Estimation adaptative de l'intensité de processus de comptage marqués

Agathe Guilloux (LSTA-UPMC)

avec **Fabienne Comte (MAP5)** et **Stéphane Gaïffas (LSTA-UPMC)**

27 août 2008

Modèle

Processus de comptage à intensité multiplicative avec covariables

Processus de comptage $\mathbf{N}^n(t) = (N_1(t), \dots, N_n(t))$ qui vérifie

$$dN^i(t) = \alpha_0(t, X_i) Y^i(z) + dM^i(t), \quad 1 \leq i \leq n,$$

pour la filtration $\mathcal{F}_t := \sigma(X_i, N^i(s), Y^i(s) : s \leq t, 1 \leq i \leq n)$ où :

- Y^i est un processus prévisible
- M^i est une $(\mathcal{F}_t)$-martingale $\in \mathcal{M}^2_{\text{loc}}$
- $\alpha_0(t, x) =$ risque instantané de saut de N^i à l'instant t avec covariable $X_i = x$
- Support $X_i \in [0, 1]^d$ (support compact)

A partir des observations i.i.d.

$$D_n := [(X_i, N^i(t), Y^i(t)) : t \in [0, T], 1 \leq i \leq n]$$

on estime α_0.

Durées de vie censurées avec covariables

Pour chaque copie i.i.d. $(X_i, T_i, C_i), 1 \leq i \leq n$, on observe

$$D_n := [(X_i, T_i^C, \delta_i) : 1 \leq i \leq n] \text{ où } T_i^C := T_i \wedge C_i \text{ et } \delta_i := I(T_i \leq C_i),$$

sous l'hypothèse (hypothèse dite *faible* en censure) :

$$T \text{ et } C \text{ indépendants } |X.$$

Dans ce cas : $N^i(t) = I(T_i^C \leq t, \delta_i = 1)$ et $Y^i(t) = I(T_i^C \geq t)$. On cherche à estimer le **taux de risque conditionel**

$$\alpha_0(t,x) = \alpha_{T|X}(t,x) = \frac{f_{T|X}(t,x)}{1 - F_{T|X}(t,x)}.$$

- Autre cas particuliers : processus de Poisson, processus de Markov, troncature...

Objets naturellement associés au modèle

1. Risque empirique :

$$R_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\int_0^T \alpha(z, X_i)^2 Y^i(z)dz - \frac{2}{n}\sum_{i=1}^{n}\int_0^T \alpha(z, X_i)dN^i(z)$$

2. Norme empirique :

$$\|\alpha\|_n^2 := \frac{1}{n}\sum_{i=1}^{n}\int_0^T \alpha(z, X_i)^2 Y^i(z)dz$$

3. Processus empirique :

$$Z(\alpha) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^T \alpha(z, X_i)dM^i(z)$$

4. Norme :

$$\|\alpha\|_\mu^2 := E\|\alpha\|_n^2 = \int\int_0^T \alpha(z, x)^2 E[Y^1(z)|X = x]dzP_X(dx)$$

Pourquoi ?

$$
\begin{aligned}
R_n(\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T \alpha(z,X_i)^2 Y^i(z)dz - \frac{2}{n}\sum_{i=1}^{n}\int_0^T \alpha(z,X_i)dN^i(z) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^T \alpha(z,X_i)^2 Y^i(z)dz - \frac{2}{n}\sum_{i=1}^{n}\int_0^T \alpha(z,X_i)\alpha_0(z,X_i)dY^i(z) \\
&- \frac{2}{n}\sum_{i=1}^{n}\int_0^T \alpha(z,X_i)dM^i(z) \\
&= \|\alpha-\alpha_0\|_n^2 - \|\alpha_0\|_n^2 - \frac{2}{\sqrt{n}}Z(\alpha)
\end{aligned}
$$

On a :

$$\mathbb{E}\left(R_n(\alpha)\right) = \|\alpha-\alpha_0\|_\mu^2 - \|\alpha_0\|_\mu^2.$$

Un exemple sans covariable et sans censure

On observe $D_n := [T_i : 1 \leq i \leq n]$ où $T_i \in \mathbb{R}_+$. On cherche à estimer le **taux de risque**

$$\alpha_0(t) = \alpha_T(t) = \frac{f_T(t)}{1 - F_T(t,)}.$$

Le risque empirique se simplifie :

$$\begin{aligned} R_n(\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T \alpha(z)^2 Y^i(z)dz - \frac{2}{n}\sum_{i=1}^{n}\int_0^T \alpha(z)dN^i(z) \\ &= \int_0^T \alpha(z)^2 \frac{1}{n} Y(z)dz - 2\int_0^T \alpha(z)\frac{1}{n}dN(z), \end{aligned}$$

la norme empirique aussi :

$$\|\alpha\|_n^2 := \int_0^T \alpha(z)^2 \frac{1}{n} Y(z)dz.$$

Projection sur une base d'histogramme

On considère ici que l'on veut estimer α_0 sur $[0,1]$.
Soit S_m le sous-espace vectoriel de $\mathbb{L}^2([0,1])$ engendré par la base d'histogramme :

$$z \mapsto \psi_k^m(z) = I\left(z \in]\frac{k-1}{2^m}, \frac{k}{2^m}]\right) \text{ pour } k = 1, \dots, 2^m - 1.$$

On peut alors orthonormaliser cette base pour le produit scalaire aléatoire : on définit

$$b_k^m = \frac{1}{n}\int_0^1 \psi_k^m(z) Y(z) dz \text{ et } x \mapsto \frac{1}{\sqrt{b_k^m}} \psi_k^m(z).$$

Soit s_m la projection orthogonale de α_0 sur S_m dénifie par :

$$s_m = \sum_{k=1}^{2^m} a_k^m \frac{\psi_k^m}{\sqrt{b_k^m}}, \text{ avec } a_k^m = \frac{1}{n}\int_0^1 \frac{\psi_k^m(z)}{\sqrt{b_k^m}} \alpha_0(z) Y(z) dz.$$

On définit le minimiseur $\hat{\alpha}^m$ du risque empirique sur S_m noté

$$\hat{\alpha}^m = \sum_{k=1}^{2^m} \hat{a}_k^m \frac{\psi_k^m}{\sqrt{b_k^m}}$$

qui vérifie

$$\begin{aligned}
\hat{\alpha}^m &= \arg\min_{h \in S_m} R_n(h) = \arg\min_{h \in S_m} \left\{ \int_0^1 h(z)^2 \frac{1}{n} Y(z) dz - 2 \int_0^1 h(z) \frac{1}{n} dN(z) \right\} \\
&= \arg\min_{h \in S_m} \left\{ ||h(z)||_n^2 - 2 \int_0^1 h(z) \frac{1}{n} dN(z) \right\}.
\end{aligned}$$

On a

$$R_n(\alpha^m) = \sum_{k=1}^{2^m} (\hat{a}_k^m)^2 - 2 \sum_{k=1}^{2^m} \int_0^1 \hat{a}_k^m \frac{\psi_k^m(z)}{\sqrt{b_k^m}} \frac{1}{n} dN(z).$$

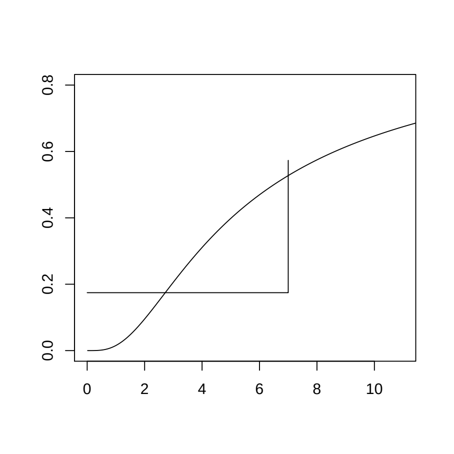
Les coefficients (a_k^m) vérifient :

$$\frac{\partial R_n(h)}{\partial \hat{a}_k^m} = 0 \Longleftrightarrow \hat{a}_k^m = \int_0^1 \frac{\psi_k^m(z)}{\sqrt{b_k^m}} \frac{1}{n} dN(z) = \frac{N_k^m}{n\sqrt{b_k^m}},$$

où N_k^m est le nombre de points du processus N dans $]\frac{k-1}{2^m}, \frac{k}{2^m}]$.

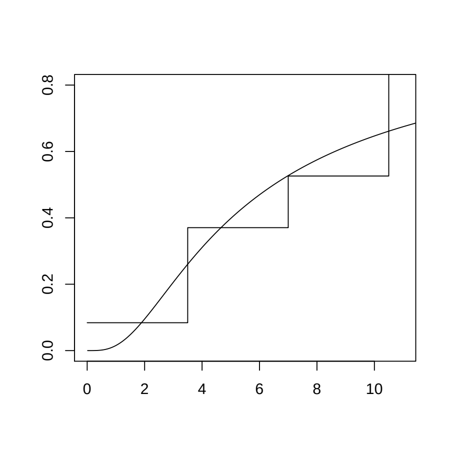
Histogramme et choix de la dimension (loi $\Gamma(5,1)$ et $n = 100$)

$m = 1, 2, 3, 4, 5$



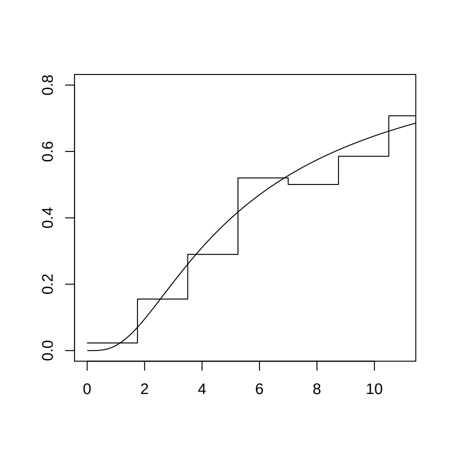
Histogramme et choix de la dimension (loi $\Gamma(5,1)$ et $n = 100$)

$m = 1, 2, 3, 4, 5$



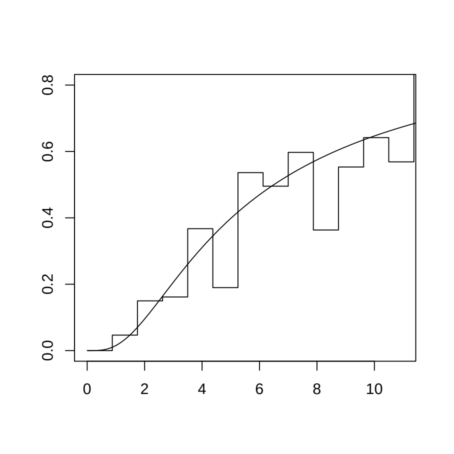
Histogramme et choix de la dimension (loi $\Gamma(5,1)$ et $n = 100$)

$m = 1, 2, 3, 4, 5$



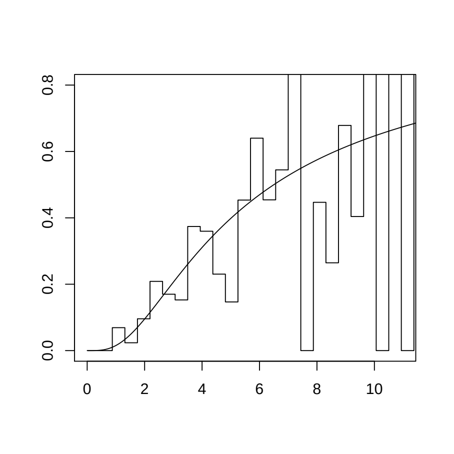
Histogramme et choix de la dimension (loi $\Gamma(5,1)$ et $n = 100$)

$m = 1, 2, 3, 4, 5$



Histogramme et choix de la dimension (loi $\Gamma(5,1)$ et $n = 100$)

$m = 1, 2, 3, 4, 5$



Un peu de biblio

- Reynaud-Bouret (2003, 2006)
- Brunel, Comte (2005)
- Huber, McGibbon (2003)
- Antoniadis, Grégoire, Nason (1999)
- Linton, Nielsen, Van de Geer (2003)
- Brunel, Comte, Lacour (2008)
- Andersen, Borgan, Gill, Keiding (1993)

Présentation pour $X \in [0,1]$

On introduit une collection d'espaces $\{S_m, m \in \mathcal{M}_n\}$ où, pour $m = (m_1, m_2)$, S_m est un espace sur $A = [0,1] \times [0,T]$ défini par :

$$S_m = F_{m_1} \otimes H_{m_2} = \{t, \quad t(y,x) = \sum_{j \in J_m} \sum_{k \in K_m} a^m_{j,k} \varphi^m_j(x) \psi^m_k(y), \ a^m_{j,k} \in \mathbb{R}\},$$

où les espaces $F_{m_1} = span\{\varphi^m_j, j \in J_{m_1}\}$ et $H_{m_2} = span\{\psi^m_k, k \in K_{m_2}\}$ vérifient les conditions habituelles de **connection de normes et d'emboîtement**.

On suppose de plus que les **covariables admettent une densité**, on a donc :

$$\begin{aligned}\|\alpha\|^2_\mu &:= \int\int_0^T \alpha(z,x)^2 E[Y^1(z)|X=x] dz P_X(dx) \\ &= \int\int_0^T \alpha(z,x)^2 d\mu(z,x) := \int\int_0^T \alpha(z,x)^2 f(z,x) dx dz.\end{aligned}$$

et que

[$\mathcal{A}1$] $\forall (x,z) \in [0,1] \times [0,T], \ d\mu(z,x) = f(z,x) dx dz \geq f_0 dx dz$.

[$\mathcal{A}2$] $\forall (x,z) \in [0,1] \times [0,T], \ \alpha(z,x) \leq \|\alpha\|_{\infty,A} < +\infty$.

On définit des premiers estimateurs comme **minimisateurs du risque empirique**

$$\hat{\alpha}_m = \arg\min_{h \in S_m} R_n(h).$$

Cela revient à résoudre le problème suivant.
Soit : $h(x,y) = \sum_{j \in J_m} \sum_{k \in K_m} \in S_m \; a_{j,k} \varphi_j^m(x) \psi_k^m(y)$:

$$\frac{\partial R_n(h)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow \sum_{j \in J_m, k \in K_m} a_{j,k} \frac{1}{n} \sum_{i=1}^{n} \varphi_j^m(X_i) \varphi_{j_0}^m(X_i) \int \psi_k^m(z) \psi_{k_0}^m(z) I_{Z_i \geq z} dz$$

$$= \frac{1}{n} \sum_{i=1}^{n} \delta_i \varphi_{j_0}^m(X_i) \psi_{k_0}^m(Z_i), \Rightarrow \forall j_0 \forall k_0, \quad \frac{\partial R_n(h)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow G_m A_m = \Upsilon_m$$

où

$$G_m = \left(\frac{1}{n} \sum_{i=1}^{n} \varphi_j^m(X_i) \varphi_l^m(X_i) \int \psi_k^m(z) \psi_p^m(z) I_{Z_i \geq z} dz \right)_{(j,k),(l,p) \in J_m \times K_m}$$

On obtient une collection d'estimateurs $\{\hat{\alpha}_m, m \geq 0\}$ et on **sélectionne la dimension par pénalisation** :

$$\hat{m} = \arg\min_{m \in \mathcal{M}_n} \left(R_n(\hat{\alpha}_m) + \mathrm{pen}(m)\right) \text{ où } \mathrm{pen}(m) = K_0 \|\alpha\|_{\infty,A} \frac{D_{m_1} D_{m_2}}{n}.$$

On peut alors montrer la borne sup adaptative :

Théorème

Supposons que $\alpha_A \in B^{\boldsymbol{\beta}}_{2,\infty}(A)$ *avec les régularités* $\boldsymbol{\beta} = (\beta_1, \beta_2)$ *telles que* $\beta_1 > 1/2$ *and* $\beta_2 > 1/2$. *Alors,*

$$\mathbb{E}\left(\|\alpha - \hat{\alpha}_{\hat{m}}\|_A^2\right) = O(n^{-\frac{2\bar{\beta}}{2\bar{\beta}+2}}).$$

où $2/\bar{\beta} = 1/\beta_1 + 1/\beta_2$.

On montre également que cette **vitesse est optimale au sens minimax**.

Etapes :

- une inégalité de concentration de type Bernstein pour le processus empirique (martingales discontinues)
- une inégalité de type Talagrand pour le processus empirique (chaînage $L^2(\mu) - L^\infty$)

Bibliographie liée à ces outils :

- Barron, A., Birgé, L., and Massart, P. (1999).
- Liptser, R. S. and Shiryayev, A. N. (1989).
- Reynaud-Bouret, P. (2006).
- van de Geer, S. (1995).

Modèle des fonctions de risque proportionnelles (multiplicative hazard model)
On pose :

$$\alpha(x,z) = \exp(\beta x)\alpha_W(z)$$

avec $\beta = 0.4$, X_i i.i.d $\mathcal{U}([0,1])$ $\alpha_W(z) = \alpha\lambda z^{\alpha-1}$ ($\lambda = 1$ et $\alpha = 3$) et $n = 1000$.

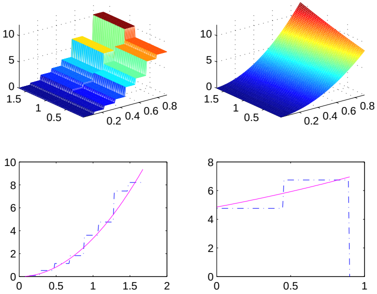


Fig.: Fonctions de risque estimée (haut gauche) et vraie (haut droit) et coupes à x (gauche) ou y (droite) fixés.

1. Découpage de l'échantillon en deux parties : entraînement et apprentissage
2. sur l'**échantillon d'entrainement** : D_m de taille $m < n$
   $\Rightarrow$calcul d'une famille d'estimateurs

   $$\mathcal{A}(\Lambda) = \{\bar{\alpha}_\lambda : \lambda \in \Lambda\}$$

   dits *faibles*, ou *non-adaptatifs*.
3. sur l'**échantillon d'apprentissage** : $D_{(m)}$ de taille $\ell = n - m$
   $\Rightarrow$associer un poids $\theta(\bar{\alpha}) \in [0,1]$ à chaque $\bar{\alpha} \in \mathcal{A}(\Lambda)$, de sorte que $\sum_{\lambda\in\Lambda} \theta(\bar{\alpha}_\lambda) = 1$.

   $$\theta(\bar{\alpha}) \approx \text{ niveau de crédibilité de } \bar{\alpha} \leftrightarrow \text{ risque empirique de } \bar{\alpha}$$

4. L'**estimateur agrégé** (adaptatif) est la combinaison convexe des estimateurs faibles suivante :

   $$\hat{\alpha}_n := \sum_{\lambda\in\Lambda} \theta(\bar{\alpha}_\lambda)\bar{\alpha}_\lambda,$$

Avantages de l'agrégation :

- On peut mélanger des estimateurs purement non-paramétriques et semi paramétriques (SIM, modèle de Cox, etc) : **adaptation en régularité et en structure**

Inconvénient :

- Plus couteux en temps de calcul qu'un algorithme de sélection

- Estimation de l'intensité du processus de comptage sur $\mathbb{R}^+$
- Covariables en temps continu : $X_i(t), t \geq 0$ ou/et en grande dimension