(u) dN_i(u)}{\sum_{j=1}^n Y_j(u)},$$

  with $\hat{S}$ the Kaplan-Meier estimator of T.

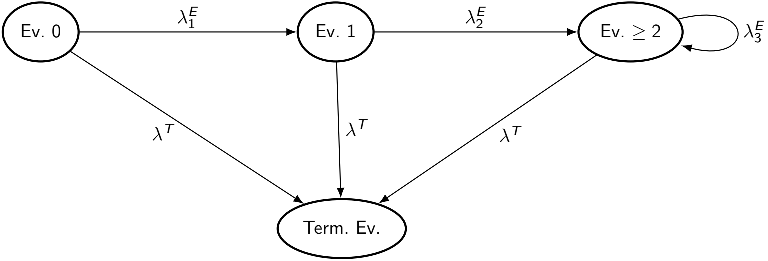
- In the Cox regression context, the regression parameter is estimated through the Cox partial likelihood :

$$\hat{\beta} = \arg\max_{\beta} \prod_{h=1}^{H} \prod_{i=1}^{n} \left(\frac{\exp(\beta Z_i(t_h))}{\sum_{j=1}^n Y_j(t_h) \exp(\beta Z_j(t_h))} \right)^{dN_i(t_h)},$$

  where $t_{(1)} < t_{(2)} < \cdots < t_{(H)}$ denote the H unique observed ordered event times.

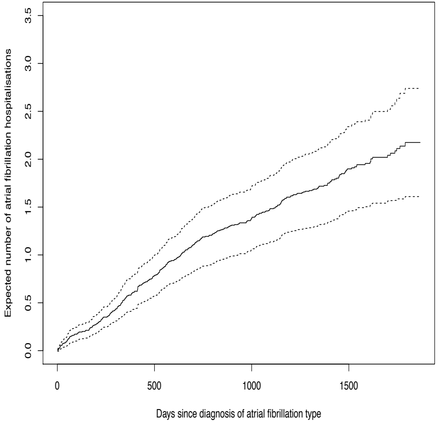
  We use a robust variance estimator.

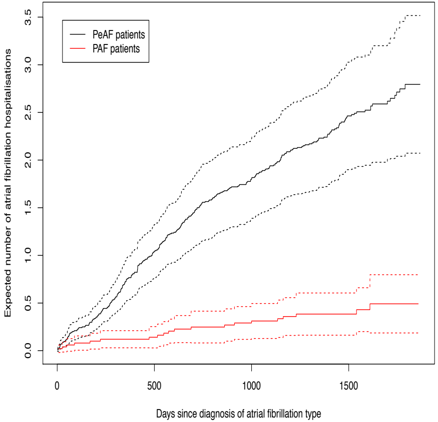
Dependence on prior counts



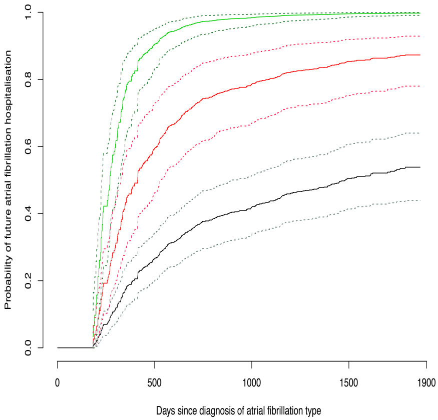
- Different rate functions λ_1^E, λ_2^E, λ_3^E modelled with a Cox model.
- A Cox model on λ^T for the terminal event.
- Predictions can be computed using the multi-state frameworks.

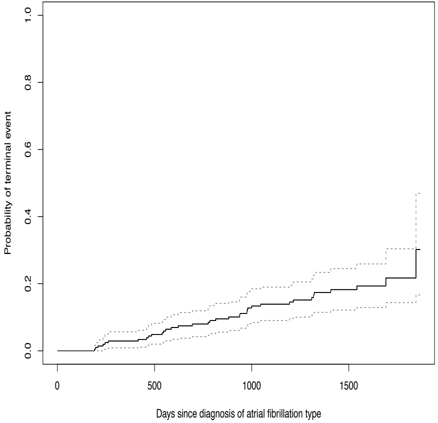
Non-parametric estimation of the cumulative mean number of AF hospitalisations





Prediction of future hospitalisations using the multi-state approach from time = 180 days





- In black : 0 hospitalisations yet
- In red : 1 hospitalisations yet
- In green : 2 or more hospitalisations yet

Statistical contributions to recurrent events data

- *Atrial fibrillation predicts atrial fibrillation - a hypothesis revisited in a clinical setting*. J. Schroder, O. Bouaziz, R. Agner, T. Martinussen, P. L. Madsen, D. Li, F. Nedaei, and U. Dixen.

Statistical contributions to recurrent events data

- *Atrial fibrillation predicts atrial fibrillation - a hypothesis revisited in a clinical setting*. J. Schroder, O. Bouaziz, R. Agner, T. Martinussen, P. L. Madsen, D. Li, F. Nedaei, and U. Dixen.
- *Is placental malaria a long-term risk factor for mild malaria attack in infancy ? Revisiting a paradigm*. O. Bouaziz, D. Courtin, G. Cottrell, J. Milet, G. Nuel, and A. Garcia, **Clinical Infectious Diseases**, 2017.

Statistical contributions to recurrent events data

- *Atrial fibrillation predicts atrial fibrillation - a hypothesis revisited in a clinical setting*. J. Schroder, O. Bouaziz, R. Agner, T. Martinussen, P. L. Madsen, D. Li, F. Nedaei, and U. Dixen.
- *Is placental malaria a long-term risk factor for mild malaria attack in infancy ? Revisiting a paradigm*. O. Bouaziz, D. Courtin, G. Cottrell, J. Milet, G. Nuel, and A. Garcia, **Clinical Infectious Diseases**, 2017.
- *Nonparametric estimation of the intensity function of a recurrent event process*. O. Bouaziz, F. Comte, A. Guilloux, **Statistica Sinica**, 2013.

Statistical contributions to recurrent events data

- *Atrial fibrillation predicts atrial fibrillation - a hypothesis revisited in a clinical setting*. J. Schroder, O. Bouaziz, R. Agner, T. Martinussen, P. L. Madsen, D. Li, F. Nedaei, and U. Dixen.
- *Is placental malaria a long-term risk factor for mild malaria attack in infancy ? Revisiting a paradigm*. O. Bouaziz, D. Courtin, G. Cottrell, J. Milet, G. Nuel, and A. Garcia, **Clinical Infectious Diseases**, 2017.
- *Nonparametric estimation of the intensity function of a recurrent event process*. O. Bouaziz, F. Comte, A. Guilloux, **Statistica Sinica**, 2013.
- *Semiparametric inference for the recurrent events process by means of a single-index model*. O. Bouaziz, S. Geffray, and O. Lopez, **Statistics**, 2015.

Statistical contributions to recurrent events data

- *Atrial fibrillation predicts atrial fibrillation - a hypothesis revisited in a clinical setting*. J. Schroder, O. Bouaziz, R. Agner, T. Martinussen, P. L. Madsen, D. Li, F. Nedaei, and U. Dixen.
- *Is placental malaria a long-term risk factor for mild malaria attack in infancy ? Revisiting a paradigm*. O. Bouaziz, D. Courtin, G. Cottrell, J. Milet, G. Nuel, and A. Garcia, **Clinical Infectious Diseases**, 2017.
- *Nonparametric estimation of the intensity function of a recurrent event process*. O. Bouaziz, F. Comte, A. Guilloux, **Statistica Sinica**, 2013.
- *Semiparametric inference for the recurrent events process by means of a single-index model*. O. Bouaziz, S. Geffray, and O. Lopez, **Statistics**, 2015.
- *A penalized algorithm for event-specific rate models for recurrent events*. O. Bouaziz and A. Guilloux. **Biostatistics**, 2014.

Perspectives

- Evaluation of the prediction performance using a Brier score for recurrent events such as :

$$\sum_{i=1}^{n} \int_0^{\tau} \left(\int_0^t \hat{S}(u)\hat{\lambda}(u|X_i)du - \int_0^t \hat{S}(u)\frac{dN_i(u)}{\sum_j Y_j(u)} \right)^2 dt.$$

- Causality in a recurrent event context.
  - For time varying HRs exchangeability does not hold anymore in standard survival analysis because the hazard rate condition on the survivors inducing a selection bias.
    The Hazards of Hazard Ratios, M.A. Hernan, **Epidemiology**, 2010.
  - What happens in terms of causality in a recurrent event context when we condition on prior events ?

Perspectives

- Evaluation of the prediction performance using a Brier score for recurrent events such as :

$$\sum_{i=1}^{n}\int_0^{\tau}\left(\int_0^t \hat{S}(u)\hat{\lambda}(u|X_i)du-\int_0^t \hat{S}(u)\frac{dN_i(u)}{\sum_j Y_j(u)}\right)^2 dt.$$

- Causality in a recurrent event context.
  - For time varying HRs exchangeability does not hold anymore in standard survival analysis because the hazard rate condition on the survivors inducing a selection bias.
    The Hazards of Hazard Ratios, M.A. Hernan, **Epidemiology**, 2010.
  - What happens in terms of causality in a recurrent event context when we condition on prior events ?

Merci de votre attention

Simulations

- Cox model : $\lambda(t|Z) = \lambda_0(t)\exp(\beta Z)$.
- The baseline

$$\lambda_0(t) = \begin{cases} 0.5 \times 10^{-2} & \text{for } t \in (0, 20], \\ 1 \times 10^{-2} & \text{for } t \in (20, 40], \\ 2 \times 10^{-2} & \text{for } t \in (40, 50], \\ 4 \times 10^{-2} & \text{for } t > 50. \end{cases}$$

- Covariate vector Z is of dimension $d = 2$ (Bernoulli and uniform)
- Regression parameter $\beta = (\log(2), \log(0.8))'$
- Two visits : $V_2 = V_1 + \mathcal{U}[0, 120]$ with $V_1 \sim \mathcal{U}[0, 60]$
  - $T_i < V_1 \implies$ left-censoring, $L_i = 0$ and $R_i = V_1$
  - $T_i > V_2 \implies$ right-censoring, $L_i = V_2$ and $R_i = +\infty$
  - $V_1 < T_i < V_2 \implies$ interval-censoring, $L_i = V_1$ and $R_i = V_2$

Simulations

- 25% left-censoring, 52% interval-censoring, 23% right-censoring.

	Midpoint		Adaptive Ridge	
n	Bias($\hat{\beta}$)	SE($\hat{\beta}$)	Bias($\hat{\beta}$)	SE($\hat{\beta}$)
200	−0.174	0.184	0.032	0.235
	0.057	0.141	−0.010	0.181
400	−0.177	0.127	0.012	0.166
	0.050	0.096	−0.014	0.120
1000	−0.171	0.075	0.007	0.099
	0.056	0.062	−0.003	0.075

- The midpoint estimator replaces (L_i, R_i) by $T_i^* = (L_i + R_i)/2$ for an interval or left-censored observation.
  - The standard Cox model is then implemented for exact and right-censored observations.

Simulations

- 25% left-censoring, 52% interval-censoring, 23% right-censoring.

	Midpoint		Adaptive Ridge		
n	$\mathrm{IBias}^2(\hat{S}_0)$	$\mathrm{IVar}(\hat{S}_0)$	$\mathrm{IBias}^2(\hat{S}_0)$	$\mathrm{IVar}(\hat{S}_0)$	$\mathrm{TV}(\hat{\lambda}_0)$
200	0.124	0.122	0.002	0.266	0.784
400	0.124	0.061	0.003	0.138	0.600
1000	0.126	0.023	0.002	0.059	0.416

$$\mathrm{IBias}^2(\hat{S}_0) = \int_0^{60} \left(\frac{1}{M} \sum_{m=1}^{M} \hat{S}_0^{(m)}(u) - S_0(u) \right)^2 du,$$

$$\mathrm{IVar}(\hat{S}_0) = \frac{1}{M} \sum_{m=1}^{M} \int_0^{60} \left(\hat{S}_0^{(m)}(u) - \frac{1}{M} \sum_{m'=1}^{M} \hat{S}_0^{(m')}(u) \right)^2 du,$$

$$\mathrm{TV}^{(m)}(\hat{\lambda}_0^{(m)}) = \sum_{k=1}^{K} (c_k - c_{k-1}) |\exp(\hat{a}_k) - \exp(a_k)|.$$