A penalized algorithm for event-specific rate models for recurrent events

Olivier Bouaziz[1] and Agathe Guilloux[2]

[1]University Paris 5, MAP5
[2]University Paris 6, LSTA

Dynstoch workshop

Bladder tumour data analysis

Data-set of Byar (1980) on bladder tumour recurrences :

- $n = 116$ patients.
- $N_i^*(t)$: number of tumour recurrences experienced by patient i before time t, where $i \in \{1, \ldots, n\}$, $t \geq 0$ (maximum$= 5$).
- $X_i(t)$: four dimensional covariates process. Number of initial tumours, size of the largest tumour, two treatment variables.
- Goal : estimation of the probability of having a tumour recurrence at any time t.

Some patients died from the bladder disease or were censored : further recurrence times are not observed.

Modeling the rate function

- Process of interest : $N^*(t),\ t \geq 0$.
- Observations :

$$\begin{cases} X_i(t) = (X_i^1(t), \ldots, X_i^p(t)) \\ T_i = D_i \wedge C_i \\ \delta_i = \mathbb{1}_{D_i \leq C_i} \\ N_i(t) = N_i^*(t \wedge T_i), i \in \{1, \ldots, n\} \end{cases}$$

Constant model of the rate function

$$\mathbb{E}\Big(dN^*(t) | D \geq t, X(t)\Big) = \mathbb{1}_{D \geq t} \rho_0(t, X(t)) dt.$$

Modeling the rate function

- Process of interest : $N^*(t),\ t \geq 0$.
- Observations :
$$\begin{cases} X_i(t) = (X_i^1(t), \ldots, X_i^p(t)) \\ T_i = D_i \wedge C_i \\ \delta_i = \mathbb{1}_{D_i \leq C_i} \\ N_i(t) = N_i^*(t \wedge T_i), i \in \{1, \ldots, n\} \end{cases}$$

Event specific model of the rate function

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t), N^*(t-) = s-1\Big) = \mathbb{1}_{D \geq t}\rho_0(t, X(t), s)dt.$$

Multiplicative model for the rate function

Constant rate model (Cox)

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t)\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t))dt$$

where

$$\rho_0(t, X(t)) = \alpha_0(t)\exp(X(t)\beta_0).$$

Event-specific rate model (PWP)

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t), N^*(t-) = s-1\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t), s)dt$$

where

$$\rho_0(t, X(t), s) = \alpha_0(t, s)\exp(X(t)\beta_0(s)).$$

Multiplicative model for the rate function

Constant rate model (Cox)

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t)\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t))dt$$

where

$$\rho_0(t, X(t)) = \alpha_0(t)\exp(X(t)\beta_0).$$

Event-specific rate model (PWP)

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t), N^*(t-) = s-1\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t), s)dt$$

where

$$\rho_0(t, X(t), s) = \alpha_0(t, s)\exp(X(t)\beta_0(s)).$$

Additive model for the rate function

Constant rate model (Aalen)

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t)\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t))dt$$

where

$$\rho_0(t, X(t)) = \alpha_0(t) + X(t)\beta_0.$$

Event-specific rate model

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t), N^*(t-) = s-1\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t), s)dt$$

where

$$\rho_0(t, X(t), s) = \alpha_0(t, s) + X(t)\beta_0(s).$$

Additive model for the rate function

Constant rate model (Aalen)

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t)\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t))dt$$

where

$$\rho_0(t, X(t)) = \alpha_0(t) + X(t)\beta_0.$$

Event-specific rate model

$$\mathbb{E}\Big(dN^*(t)|D \geq t, X(t), N^*(t-) = s-1\Big) = \mathbb{1}_{D\geq t}\rho_0(t, X(t), s)dt$$

where

$$\rho_0(t, X(t), s) = \alpha_0(t, s) + X(t)\beta_0(s).$$

Estimation procedure

A key relation

Under some assumptions, we have :

$$\mathbb{E}\Big(dN(t)|T \geq t, X(t), N(t-) = s-1\Big)$$
$$= \mathbb{E}\Big(dN^*(t)|D \geq t, X(t), N^*(t-) = s-1\Big)$$
$$= \mathbb{1}_{D \geq t}\rho_0(t, X(t), s)dt.$$

Estimation of $\beta_0(s)$ is performed from observations $\{N_i(t), i = 1 \ldots, n\}$ in stratum $s-1$.

Estimation criterion

Suppose that $N(t) \leq B$ almost surely. In the event specific models, we want to perform estimation of
$\beta_0 = (\beta_0^1(1), \ldots, \beta_0^1(B), \ldots, \beta_0^p(1), \ldots, \beta_0^p(B))^\top$.

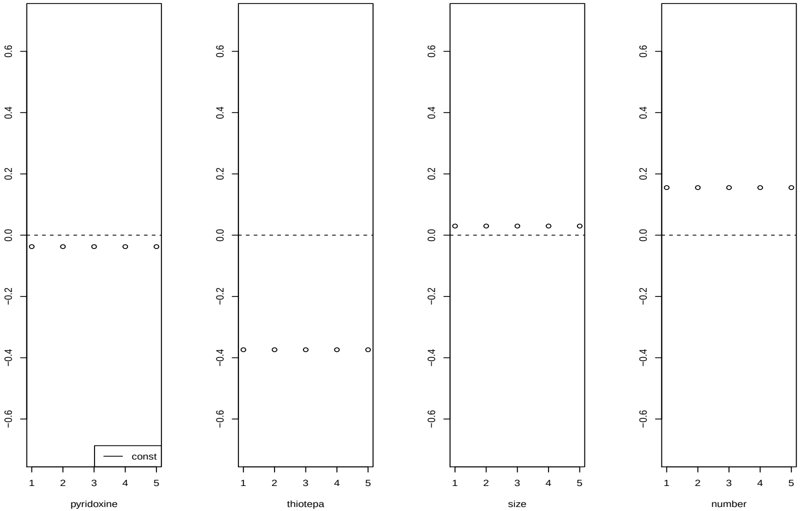
Definition of $\hat{\beta}_{ES}$

$$\hat{\beta}_{ES} = \arg\min_{\beta \in \mathbb{R}^{pB}} \Gamma_n(\beta),$$

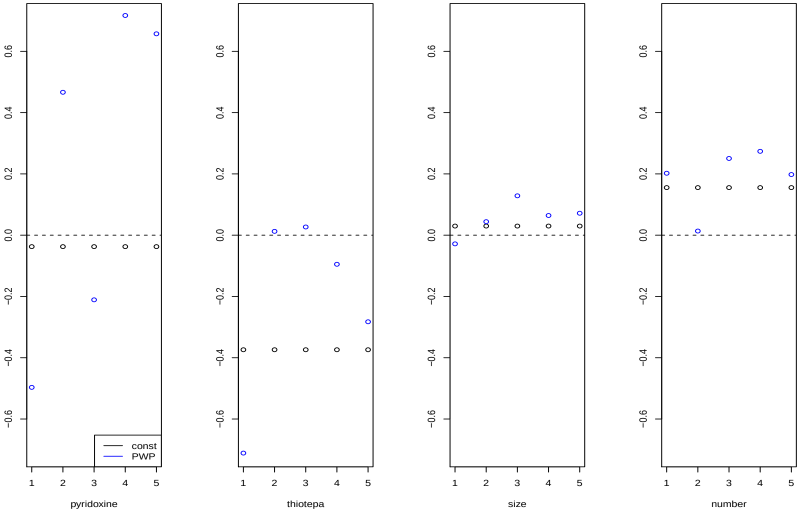
where

- $\Gamma_n(\beta)$ is a (partial) maximum likelihood estimator in the multiplicative model.
- $\Gamma_n(\beta)$ is a (partial) least squares estimator in the additive model.

Multiplicative model



Multiplicative model



Overparametrization

- $n = 116$ patients.
- $B = 5$ maximum of tumour recurrences per patient.
- $p = 4$ covariates.

$$\sqrt{n} \simeq 10.77 < p \times B = 20.$$

The event specific model is overparametrized !

Overparametrization

$$\beta_0 = (\beta_0^1(1), \ldots, \beta_0^1(B), \ldots, \beta_0^p(1), \ldots, \beta_0^p(B)).$$

Event-specific estimator

- $\hat{\beta}_{ES} \to \beta_0$, in probability
- But fluctuates too much when n is small.

Constant estimator

- $\hat{\beta}_{const} \not\to \beta_0$ in probability
- But is easier to interpret.

How to define an estimator that "fluctuates" less but is still consistent ?
For each covariate $X^j, j = 1, \ldots, p$, we want the total-variation $\sum_{s=2}^{B} |\beta^j(s) - \beta^j(s-1)|$ to be "small".

Overparametrization

$$\beta_0 = (\beta_0^1(1), \ldots, \beta_0^1(B), \ldots, \beta_0^p(1), \ldots, \beta_0^p(B)).$$

Event-specific estimator

- $\hat{\beta}_{ES} \to \beta_0$, in probability
- But fluctuates too much when n is small.

Constant estimator

- $\hat{\beta}_{const} \not\to \beta_0$ in probability
- But is easier to interpret.

How to define an estimator that "fluctuates" less but is still consistent ?

For each covariate $X^j, j = 1, \ldots, p$, we want the total-variation $\sum_{s=2}^{B} |\beta^j(s) - \beta^j(s-1)|$ to be "small".

Overparametrization

$$\beta_0 = (\beta_0^1(1), \dots, \beta_0^1(B), \dots, \beta_0^p(1), \dots, \beta_0^p(B)).$$

Event-specific estimator

- $\hat{\beta}_{ES} \to \beta_0$, in probability
- But fluctuates too much when n is small.

Constant estimator

- $\hat{\beta}_{const} \not\to \beta_0$ in probability
- But is easier to interpret.

How to define an estimator that "fluctuates" less but is still consistent ?
For each covariate $X^j, j = 1, \dots, p$, we want the total-variation $\sum_{s=2}^{B} |\beta^j(s) - \beta^j(s-1)|$ to be "small".

Total variation penalization

A penalty is introduced to constrain $\hat{\beta}^j_{ES}$ to be piecewise constant.

Definition of $\hat{\beta}_{TV}$

$$\hat{\beta}_{TV} = \underset{\beta \in \mathbb{R}^{pB}}{\arg\min} \left\{ \Gamma_n(\beta) + \frac{\lambda_n}{n} \sum_{j=1}^{p} \sum_{s=2}^{B} |\beta^j(s) - \beta^j(s-1)| \right\}.$$

Total variation penalization

A penalty is introduced to constrain $\hat{\beta}^j_{ES}$ to be piecewise constant.

Definition of $\hat{\beta}_{TV}$

$$\hat{\beta}_{TV} = \underset{\beta \in \mathbb{R}^{pB}}{\arg\min} \left\{ \Gamma_n(\beta) + \frac{\lambda_n}{n} \sum_{j=1}^{p} \sum_{s=2}^{B} \underbrace{|\beta^j(s) - \beta^j(s-1)|}_{=\Delta\beta^j_{TV}(s)} \right\}.$$

- If $\lambda_n = 0$, $\hat{\beta}_{TV} = \hat{\beta}_{ES}$.
- If $\lambda_n / n = \infty$, $\hat{\beta}_{TV} = \hat{\beta}_{Const}$.

Link to the Lasso

D : block matrix of size $(pB \times pB)$

$$D = \begin{pmatrix} T_B & O_B & \cdots & O_B \\ O_B & T_B & \cdots & O_B \\ \cdots & \cdots & \cdots & \cdots \\ O_B & O_B & \cdots & T_B \end{pmatrix} \text{ with } T_B = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

The minimization problems can then be rewritten as a Lasso algorithm :

$$\hat{\beta}_{TV} = D\hat{\gamma}_{TV} \text{ with}$$

$$\hat{\gamma}_{TV} = \operatorname*{arg\,min}_{\gamma \in \mathbb{R}^{pB}} \left\{ \Gamma_n(D\gamma) + \frac{\lambda_n}{n} \sum_{j=1}^{p} \sum_{s=2}^{B} |\gamma^j(s)| \right\}$$

where $\hat{\gamma}_{TV} = (\hat{\beta}^1_{TV}(1), \Delta\hat{\beta}^1_{TV}(2), \ldots, \Delta\hat{\beta}^1_{TV}(B), \ldots, \Delta\hat{\beta}^p_{TV}(B))^\top$.

Asymptotic results

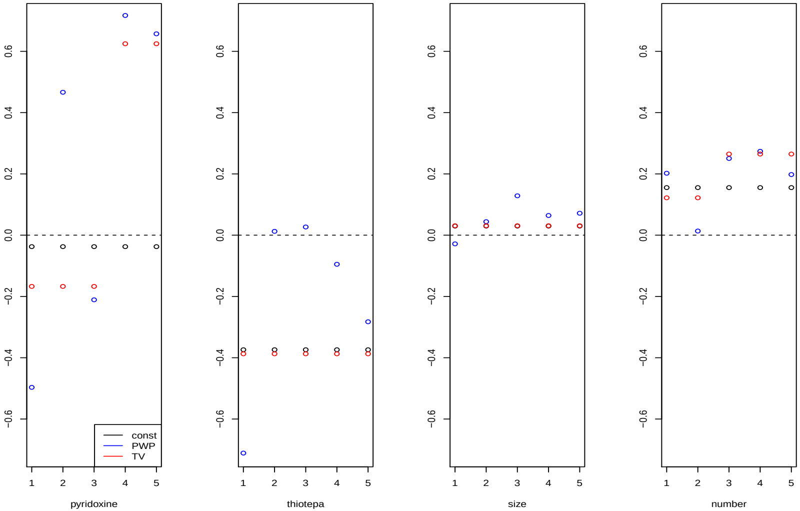
Theorem

- If $\lambda_n/n \to 0$ then

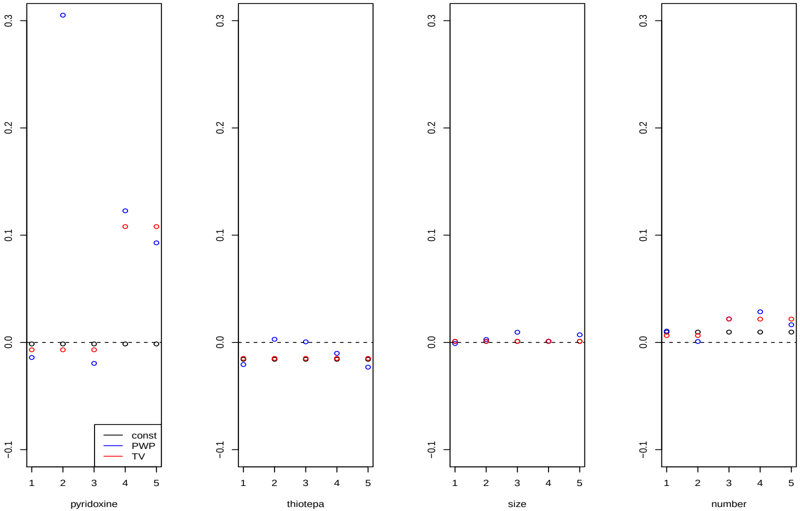
$$\hat{\beta}_{TV} \underset{n\to\infty}{\to} \beta_0 \text{ in probability.}$$

- If $\lambda_n/\sqrt{n} \to \lambda_0 \geq 0$ then $\hat{\beta}_{TV}$ converges in law to a gaussian process.

Multiplicative model



Additive model



Simulation study

Simulations

- $B = 5$
- $p = 4$, $X^j \sim$ Uniform, $j = 1, \dots, 4$.
- $n = 50 (= 2.5pB)$ to $n = 1000 \simeq (pB)^{2.3}$
- $\beta_0^1 = (0, 0, b_1, b_1, 0)$
- $\beta_0^2 = (b_2, b_2, b_2, b_2, b_2)$
- $\beta_0^3 = (1, 2, 3, 4, 5)$.
- $\beta_0^4 = (0, 0, 0, 0, 0)$.
- $D, C \sim$ Weibull.
  15% to 30% of individuals experience the fifth recurrent event.

Simulation study

Monte-Carlo experiment : $M = 200$ experiences.

$$\text{MSE} = \frac{1}{M}\sum_{m=1}^{M}\frac{\|\hat{\beta}_m - \beta_0\|^2}{\|\beta_0\|^2}.$$

Detection of false positive and false negative :

$$\text{FP}(\hat{\beta}_m) = \text{Card}\left(j \in \{1,\ldots,p\} \text{ s.t. } \text{TV}(\hat{\beta}^j) \neq 0 \text{ and } \text{TV}(\beta_0^j) = 0\right)$$

and

$$\text{FN}(\hat{\beta}_m) = \text{Card}\left(j \in \{1,\ldots,p\} \text{ s.t. } \text{TV}(\hat{\beta}^j) = 0 \text{ and } \text{TV}(\beta_0^j) \neq 0\right).$$

Simulation results in the multiplicative model

30%	Unconstrained			Constant			TV			two-step TV		
n	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	0·100	2	0	0·412	0	2	0·054	1·44	0·03	0·044	0·82	0·02
100	0·030	2	0	0·415	0	2	0·025	1·54	0	0·019	0·76	0
500	0·006	2	0	0·413	0	2	0·008	1·76	0	0·006	0·30	0
1000	0·005	2	0	0·415	0	2	0·006	1·81	0	0·006	0·05	0

15%	Unconstrained			Constant			TV			two-step TV		
n	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0·440	0	2	0·161	1·37	0·185	0·137	0·82	0·19
100	0·566	2	0	0·434	0	2	0·053	1·55	0·005	0·042	0·88	0
500	0·014	2	0	0·433	0	2	0·016	1·84	0	0·012	1·06	0
1000	0·009	2	0	0·433	0	2	0·011	1·89	0	0·010	0·68	0

Simulation results in the additive model

30%	Unconstrained			Constant			TV			two-step TV		
n	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	4·986	2	0	0·416	0	2	0·467	0·98	0·58	1·142	0·65	0·81
100	0·935	2	0	0·351	0	2	0·254	1·38	0·21	0·353	0·86	0·48
500	0·135	2	0	0·309	0	2	0·079	1·91	0·01	0·094	1·44	0·08
1000	0·071	2	0	0·299	0	2	0·049	1·98	0	0·05	1·64	0

15%	Unconstrained			Constant			TV			two-step TV		
n	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0·505	0	2	0·781	0·95	0·81	2·368	0·86	0·97
100	4·114	2	0	0·393	0	2	0·707	1·450	0·27	0·84	1·11	0·52
500	0·339	2	0	0·330	0	2	0·154	1·975	0·01	0·19	1·67	0·06
1000	0·171	2	0	0·320	0	2	0·097	1·995	0	0·12	1·80	0·02

Implementation

R-packages and functions :

- constant and event-specific estimators calculated using the **coxph** function (R package **survival**) and **ahaz** function (R package **ahaz**).
- penalized estimators calculated through the **coxnet** function (R package **glmnet**) and **ahazpen** function (R package **ahaz**).

Our programs :
http ://www.lsta.upmc.fr/guilloux.php ?main=publications

Thanks for your attention !

Implementation

R-packages and functions :

- constant and event-specific estimators calculated using the **coxph** function (R package **survival**) and **ahaz** function (R package **ahaz**).
- penalized estimators calculated through the **coxnet** function (R package **glmnet**) and **ahazpen** function (R package **ahaz**).

Our programs :
http ://www.lsta.upmc.fr/guilloux.php ?main=publications

Thanks for your attention !