Estimation adaptative de la fonction de répartition conditionnellement à une covariable fonctionnelle

Gaëlle Chagny [1] Angelina Roche [2]

[1]LMRS – Université de Rouen

[2]I3M – Université Montpellier II

Onzième Colloque Jeunes Probabilistes et Statisticiens,
Forges-les-Eaux, 6 au 11 avril 2014.

Cadre : statistique pour données fonctionnelles

- *Données fonctionnelles :*

  Les données sont des réalisations d'une fonction aléatoire X.

  $\longrightarrow$ cadre différent du cadre classique de la statistique qui consiste à étudier des vecteurs de $\mathbb{R}^d$.

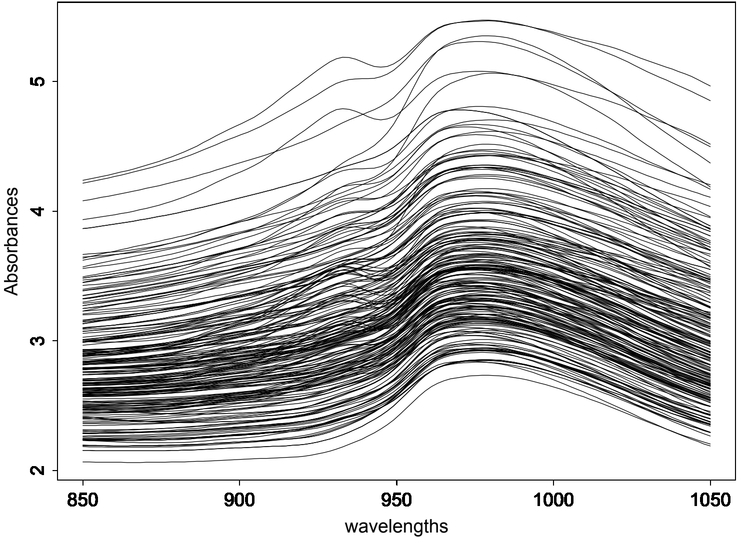
  - Hypothèse: $X \in \mathbb{H}$, où $(\mathbb{H}, \|\cdot\|, \langle\cdot,\cdot\rangle)$ espace de Hilbert séparable.
  - Exemple: $X : [a, b] \to \mathbb{R}$ et $\mathbb{H} = \mathbb{L}^2([a, b])$ avec

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt \text{ et } \|f\| = \sqrt{\langle f, f\rangle}.$$

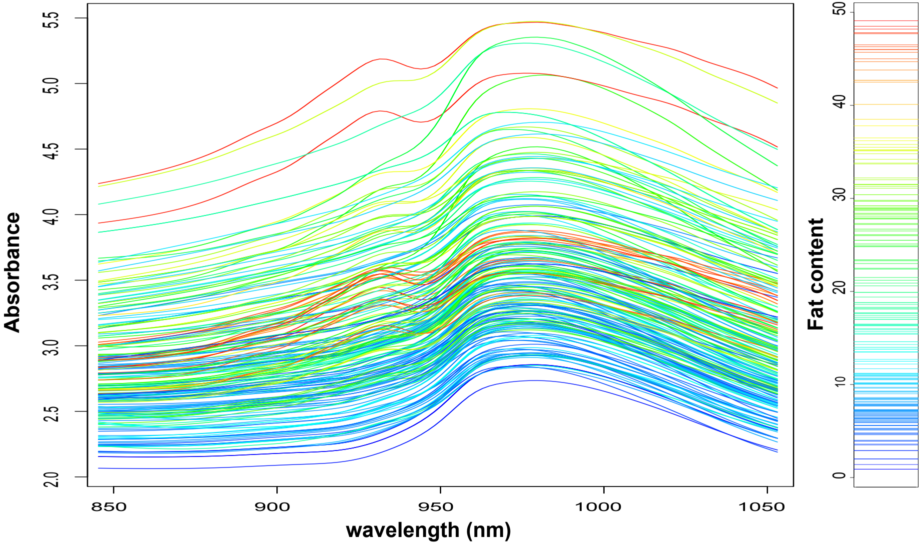
- *Contexte :*

  Étude du lien entre $Y \in \mathbb{R}$ et $X \in \mathbb{H}$ à l'aide d'un échantillon $\{(X_i, Y_i), i = 1, \ldots, n\}$ *i.i.d.* de copies de (X, Y).

Exemples de données fonctionnelles



Exemples de données fonctionnelles



Plan

1. Estimateurs à noyaux de la fonction de répartition conditionnelle
2. Sélection de la fenêtre
3. Simulations

Plan

Objectif

Objectif

Estimer la fonction de répartition conditionnelle :

$$F^X : y \mapsto \mathbb{P}(Y \leq y|X),$$

où X est une v.a. à valeur dans $\mathbb{H}$ et Y v.a.r.

Pas d'hypothèse sur la forme de la relation de dépendance entre X et Y.
$\longrightarrow$ *cadre non-paramétrique*.

Travaux existant: estimation de $F^X : y \mapsto \mathbb{P}(Y \leq y|X)$

Cas particulier: Si X et Y indépendants, $F^X(y) = \mathbb{P}(Y \leq y)$.

$$\text{Estimateur : } \widehat{F}(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq y\}}.$$

Cas général : (Ferraty *et al.*, 2006,2010) Estimateur à noyau:

$\hat{F}_h^x(y) := \sum_{i=1}^{n} W_h^{(i)} \mathbf{1}_{\{Y_i \leq y\}}$ avec $W_h^{(i)} = \frac{K(\|X_i - x\|/h)}{\sum_{i=1}^{n} K(\|X_i - x\|/h)}$ où K est un noyau

i.e. $K : \mathbb{R} \to \mathbb{R}_+$, $\int_{\mathbb{R}} K(t)dt = 1$.

$\longrightarrow$ *Estimateur convergent lorsque* h *est bien choisi.*

Travaux existant: estimation de $F^X : y \mapsto \mathbb{P}(Y \leq y|X)$

Cas particulier: Si X et Y indépendants, $F^X(y) = \mathbb{P}(Y \leq y)$.

$$\text{Estimateur : } \widehat{F}(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq y\}}.$$

Cas général : (Ferraty *et al.*, 2006,2010) Estimateur à noyau:

$\hat{F}_h^x(y) := \sum_{i=1}^{n} W_h^{(i)} \mathbf{1}_{\{Y_i \leq y\}}$ avec $W_h^{(i)} = \dfrac{K(\|X_i - x\|/h)}{\sum_{i=1}^{n} K(\|X_i - x\|/h)}$ où K est un noyau

i.e. $K : \mathbb{R} \to \mathbb{R}_+$, $\int_{\mathbb{R}} K(t)dt = 1$.

$\longrightarrow$ *Estimateur convergent lorsque* h *est bien choisi.*

Travaux existant: estimation de $F^X : y \mapsto \mathbb{P}(Y \leq y | X)$

Cas particulier: Si X et Y indépendants, $F^X(y) = \mathbb{P}(Y \leq y)$.

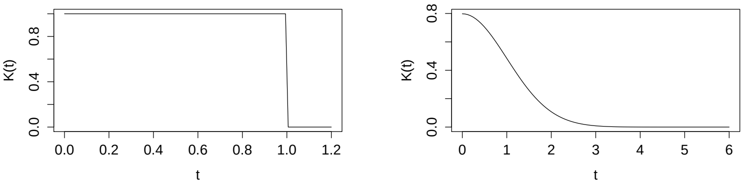
$$\text{Estimateur : } \widehat{F}(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq y\}}.$$

Cas général : (Ferraty *et al.*, 2006,2010) Estimateur à noyau:

$\hat{F}_h^x(y) := \sum_{i=1}^{n} W_h^{(i)} \mathbf{1}_{\{Y_i \leq y\}}$ avec $W_h^{(i)} = \frac{K(\|X_i - x\|/h)}{\sum_{i=1}^{n} K(\|X_i - x\|/h)}$ où K est un noyau

i.e. $K : \mathbb{R} \to \mathbb{R}_+$, $\int_{\mathbb{R}} K(t)dt = 1$.

$\longrightarrow$ *Estimateur convergent lorsque* h *est bien choisi.*



Travaux existant: estimation de $F^X : y \mapsto \mathbb{P}(Y \leq y|X)$

Cas particulier: Si X et Y indépendants, $F^X(y) = \mathbb{P}(Y \leq y)$.

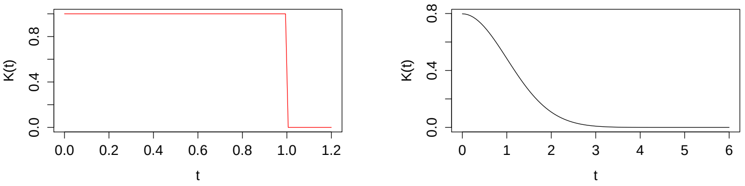
$$\text{Estimateur : } \widehat{F}(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq y\}}.$$

Cas général : (Ferraty *et al.*, 2006,2010) Estimateur à noyau:

$\hat{F}_h^x(y) := \sum_{i=1}^{n} W_h^{(i)} \mathbf{1}_{\{Y_i \leq y\}}$ avec $W_h^{(i)} = \frac{K(\|X_i - x\|/h)}{\sum_{i=1}^{n} K(\|X_i - x\|/h)}$ où K est un noyau

i.e. $K : \mathbb{R} \to \mathbb{R}_+$, $\int_{\mathbb{R}} K(t)dt = 1$.

$\longrightarrow$ *Estimateur convergent lorsque* h *est bien choisi.*



Choix de la fenêtre

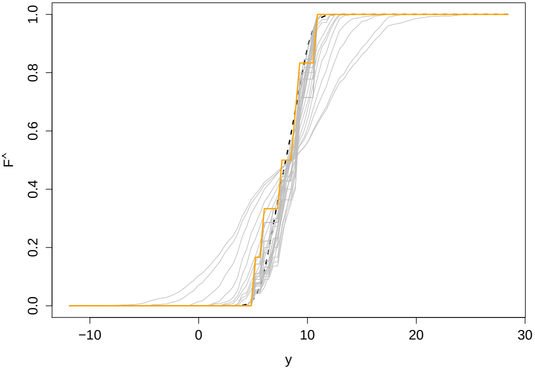


Figure: En gris, estimateurs de la famille $\widehat{F}_h$, $h \in \mathcal{H}_n$ (où $\mathcal{H}_n$ est notre collection de fenêtres), en orange $\widehat{F}_h$ avec $h \approx 0.18$.

Choix de la fenêtre

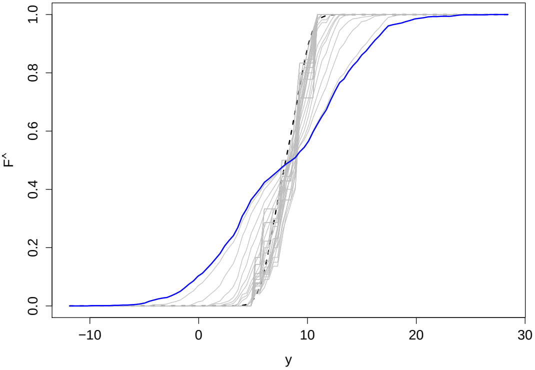


Figure: En gris, estimateurs de la famille $\widehat{F}_h$, $h \in \mathcal{H}_n$ (où $\mathcal{H}_n$ est notre collection de fenêtres), en bleu $\widehat{F}_h$ avec $h \approx 5.19$.

Choix de la fenêtre

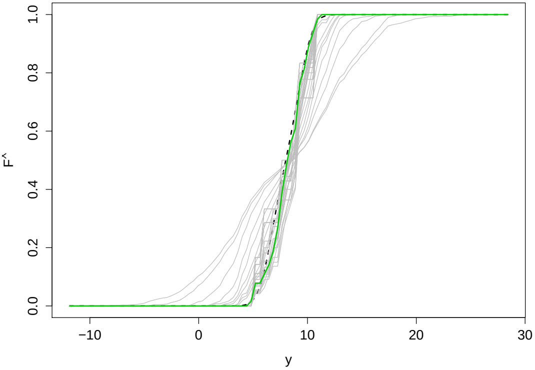


Figure: En gris, estimateurs de la famille $\widehat{F}_h$, $h \in \mathcal{H}_n$ (où $\mathcal{H}_n$ est notre collection de fenêtres), en vert $\widehat{F}_h$ avec $h \approx 0.16$.

Choix de la fenêtre

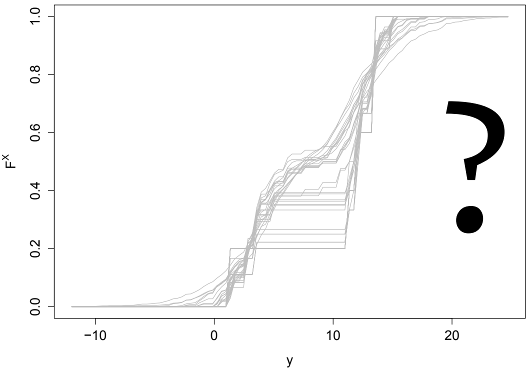


Figure: En gris, estimateurs de la famille $\widehat{F}_h$, $h \in \mathcal{H}_n$ (où $\mathcal{H}_n$ est notre collection de fenêtres).

Questions :

- Comment choisir h ?
- Propriétés de $\widehat{F}_h^x$ lorsque n fini ?

Étude des propriétés de $\hat{F}_h^x(y)$: risque considéré

Risque quadratique intégré :

$$\mathbb{E}\left[\left\|F^{X'} - \hat{F}_h^{X'}\right\|_D^2 \mathbf{1}_{\{X' \in B\}}\right] = \int_B \int_D \left(F^x(y) - \hat{F}_h^x(y)\right)^2 dP_X(x)dy,$$

où

- X' est une copie indépendante de X;
- D est un compact de $\mathbb{R}$;
- B est un sous-ensemble borné de $\mathbb{H}$.

Étude des propriétés de $\hat{F}_h^x(y)$: hypothèses

Hypothèses...

H_K **... sur le noyau**

- K est à support dans $[0,1]$,
- $\forall t \in [0,1], \;\; 0 < c_K \leq K(t) \leq C_K < +\infty;$

H_F **... sur la répartition conditionnelle**

- L'application $x \mapsto F^x$ est β-höldérienne :

$$\exists C_D > 0, \;\; \forall x, x' \in \mathbb{H}, \;\; \|F^x - F^{x'}\|_D \leq C_D \|x - x'\|^{\beta};$$

H_φ **... sur le processus X**

- via les probabilités de petites boules:

$$\varphi^{X'}(h) := \mathbb{P}(\|X - X'\| \leq h | X') \text{ et } \varphi(h) := \mathbb{P}(\|X\| \leq h).$$

- $\exists c_\varphi, C_\varphi > 0$, telles que :

$$\forall h > 0, \;\; c_\varphi \varphi(h) \leq \varphi^{X'}(h) \leq C_\varphi \varphi(h), \text{ p.s. sur } \{X' \in B\}.$$

Étude des propriétés de $\hat{F}_h^x(y)$: majoration à h fixé

Proposition (Chagny et Roche, 2014)

Sous les hypothèses H_K, H_F et H_φ,

$$\mathbb{E}\left[\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right] \leq C\left(h^{2\beta} + \frac{1}{n\varphi(h)}\right),$$

où $C > 0$ dépend uniquement de c_K, C_K, c_φ, C_φ, $|D|$ et C_D.

Problème:

h optimal dépend de β inconnu $\longrightarrow$ comment choisir h en pratique ?

Étude des propriétés de $\hat{F}_h^x(y)$ *: majoration à* h *fixé*

Proposition (Chagny et Roche, 2014)

Sous les hypothèses H_K, H_F et H_φ,

$$\mathbb{E}\left[\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right] \leq C\left(h^{2\beta} + \frac{1}{n\varphi(h)}\right),$$

où $C > 0$ dépend uniquement de c_K, C_K, c_φ, C_φ, $|D|$ et C_D.

Problème:

h optimal dépend de β inconnu $\longrightarrow$ comment choisir h en pratique ?

Plan

1. *Estimateurs à noyaux de la fonction de répartition conditionnelle*

2. *Sélection de la fenêtre*
   - Methode inspirée de Goldenshluger-Lepski
   - Majoration du risque de l'estimateur adaptatif
   - Vitesses de convergence : optimalité au sens minimax

3. *Simulations*

Méthode "type Goldenshluger-Lepski"

Objectif:

Sélectionner un estimateur ayant des propriétés comparables à celles de l'*oracle* $\widehat{F}_{h^*}$ où

$$h^* = \arg\min_{h>0} \mathbb{E}\left[\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right]$$

Critère imitant la décomposition biais-variance du risque:

$$\hat{h} = \arg\min_{h\in\mathcal{H}_n}\left\{\widehat{A}(h) + \widehat{V}(h)\right\}, \mathcal{H}_n \subset \mathbb{R}_+^* \text{ collection finie,}$$

où:

- $\widehat{V}(h) = \kappa \dfrac{\ln n}{n\widehat{\varphi}(h)}$ où $\kappa > 0$ et $\widehat{\varphi}(h) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{\|X_i\|\leq h\}}$.
- $\widehat{A}(h) = \max_{h'\in\mathcal{H}_n}\left(\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h'\vee h}^{X'}\|_D^2 - \widehat{V}(h')\right)_+$.

Méthode "type Goldenshluger-Lepski"

Objectif:

Sélectionner un estimateur ayant des propriétés comparables à celles de l'*oracle* $\widehat{F}_{h^*}$ où

$$h^* = \arg\min_{h>0} \underbrace{\mathbb{E}\left[\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right]}_{\leq C\left(h^{2\beta} + \frac{1}{n\varphi(h)}\right)}$$

Critère imitant la décomposition biais-variance du risque:

$$\hat{h} = \arg\min_{h\in\mathcal{H}_n}\left\{\widehat{A}(h) + \widehat{V}(h)\right\}, \mathcal{H}_n \subset \mathbb{R}_+^* \text{ collection finie,}$$

où:

- $\widehat{V}(h) = \kappa\dfrac{\ln n}{n\widehat{\varphi}(h)}$ où $\kappa > 0$ et $\widehat{\varphi}(h) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{\|X_i\|\leq h\}}$.
- $\widehat{A}(h) = \max_{h'\in\mathcal{H}_n}\left(\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h'\vee h}^{X'}\|_D^2 - \widehat{V}(h')\right)_+$.

Méthode "type Goldenshluger-Lepski"

Objectif:

Sélectionner un estimateur ayant des propriétés comparables à celles de l'*oracle* $\widehat{F}_{h^*}$ où

$$h^* = \arg\min_{h>0} \quad \underbrace{\mathbb{E}\left[\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right]}_{\leq C\left(h^{2\beta} + \frac{1}{n\varphi(h)}\right)}$$

Critère imitant la décomposition biais-variance du risque:

$$\hat{h} = \arg\min_{h \in \mathcal{H}_n} \left\{\widehat{A}(h) + \widehat{V}(h)\right\}, \mathcal{H}_n \subset \mathbb{R}_+^* \text{ collection finie,}$$

où:

- $\widehat{V}(h) = \kappa \dfrac{\ln n}{n\widehat{\varphi}(h)}$ où $\kappa > 0$ et $\widehat{\varphi}(h) := \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{\|X_i\| \leq h\}}$.
- $\widehat{A}(h) = \max_{h' \in \mathcal{H}_n} \left(\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h' \vee h}^{X'}\|_D^2 - \widehat{V}(h')\right)_+$.

Majoration du risque de l'estimateur adaptatif

Majoration du risque adaptatif (Chagny et Roche, 2014)

Sous des conditions portant sur la collection $\mathcal{H}_n$ et sur la constante κ, si les hypothèses H_K, H_F et H_φ sont vérifiées :

$$\mathbb{E}\left[\left\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right] \leq C'\left(h^{2\beta} + \frac{\ln n}{n\varphi(h)}\right) + \frac{C}{n}.$$

$\longrightarrow$ Estimateur optimal au sens de l'oracle, à la perte log près.

Vitesses de convergence ?

Majoration du risque de l'estimateur adaptatif

Majoration du risque adaptatif (Chagny et Roche, 2014)

Sous des conditions portant sur la collection $\mathcal{H}_n$ et sur la constante κ, si les hypothèses H_K, H_F et H_φ sont vérifiées :

$$\mathbb{E}\left[\left\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right] \leq C'\left(h^{2\beta} + \frac{\ln n}{n\varphi(h)}\right) + \frac{C}{n}.$$

$\longrightarrow$ Estimateur optimal au sens de l'oracle, à la perte log près.

Vitesses de convergence ?

Hypothèses sur la probabilité de petite boule φ

- **Rappel:** Concentration du processus X en l'origine

$$\varphi(h) = \mathbb{P}\left(\|X\| \leq h\right), \; h > 0.$$

- **Hypothèses**

$H_{X,L}$ Il existe $\gamma_1, \gamma_2 \in \mathbb{R}$ et $\alpha > 0$ tels que

$$c_1 h^{\gamma_1} \exp(-c_2 h^{-\alpha}) \leq \varphi(h) \leq C_1 h^{\gamma_2} \exp(-c_2 h^{-\alpha}),$$

$H_{X,M}$ Il existe $\gamma_1, \gamma_2 \in \mathbb{R}$ et $\alpha > 1$ tels que

$$c_1 h^{\gamma_1} \exp(-c_2 \ln^{\alpha}(1/h)) \leq \varphi(h) \leq C_1 h^{\gamma_2} \exp(-c_2 \ln^{\alpha}(1/h)),$$

$H_{X,F}$ Il existe $\gamma > 0$ tel que

$$c_1 h^{\gamma} \leq \varphi(h) \leq C_1 h^{\gamma}.$$

- X mouvement brownien vérifie $H_{X,L}$ avec $\alpha = 2$;
- $X \in \mathbb{R}^d$ vecteur aléatoire vérifie $H_{X,F}$ avec $\gamma = d$.

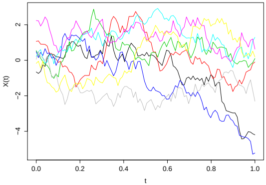
Vitesses

	$H_{X,L}$ (vitesse lente)	$H_{X,M}$ (vitesse intermédiaire)	$H_{X,F}$ (vitesse rapide)
(a) Vitesses pour $\widehat{F}_{h^*}$ (bornes sup.)	$(\ln(n))^{-2\beta/\alpha}$	$\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha}(n)\right)$	$n^{-\frac{2\beta}{2\beta+\gamma}}$
(b) Vitesses pour $\widehat{F}_{\widehat{h}}$ (bornes sup.)	$(\ln(n))^{-2\beta/\alpha}$	$\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha}(n)\right)$	$\left(\frac{n}{\ln(n)}\right)^{-\frac{2\beta}{2\beta+\gamma}}$
(c) Risque minimax (bornes inf.)	$(\ln(n))^{-2\beta/\alpha}$	$\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha}(n)\right)$	$n^{-\frac{2\beta}{2\beta+\gamma}}$

Plan

Simulation de X

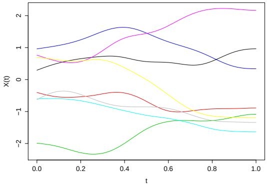
$H_{X,L}$



$$X(t) = W(t) + \xi_0$$

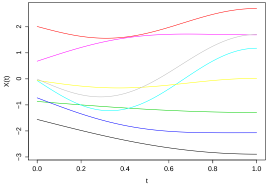
$(W(t)$ mouvement brownien$)$

$H_{X,M}$



$$X(t) = \xi_0 + \sqrt{2}\sum_{j=1}^{150} \frac{e^{-j}}{\sqrt{j}} \sin(\pi(j-0.5)t)$$

$H_{X,F}$



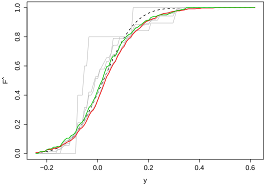
$$X(t) = \xi_0 + \sqrt{2}\xi_1 \sin(-\pi t/2) + \xi_2 \sin(\pi t/2)/\sqrt{2}$$

avec $(\xi_j)_{j\geq 0}$ i.i.d. $\mathcal{N}(0,1)$.

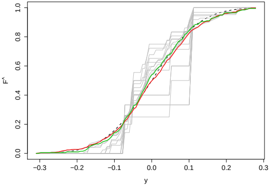
Résultats : modèle de régression

$Y_i = \left(\int_0^1 \beta(t) X_i(t) dt\right)^2 + \varepsilon_i$ ($i = 1, \ldots, 500$) avec $\beta(t) = \sin(4\pi t)$ et $\varepsilon_i \sim \mathcal{N}(0, 0.1)$.

$H_{X,L}$



$H_{X,M}$



$H_{X,F}$

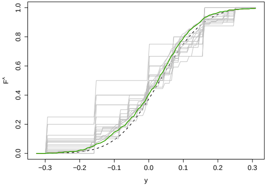
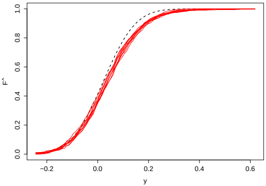


Figure: Légende : en gris, estimateurs de la famille $\widehat{F}_h$, $h \in \mathcal{H}_n$, en vert, meilleur estimateur, en rouge, estimateur sélectionné $\widehat{F}_{\widehat{h}}$.

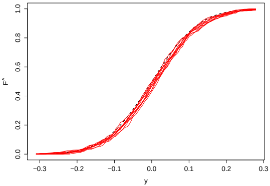
Résultats : modèle de régression

$Y_i = \left(\int_0^1 \beta(t) X_i(t) dt \right)^2 + \varepsilon_i$ ($i = 1, \ldots, 500$) avec $\beta(t) = \sin(4\pi t)$ et $\varepsilon_i \sim \mathcal{N}(0, 0.1)$.

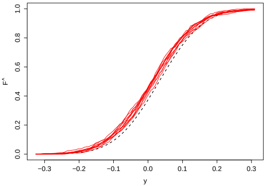
$H_{X,L}$



$H_{X,M}$



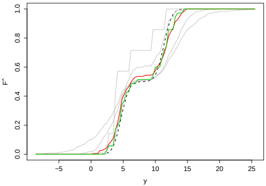
$H_{X,F}$



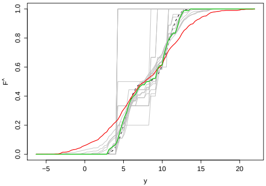
Résultats : modèle de mélange gaussien

$$Y_i | X_i = x \sim 0.5\mathcal{N}(8 - 4\|x\|, 1) + 0.5\mathcal{N}(8 + 4\|x\|, 1),\ i = 1, \ldots, 500.$$

$H_{X,L}$



$H_{X,M}$



$H_{X,F}$

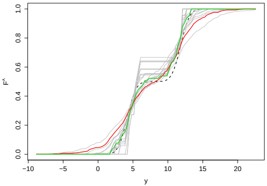
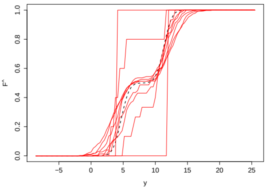


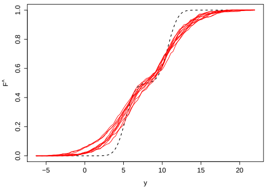
Figure: Légende : en gris, estimateurs de la famille $\widehat{F}_h$, $h \in \mathcal{H}_n$, en vert, meilleur estimateur, en rouge, estimateur sélectionné $\widehat{F}_{\widehat{h}}$.

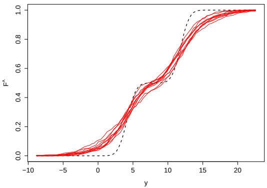
Résultats : modèle de mélange gaussien

$$Y_i | X_i = x \sim 0.5\mathcal{N}(8 - 4\|x\|, 1) + 0.5\mathcal{N}(8 + 4\|x\|, 1),\ i = 1, \ldots, 500.$$

$H_{X,L}$ $H_{X,M}$ $H_{X,F}$







Merci pour votre attention !

Chagny, G. et Roche A. (2014). Adaptive and Minimax estimation of the Cumulative Distribution Function given a functional covariate, *hal-00931228*.

Ferraty, F., Laksaci, A., Tadj, A. et Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables, *J. Stat. Plan. Inference*, 140(2):335–352.

Ferraty, F., Laksaci, A. et Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.*, 9(1):47–76.

Méthode de Goldenshluger-Lepski et stratégies connexes

Chagny, G. (2013). Penalization versus Goldenshluger-Lepski strategies in warped bases regression, *ESAIM Probab. Statist.*, 17:328–358.

Goldenshluger, A. et Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality, *Ann. Statist.*, 39(3):1608–1632.

Analyse des données fonctionnelles

Ferraty, F. et Vieu, P. (2006). *Nonparametric Functional Data Analysis*, Springer Series in Statistics, Springer, New York.

Ramsay, J.O. et Silverman, B.W. (2005). *Functional Data Analysis*, Springer Series in Statistics, Springer, 2nd ed.