

Estimation adaptative dans le cadre d'une modélisation poissonienne d'une interaction biologique

Soutenance de stage de M2 Probabilités et Statistiques

Laure Sansonnet

Université Paris-Sud 11

16 septembre 2009 - Workshop ATLAS

Plan

- 1 Du cadre biologique à la modélisation mathématique

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation
- 5 Application aux données génomiques

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation
- 5 Application aux données génomiques
- 6 Conclusion et perspectives

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation
- 5 Application aux données génomiques
- 6 Conclusion et perspectives

Le modèle

- Les points dénommés « parents » :

Le modèle

- Les points dénommés « parents » :
processus homogène de Poisson N^0 sur $[0; T]$

Le modèle

- Les points dénommés « parents » :
processus homogène de Poisson N^0 sur $[0; T]$
observations : U_1, \dots, U_n v.a. i.i.d. de loi $\mathcal{U}([0; T])$

Le modèle

- Les points dénommés « parents » :
processus homogène de Poisson N^0 sur $[0; T]$
observations : U_1, \dots, U_n v.a. i.i.d. de loi $\mathcal{U}([0; T])$
- Les points dénommés « enfants » :

Le modèle

- Les points dénommés « parents » :
processus homogène de Poisson N^0 sur $[0; T]$
observations : U_1, \dots, U_n v.a. i.i.d. de loi $\mathcal{U}([0; T])$
- Les points dénommés « enfants » :
 $\forall i \in \{1, \dots, n\}$, processus de Poisson N^i d'intensité $h(t - U_i)$
par rapport à la mesure de Lebesgue sur \mathbb{R}

Le modèle

- Les points dénommés « parents » :
processus homogène de Poisson N^0 sur $[0; T]$
observations : U_1, \dots, U_n v.a. i.i.d. de loi $\mathcal{U}([0; T])$
- Les points dénommés « enfants » :
 $\forall i \in \{1, \dots, n\}$, processus de Poisson N^i d'intensité $h(t - U_i)$
par rapport à la mesure de Lebesgue sur \mathbb{R}
observations : processus agrégé $N = \sum_{i=1}^n N^i$, d'intensité

$$\sum_{i=1}^n h(t - U_i)$$

Le modèle

- Les points dénommés « parents » :
processus homogène de Poisson N^0 sur $[0; T]$
observations : U_1, \dots, U_n v.a. i.i.d. de loi $\mathcal{U}([0; T])$
- Les points dénommés « enfants » :
 $\forall i \in \{1, \dots, n\}$, processus de Poisson N^i d'intensité $h(t - U_i)$
par rapport à la mesure de Lebesgue sur \mathbb{R}
observations : processus agrégé $N = \sum_{i=1}^n N^i$, d'intensité

$$\sum_{i=1}^n h(t - U_i)$$
- But :

Le modèle

- Les points dénommés « parents » :
processus homogène de Poisson N^0 sur $[0; T]$
observations : U_1, \dots, U_n v.a. i.i.d. de loi $\mathcal{U}([0; T])$
- Les points dénommés « enfants » :
 $\forall i \in \{1, \dots, n\}$, processus de Poisson N^i d'intensité $h(t - U_i)$
par rapport à la mesure de Lebesgue sur \mathbb{R}
observations : processus agrégé $N = \sum_{i=1}^n N^i$, d'intensité

$$\sum_{i=1}^n h(t - U_i)$$
- But : estimer h .

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques**
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation
- 5 Application aux données génomiques
- 6 Conclusion et perspectives

La base d'ondelettes de Haar

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$
- $L_2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$
- $L_2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$
 - $V_0 = \{f \in L_2(\mathbb{R}) : f \text{ est constante sur }]k; k+1], k \in \mathbb{Z}\},$

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$
- $L_2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$
 - $V_0 = \{f \in L_2(\mathbb{R}) : f \text{ est constante sur }]k; k+1], k \in \mathbb{Z}\}$,
de base orthonormée $(\psi_{-1k})_{k \in \mathbb{Z}}$:

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$
- $L_2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$
 - $V_0 = \{f \in L_2(\mathbb{R}) : f \text{ est constante sur }]k; k+1], k \in \mathbb{Z}\}$,
de base orthonormée $(\psi_{-1k})_{k \in \mathbb{Z}}$:
 $\forall k \in \mathbb{Z}, \psi_{-1k}(x) = \mathbb{1}_{]k; k+1]}(x)$

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$
- $L_2(\mathbb{R}) = V_0 \oplus \overline{\bigoplus_{j=0}^{\infty} W_j}$
 - $V_0 = \{f \in L_2(\mathbb{R}) : f \text{ est constante sur }]k; k+1], k \in \mathbb{Z}\}$,
de base orthonormée $(\psi_{-1k})_{k \in \mathbb{Z}}$:
$$\forall k \in \mathbb{Z}, \psi_{-1k}(x) = \mathbb{1}_{]k; k+1]}(x)$$
 - $\forall j \geq 0, W_j = \overline{\text{Vect}\{\psi_{jk}, k \in \mathbb{Z}\}}$:

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$

- $L_2(\mathbb{R}) = V_0 \oplus \overline{\bigoplus_{j=0}^{\infty} W_j}$

- $V_0 = \{f \in L_2(\mathbb{R}) : f \text{ est constante sur }]k; k+1], k \in \mathbb{Z}\}$,
de base orthonormée $(\psi_{-1k})_{k \in \mathbb{Z}}$:

$$\forall k \in \mathbb{Z}, \psi_{-1k}(x) = \mathbb{1}_{]k; k+1]}(x)$$

- $\forall j \geq 0, W_j = \overline{\text{Vect}\{\psi_{jk}, k \in \mathbb{Z}\}}$:

$$\forall j \geq 0, \forall k \in \mathbb{Z},$$

$$\psi_{jk}(x) = 2^{j/2} \times \{ \mathbb{1}_{]2k+1; 2k+2]}(x) - \mathbb{1}_{]2k; 2k+1]}(x) \},$$

$$\text{où } J = j + 1$$

La base d'ondelettes de Haar

- ondelette père : $\phi(x) = \mathbb{1}_{]0;1]}(x)$
- ondelette mère : $\psi(x) = \mathbb{1}_{]1/2;1]}(x) - \mathbb{1}_{]0;1/2]}(x)$

- $L_2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$

- $V_0 = \{f \in L_2(\mathbb{R}) : f \text{ est constante sur }]k; k+1], k \in \mathbb{Z}\}$,
de base orthonormée $(\psi_{-1k})_{k \in \mathbb{Z}}$:

$$\forall k \in \mathbb{Z}, \psi_{-1k}(x) = \mathbb{1}_{]k; k+1]}(x)$$

- $\forall j \geq 0, W_j = \overline{\text{Vect}\{\psi_{jk}, k \in \mathbb{Z}\}}$:

$$\forall j \geq 0, \forall k \in \mathbb{Z},$$

$$\psi_{jk}(x) = 2^{j/2} \times \{ \mathbb{1}_{]2k+1; 2k+2]}(x) - \mathbb{1}_{]2k; 2k+1]}(x) \},$$

$$\text{où } J = j + 1$$

- $\forall f \in L_2(\mathbb{R}), f = \sum_{k \in \mathbb{Z}} \beta_{-1k} \psi_{-1k} + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \beta_{jk} \psi_{jk},$

$$\text{où } \forall j \geq -1, \forall k \in \mathbb{Z}, \beta_{jk} = \int f(x) \psi_{jk}(x) dx.$$

Pour la sélection de modèles

Notations

- $\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}$
- $\forall \lambda \in \Lambda, \varphi_\lambda = \psi_{jk}$
- $\forall \lambda \in \Lambda, \beta_\lambda = \beta_{jk}$

Pour la sélection de modèles

Notations

- $\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}$
- $\forall \lambda \in \Lambda, \varphi_\lambda = \psi_{jk}$
- $\forall \lambda \in \Lambda, \beta_\lambda = \beta_{jk}$

$$h = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda, \text{ avec } \beta_\lambda = \int h(x) \varphi_\lambda(x) dx.$$

Pour la sélection de modèles

Notations

- $\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}$
- $\forall \lambda \in \Lambda, \varphi_\lambda = \psi_{jk}$
- $\forall \lambda \in \Lambda, \beta_\lambda = \beta_{jk}$

$$h = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda, \text{ avec } \beta_\lambda = \int h(x) \varphi_\lambda(x) dx.$$

But :

Pour la sélection de modèles

Notations

- $\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}$
- $\forall \lambda \in \Lambda, \varphi_\lambda = \psi_{jk}$
- $\forall \lambda \in \Lambda, \beta_\lambda = \beta_{jk}$

$$h = \sum_{\lambda \in \Lambda} \beta_\lambda \varphi_\lambda, \text{ avec } \beta_\lambda = \int h(x) \varphi_\lambda(x) dx.$$

But : estimer les β_λ .

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ sous-ensemble déterministe connu de Λ

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ sous-ensemble déterministe connu de Λ
- $(\eta_\lambda)_{\lambda \in \Gamma}$

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ *sous-ensemble déterministe connu de Λ*
- $(\eta_\lambda)_{\lambda \in \Gamma}$

Règle de seuillage :

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ sous-ensemble déterministe connu de Λ
- $(\eta_\lambda)_{\lambda \in \Gamma}$

Règle de seuillage : $\tilde{\beta} = (\hat{\beta}_\lambda \mathbb{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbb{1}_{\lambda \in \Gamma})_{\lambda \in \Gamma}$.

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ sous-ensemble déterministe connu de Λ
- $(\eta_\lambda)_{\lambda \in \Gamma}$

Règle de seuillage : $\tilde{\beta} = (\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbf{1}_{\lambda \in \Gamma})_{\lambda \in \Gamma}$.

$\varepsilon > 0$ fixé.

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ sous-ensemble déterministe connu de Λ
- $(\eta_\lambda)_{\lambda \in \Gamma}$

Règle de seuillage : $\tilde{\beta} = (\hat{\beta}_\lambda \mathbb{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbb{1}_{\lambda \in \Gamma})_{\lambda \in \Gamma}$.

$\varepsilon > 0$ fixé. Trouver $(H_\lambda)_{\lambda \in \Gamma}$, $\kappa \in [0; 1[$, $\omega \in [0; 1]$ et $\mu > 0$ vérifiant les propriétés suivantes :

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ sous-ensemble déterministe connu de Λ
- $(\eta_\lambda)_{\lambda \in \Gamma}$

Règle de seuillage : $\tilde{\beta} = (\hat{\beta}_\lambda \mathbb{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbb{1}_{\lambda \in \Gamma})_{\lambda \in \Gamma}$.

$\varepsilon > 0$ fixé. Trouver $(H_\lambda)_{\lambda \in \Gamma}$, $\kappa \in [0; 1[$, $\omega \in [0; 1]$ et $\mu > 0$ vérifiant les propriétés suivantes :

$$(A1) \quad \forall \lambda \in \Gamma, \mathbb{P}(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda) \leq \omega.$$

Résultats théoriques

Théorème (hypothèses)

Pour estimer $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$

Sont données :

- $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, Γ sous-ensemble déterministe connu de Λ
- $(\eta_\lambda)_{\lambda \in \Gamma}$

Règle de seuillage : $\tilde{\beta} = (\hat{\beta}_\lambda \mathbb{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbb{1}_{\lambda \in \Gamma})_{\lambda \in \Gamma}$.

$\varepsilon > 0$ fixé. Trouver $(H_\lambda)_{\lambda \in \Gamma}$, $\kappa \in [0; 1[$, $\omega \in [0; 1]$ et $\mu > 0$ vérifiant les propriétés suivantes :

$$(A1) \quad \forall \lambda \in \Gamma, \mathbb{P}(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda) \leq \omega.$$

$$(A2) \quad \exists 1 < p, q < \infty, \exists R > 0, \forall \lambda \in \Gamma, \\ \left(\mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2p}) \right)^{\frac{1}{p}} \leq R \max \left(H_\lambda, H_\lambda^{\frac{1}{p}} \varepsilon^{1/q} \right).$$

$$(A3) \quad \exists \theta, \forall \lambda \in \Gamma, H_\lambda < \theta \varepsilon, \\ \mathbb{P}(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda, |\hat{\beta}_\lambda| > \eta_\lambda) \leq H_\lambda \mu.$$

Résultats théoriques

Théorème (conclusion)

Alors, l'estimateur $\tilde{\beta}$ satisfait

Résultats théoriques

Théorème (conclusion)

Alors, l'estimateur $\tilde{\beta}$ satisfait

$$\mathbb{E}(\|\tilde{\beta} - \beta\|_{\ell_2}^2) \leq C \times \mathbb{E} \left(\inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + \sum_{\lambda \in m} (\hat{\beta}_\lambda - \beta_\lambda)^2 + \sum_{\lambda \in m} \eta_\lambda^2 \right\} \right) + LD \sum_{\lambda \in \Gamma} H_\lambda.$$

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle**
- 4 Procédure d'implémentation
- 5 Application aux données génomiques
- 6 Conclusion et perspectives

L'estimateur

- $\hat{\beta}_{jk} = \frac{\Gamma(\psi_{jk})}{n},$

où $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right] dN_t$

L'estimateur

- $\hat{\beta}_{jk} = \frac{\Gamma(\psi_{jk})}{n},$

où $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right] dN_t$

avec $\mathbb{E}_u(\psi_{jk}(t - u)) = \frac{1}{T} \int_0^T \psi_{jk}(t - u) du$

L'estimateur

- $\hat{\beta}_{jk} = \frac{\Gamma(\psi_{jk})}{n},$

où $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right] dN_t$

avec $\mathbb{E}_u(\psi_{jk}(t - u)) = \frac{1}{T} \int_0^T \psi_{jk}(t - u) du$

- Calcul de l'espérance

L'estimateur

- $\hat{\beta}_{jk} = \frac{\Gamma(\psi_{jk})}{n},$

où $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right] dN_t$

avec $\mathbb{E}_u(\psi_{jk}(t - u)) = \frac{1}{T} \int_0^T \psi_{jk}(t - u) du$

- Calcul de l'espérance
estimateur sans biais :

L'estimateur

- $\hat{\beta}_{jk} = \frac{\Gamma(\psi_{jk})}{n},$

où $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right] dN_t$

avec $\mathbb{E}_u(\psi_{jk}(t - u)) = \frac{1}{T} \int_0^T \psi_{jk}(t - u) du$

- Calcul de l'espérance

estimateur sans biais : $\mathbb{E}(\hat{\beta}_{jk}) = \int \psi_{jk} h = \beta_{jk}$

L'estimateur

- $\hat{\beta}_{jk} = \frac{\Gamma(\psi_{jk})}{n},$

où $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right] dN_t$

avec $\mathbb{E}_u(\psi_{jk}(t - u)) = \frac{1}{T} \int_0^T \psi_{jk}(t - u) du$

- Calcul de l'espérance

estimateur sans biais : $\mathbb{E}(\hat{\beta}_{jk}) = \int \psi_{jk} h = \beta_{jk}$

- Calcul de la variance

L'estimateur

- $\hat{\beta}_{jk} = \frac{\Gamma(\psi_{jk})}{n},$

où $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right] dN_t$

avec $\mathbb{E}_u(\psi_{jk}(t - u)) = \frac{1}{T} \int_0^T \psi_{jk}(t - u) du$

- Calcul de l'espérance

estimateur sans biais : $\mathbb{E}(\hat{\beta}_{jk}) = \int \psi_{jk} h = \beta_{jk}$

- Calcul de la variance

$$\mathbb{E}(V(\psi_{jk})) \leq n \int \psi_{jk}^2 h + \frac{n(n-1)}{T} \int h,$$

où $V(\psi_{jk}) = \text{Var}(\Gamma(\psi_{jk}) | U_1, \dots, U_n)$

À calculer

- $$\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$$

À calculer

- $\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$
- $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \lambda(\psi_{jk})(t) dN_t$

À calculer

- $\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$
- $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \lambda(\psi_{jk})(t) dN_t$
- $\hat{V}(\psi_{jk}) = \int_{\mathbb{R}} (\lambda(\psi_{jk})(t))^2 dN_t$

À calculer

- $\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$
- $\Gamma(\psi_{jk}) = \int_{\mathbb{R}} \lambda(\psi_{jk})(t) dN_t$
- $\hat{V}(\psi_{jk}) = \int_{\mathbb{R}} (\lambda(\psi_{jk})(t))^2 dN_t$
- $B(\psi_{jk}) = \|\lambda(\psi_{jk})\|_{\infty}$

Le seuil

$$\mathbb{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > \sqrt{2\alpha \tilde{V}(\psi_{jk}/n)} + \frac{B(\psi_{jk}/n)\alpha}{3} + \eta \int h \right)$$

Le seuil

$$\begin{aligned} \mathbb{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > \sqrt{2\alpha \tilde{V}(\psi_{jk}/n)} + \frac{B(\psi_{jk}/n)\alpha}{3} + \eta \int h \right) \\ \leq (5 + 2 \times 2.77)e^{-\alpha} = \omega \end{aligned}$$

Le seuil

$$\begin{aligned} \mathbb{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > \sqrt{2\alpha \tilde{V}(\psi_{jk}/n)} + \frac{B(\psi_{jk}/n)\alpha}{3} + \eta \int h \right) \\ \leq (5 + 2 \times 2.77)e^{-\alpha} = \omega \end{aligned}$$

- $\alpha = \gamma \ln(T)$

Le seuil

$$\mathbb{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > \sqrt{2\alpha \tilde{V}(\psi_{jk}/n)} + \frac{B(\psi_{jk}/n)\alpha}{3} + \eta \int h \right) \\ \leq (5 + 2 \times 2.77)e^{-\alpha} = \omega$$

- $\alpha = \gamma \ln(T)$
- $\eta = \frac{2\alpha}{3T} 2^{-j/2} + 2(1 + \varepsilon)^{3/2} \sqrt{\frac{2}{T} + \frac{2}{T^2} 2^{-j}} \sqrt{\alpha} + 4\eta(\varepsilon) \sqrt{\frac{1}{T} \alpha} + \beta(\varepsilon) \sqrt{\frac{40}{nT}} \alpha^{3/2} + \gamma(\varepsilon) \frac{8}{n} 2^{j/2} \alpha^2$
- $\eta = \eta_{bis} / \sqrt{T}$

Le seuil

$$\mathbb{P} \left(|\hat{\beta}_{jk} - \beta_{jk}| > \sqrt{2\alpha \tilde{V}(\psi_{jk}/n)} + \frac{B(\psi_{jk}/n)\alpha}{3} + \eta \int h \right) \\ \leq (5 + 2 \times 2.77)e^{-\alpha} = \omega$$

- $\alpha = \gamma \ln(T)$
- $\eta = \frac{2\alpha}{3T} 2^{-j/2} + 2(1 + \varepsilon)^{3/2} \sqrt{\frac{2}{T} + \frac{2}{T^2} 2^{-j}} \sqrt{\alpha} + 4\eta(\varepsilon) \sqrt{\frac{1}{T} \alpha} + \beta(\varepsilon) \sqrt{\frac{40}{nT}} \alpha^{3/2} + \gamma(\varepsilon) \frac{8}{n} 2^{j/2} \alpha^2$
- $\eta = \eta_{bis} / \sqrt{T}$
- $\int h \rightarrow N/n$

Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation**
- 5 Application aux données génomiques
- 6 Conclusion et perspectives

Algorithme

- But :

Algorithme

- But : calculer

$$\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$$

Algorithme

- But : calculer

$$\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$$

- partie « constante par morceaux » : $\sum_{i=1}^n \psi_{jk}(t - U_i)$

Algorithme

- But : calculer

$$\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$$

- partie « constante par morceaux » : $\sum_{i=1}^n \psi_{jk}(t - U_i)$

on la calcule au niveau J_{max}

→ algorithme de descente

Algorithme

- But : calculer

$$\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$$

- partie « constante par morceaux » : $\sum_{i=1}^n \psi_{jk}(t - U_i)$

on la calcule au niveau J_{max}

→ algorithme de descente

- partie déterministe (affine par morceaux) :
 $(n-1) \mathbb{E}_u(\psi_{jk}(t - u))$

Algorithme

- But : calculer

$$\lambda(\psi_{jk})(t) = \sum_{i=1}^n \left[\psi_{jk}(t - U_i) - \frac{n-1}{n} \mathbb{E}_u(\psi_{jk}(t - u)) \right]$$

- partie « constante par morceaux » : $\sum_{i=1}^n \psi_{jk}(t - U_i)$

on la calcule au niveau J_{max}

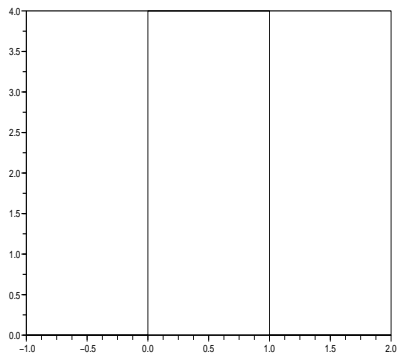
→ algorithme de descente

- partie déterministe (affine par morceaux) : $(n-1)\mathbb{E}_u(\psi_{jk}(t - u))$
- calcul du seuil

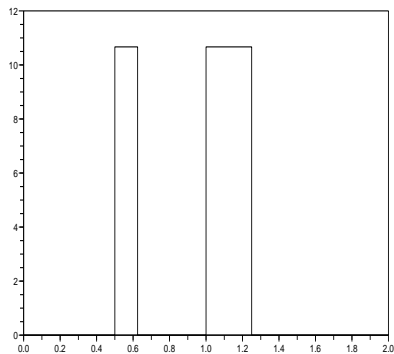
Deux fonctions tests

Deux fonctions tests

Uniform



FctTest



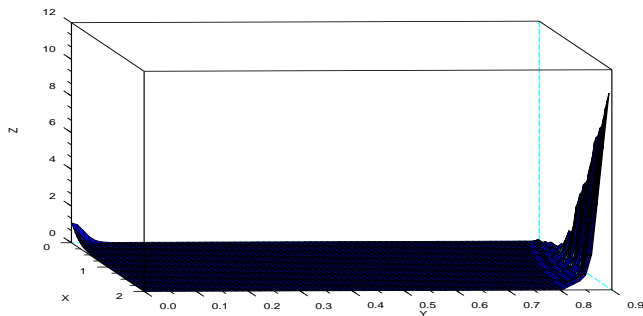
Grille en γ et η

Pour Uniform

Grille en γ et η

Pour Uniform

- $T = 2000$; $Jmax = 7$; $A = 1$
- axe des $X \rightarrow \gamma$, allant de 0 à 2
- axe des $Y \rightarrow \eta$, avec $\eta = \eta_{bis}/\sqrt{T}$, η_{bis} allant de 0 à 40
- axe des $Z \rightarrow$ le risque



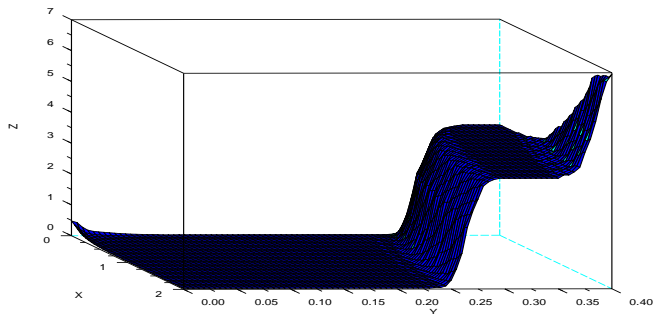
Grille en γ et η

Pour FctTest

Grille en γ et η

Pour FctTest

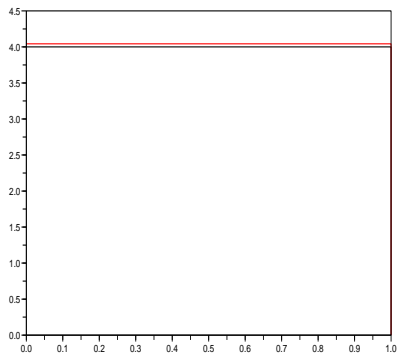
- $T = 10000$; $Jmax = 5$; $A = 10$
- axe des $X \rightarrow \gamma$, allant de 0 à 2
- axe des $Y \rightarrow \eta$, avec $\eta = \eta_{bis}/\sqrt{T}$, η_{bis} allant de 0 à 40
- axe des $Z \rightarrow$ le risque



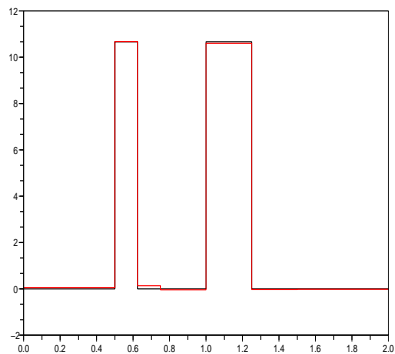
Les reconstructions

Les reconstructions

Uniform

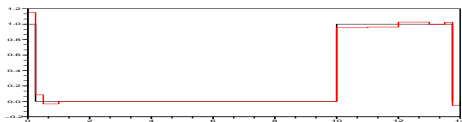
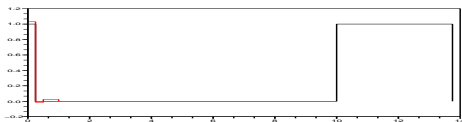
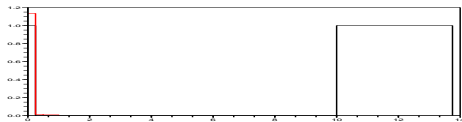


FctTest



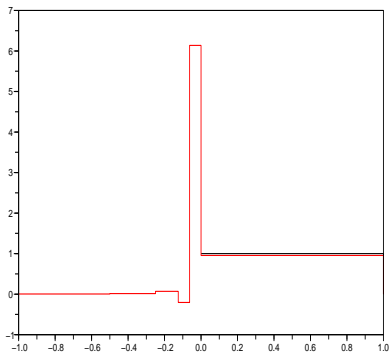
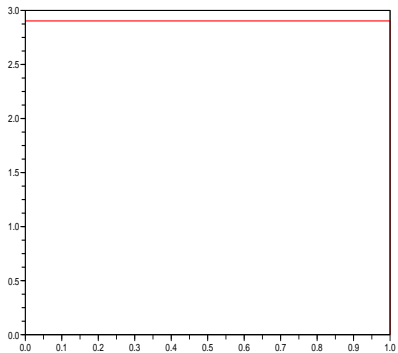
Robustesse de la procédure

- si on se trompe de A



Robustesse de la procédure

- s'il y a de l'apparition spontanée



Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation
- 5 Application aux données génomiques**
- 6 Conclusion et perspectives

E. coli : influence sites promoteurs/gènes

- données : bactérie *E. coli* (*Escherichia coli*)
- parents = tataat ; enfants = gènes

E. coli : influence sites promoteurs/gènes

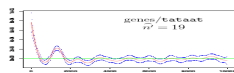
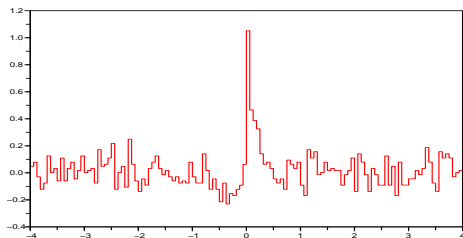
- données : bactérie *E. coli* (*Escherichia coli*)
- parents = tataat ; enfants = gènes
- 4290 gènes et 1036 occurrences de tataat

E. coli : influence sites promoteurs/gènes

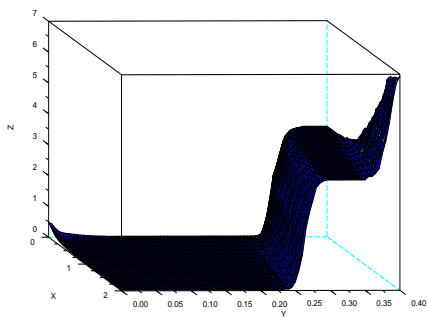
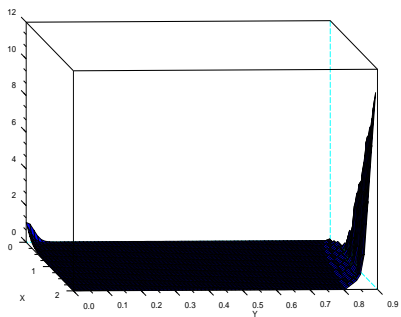
- données : bactérie *E. coli* (*Escherichia coli*)
- parents = tataat ; enfants = gènes
- 4290 gènes et 1036 occurrences de tataat
- $T = 9282(\times 1000)$; $J_{max} = 4$; $A = 4$

E. coli : influence sites promoteurs/gènes

- données : bactérie *E. coli* (*Escherichia coli*)
- parents = tataat ; enfants = gènes
- 4290 gènes et 1036 occurrences de tataat
- $T = 9282(\times 1000)$; $J_{max} = 4$; $A = 4$



Grilles en γ et η



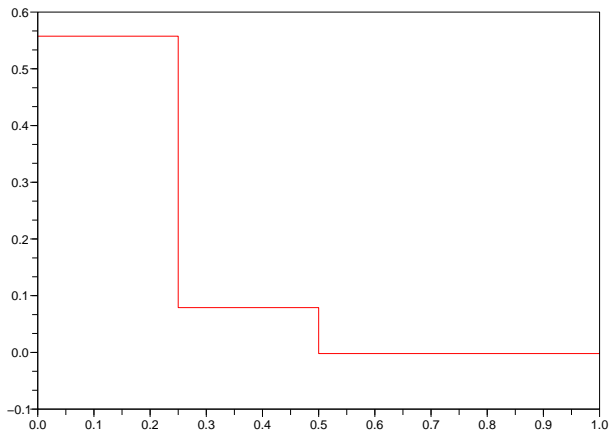
E. coli

E. coli

- $\gamma = 0.05$
- $\eta \simeq 0.03$, i.e. $\eta_{bis} = 3$

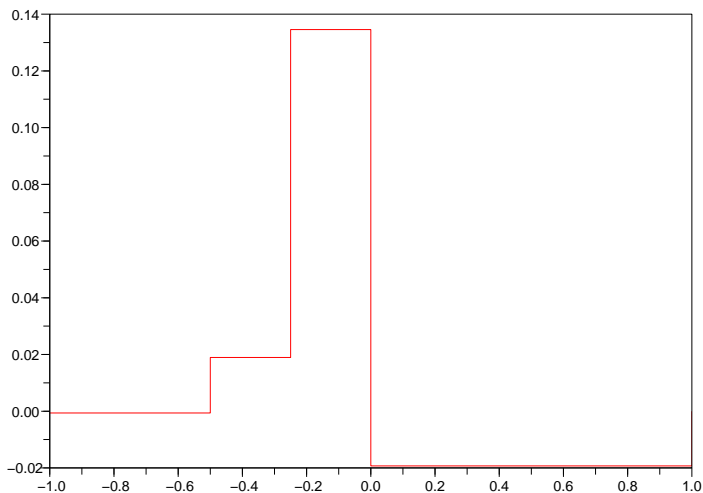
E. coli

- $\gamma = 0.05$
- $\eta \simeq 0.03$, i.e. $\eta_{bis} = 3$



Inversion parents/enfants

Inversion parents/enfants



Plan

- 1 Du cadre biologique à la modélisation mathématique
- 2 Les outils de prédilection et résultats théoriques
- 3 Estimation et seuillage pour notre modèle
- 4 Procédure d'implémentation
- 5 Application aux données génomiques
- 6 Conclusion et perspectives**

Perspectives

- Hypothèses (A2) et (A3) du théorème général à vérifier dans le cadre de notre modèle → début de thèse

Perspectives

- Hypothèses (A2) et (A3) du théorème général à vérifier dans le cadre de notre modèle → début de thèse
- Amélioration du modèle : prendre en compte une apparition spontanée d'enfants, selon un processus de Poisson

Perspectives

- Hypothèses (A2) et (A3) du théorème général à vérifier dans le cadre de notre modèle → début de thèse
- Amélioration du modèle : prendre en compte une apparition spontanée d'enfants, selon un processus de Poisson
- Méconnaissance a priori du support de la fonction h ... : estimation possible sans connaître A ?