

Performances Statistiques d'estimateurs non-linéaires

Michaël CHICHIGNOUD

LATP
Université de Provence

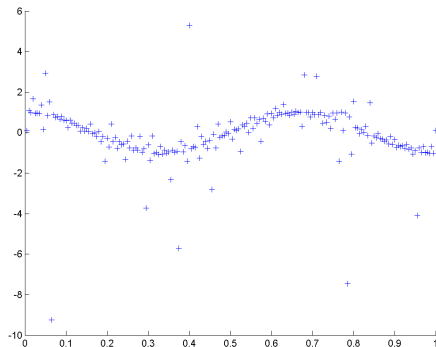
Marseille, 25 novembre 2010

Plan

- 1 Régression non-paramétrique
 - Motivation
 - Méthode des polynômes locaux
- 2 Estimateur Bayésien
 - Approche Minimax
 - Adaptation : Méthode de Lepski
 - Inégalités exponentielles
- 3 Critère de Huber
 - Fonction de Huber
 - Inégalités exponentielles
- 4 Conclusions et Perspectives

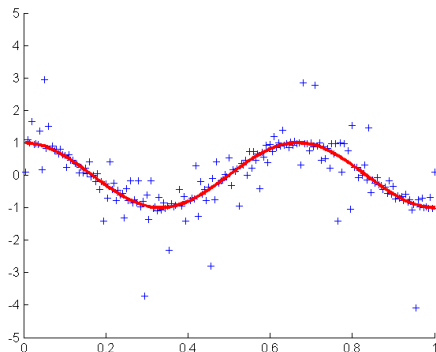
Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



Statistique mathématique

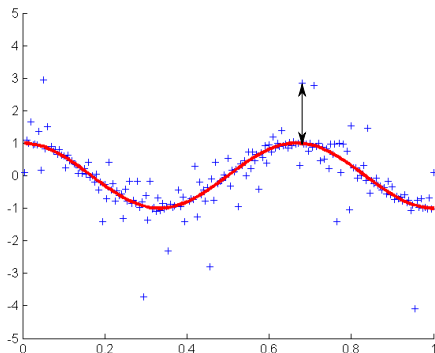
On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



Représentation : $Y_i \sim f(X_i)$

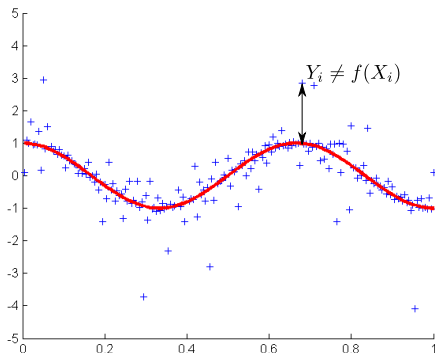
Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



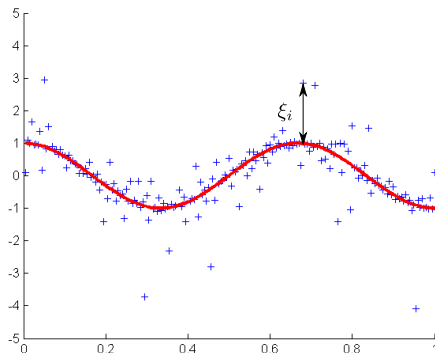
Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



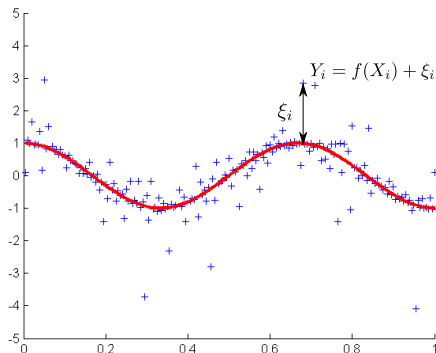
Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



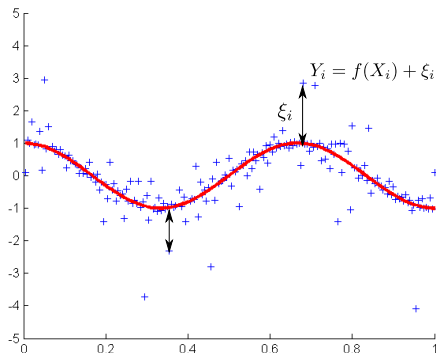
Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



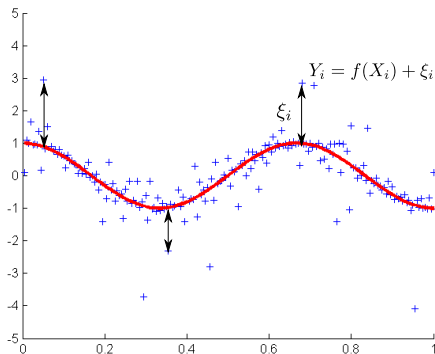
Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



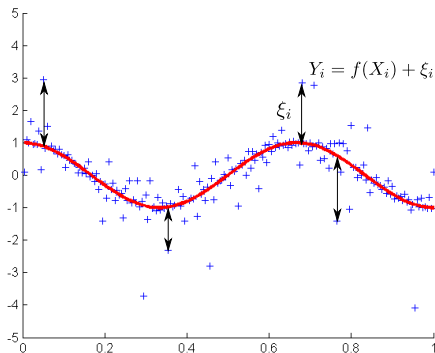
Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



Statistique mathématique

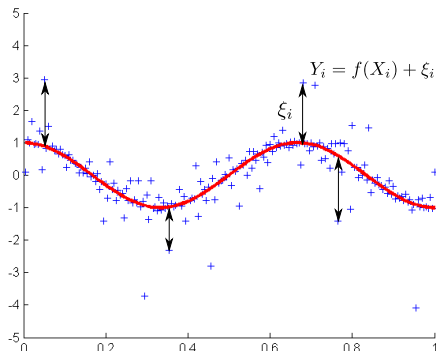
On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



Modèle classique : $Y_i = f(X_i) + \xi_i$ où ξ_i est aléatoire.

Statistique mathématique

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, où $X_i \in [0, 1]$ et $Y_i \in \mathbb{R}$



Modèle classique : $Y_i = f(X_i) + \xi_i$ où ξ_i est aléatoire.

Estimateurs linéaires $\Rightarrow \mathbb{E}_f \xi_j = 0$ et $\mathbb{E}_f \xi_j^2 < \infty$, Tsybakov [2008].

Si $f \equiv \theta$

- $\xi_j \sim \mathcal{N}(0, 1)$: $\hat{\theta}_n - \theta \approx \frac{1}{\sqrt{n}}$, $\hat{\theta}_n = \bar{Y}_n$,
- $\xi_j \sim \mathcal{U}_{[-1,1]}$: $\hat{\theta}_n - \theta \approx \frac{1}{n}$, $\hat{\theta}_n = \frac{\max_i Y_i + \min_i Y_i}{2}$,
- $\xi_j \sim \text{Cauchy}(0, 1)$: $\hat{\theta}_n - \theta \approx \frac{1}{\sqrt{n}}$, $\hat{\theta}_n = \text{Med}(Y^{(n)})$.

But de la thèse : étudier les modèles statistiques pour lesquels les estimateurs linéaires ne sont pas optimaux.

Estimateurs linéaires $\Rightarrow \mathbb{E}_f \xi_j = 0$ et $\mathbb{E}_f \xi_j^2 < \infty$, Tsybakov [2008].

Si $f \equiv \theta$

- $\xi_j \sim \mathcal{N}(0, 1)$: $\hat{\theta}_n - \theta \approx \frac{1}{\sqrt{n}}$, $\hat{\theta}_n = \bar{Y}_n$,
- $\xi_j \sim \mathcal{U}_{[-1,1]}$: $\hat{\theta}_n - \theta \approx \frac{1}{n}$, $\hat{\theta}_n = \frac{\max_i Y_i + \min_i Y_i}{2}$,
- $\xi_j \sim \text{Cauchy}(0, 1)$: $\hat{\theta}_n - \theta \approx \frac{1}{\sqrt{n}}$, $\hat{\theta}_n = \text{Med}(Y^{(n)})$.

But de la thèse : étudier les modèles statistiques pour lesquels les estimateurs linéaires ne sont pas optimaux.

Régression générale

Observations indépendantes:

$$(Y_1, \dots, Y_n)_{i=1, \dots, n} \in \mathbb{R}^n, \quad n \in \mathbb{N}^*,$$

où

$$\mathbb{P}_{Y_i}(dx) = g(x, f(X_i)) dx, \quad i = 1, \dots, n,$$

$f : [0, 1]^d \rightarrow \mathbb{R}$ régulière et $\forall i$

$$X_i \in \left\{ 1/n^{1/d}, 2/n^{1/d}, \dots, 1 \right\}^d.$$

Espace de Hölder isotrope

Pour tout $(p_1, \dots, p_d) \in \mathbb{N}^d$, posons $\vec{p} = (p_1, \dots, p_d)$ et $|\vec{p}| = p_1 + \dots + p_d$.

Définition

Soient $\beta > 0$, $L > 0$ et $M > 0$. La classe de Hölder $\mathbb{H}_d(\beta, L, M)$ est l'ensemble des fonctions $f : [0, 1]^d \rightarrow \mathbb{R}$ telles que toutes ses dérivées d'ordre $|\vec{p}|$ existent sur $[0, 1]^d$ et $\forall x, y \in [0, 1]^d$, $\forall |\vec{p}| = \lfloor \beta \rfloor$ vérifient

$$\left| \frac{\partial^{|\vec{p}|} f(x)}{\partial x_1^{p_1} \dots \partial x_d^{p_d}} - \frac{\partial^{|\vec{p}|} f(y)}{\partial y_1^{p_1} \dots \partial y_d^{p_d}} \right| \leq L \|x - y\|_1^{\beta - \lfloor \beta \rfloor},$$

$$\sum_{m=0}^{\lfloor \beta \rfloor} \sum_{|\vec{p}|=m} \sup_{x \in [0, 1]^d} \left| \frac{\partial^{|\vec{p}|} f(x)}{\partial x_1^{p_1} \dots \partial x_d^{p_d}} \right| \leq M.$$

Exemples

- *Régression Gaussienne.* On observe

$$Y_i = f(X_i) + \xi_i$$

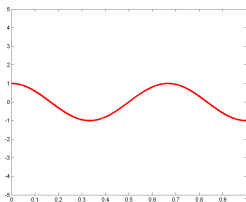
avec $\xi_i \sim \mathcal{N}(0, 1)$ i.i.d.

Exemples: Problèmes standards en traitement du Signal et en imagerie.

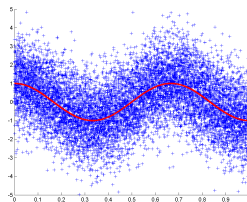
- *Régression inhomogène de Poisson.* La loi de Y_i s'écrit

$$g(k, f_i) = \mathbb{P}_f(Y_i = k) = \frac{[f(X_i)]^k}{k!} \exp\{-f(X_i)\}, \quad k \in \mathbb{N}.$$

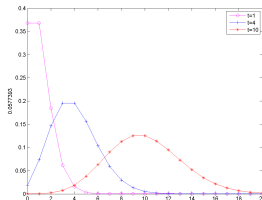
Exemples: caméra vidéo, Tomographie, etc.



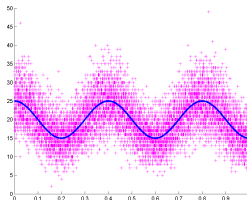
(a) Fonction de régression.



(b) Régression Gaussienne, $\sigma = 1$.



(c) Densités de poisson.



(d) Observations de Poisson.

- *Régression α* . Soit $0 < \alpha < 1/2$ et

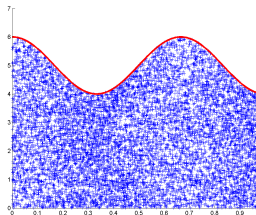
$$Y_i = f(X_i) + \epsilon_i,$$

où $(\epsilon_i)_i$ sont i.i.d. de densité $g_\epsilon(x) = C_\alpha e^{-|x|^\alpha}$, $x \in \mathbb{R}$.

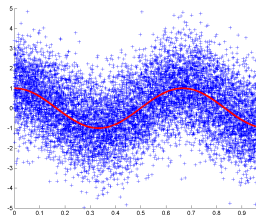
- *Régression multiplicative uniforme*.

$$Y_i = f(X_i) \times U_i, \quad U_i \sim \mathcal{U}_{[0,1]}.$$

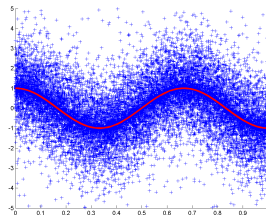
Exemple : modèle de frontière (Simar and Wilson [2000] ou Passy and al. [2007])



Régression Uniforme.

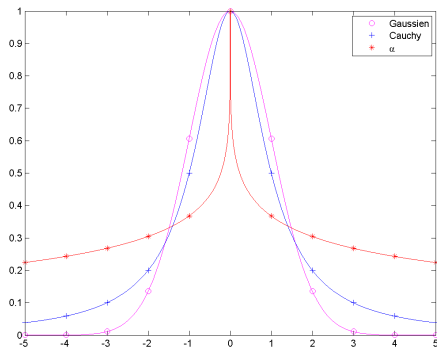


Régression gaussienne.



Régression $\alpha = 1/4$

Densité



Estimation de f

Soit \hat{f}_n une reconstruction de f fait à partir des données.

$$\hat{f}_n - f$$

Estimation de f

Soit \hat{f}_n une reconstruction de f fait à partir des données.

$$\ell(\hat{f}_n, f)$$

Estimation de f

Soit \hat{f}_n une reconstruction de f fait à partir des données.

$$\mathbb{E}_f \ell(\hat{f}_n, f)$$

Estimation de f

Soit \hat{f}_n une reconstruction de f fait à partir des données.

$$\mathbb{E}_f \ell(\hat{f}_n, f) \leq \varphi_n$$

Estimation de f

Soit \hat{f}_n une reconstruction de f fait à partir des données.

$$\sup_{f \in \mathbb{H}_d(\beta, L, M)} \mathbb{E}_f \ell(\hat{f}_n, f) \leq \varphi_{n, \gamma}(\beta)$$

On choisit $\ell(\hat{f}_n, f) = |\hat{f}_n(y) - f(y)|$.

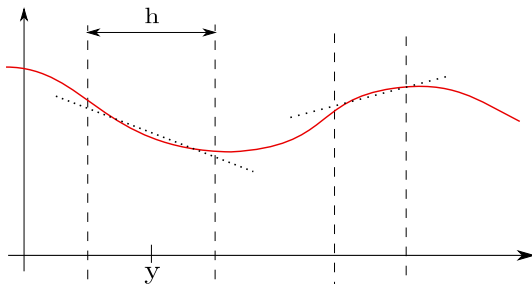
Estimation de f

Soit \hat{f}_n une reconstruction de f fait à partir des données.

$$\sup_{f \in \mathbb{H}_d(\beta, L, M)} \mathbb{E}_f \ell(\hat{f}_n, f) \leq \varphi_{n, \gamma}(\beta)$$

On choisit $\ell(\hat{f}_n, f) = |\hat{f}_n(y) - f(y)|$.

Méthode des polynômes locaux



Approximation par un polynôme d'ordre 1.

Méthode des polynômes locaux

Le but est d'estimer la fonction f en un point $y \in [0, 1]^d$.

Idée : Sur un voisinage de y de taille h , noté $V_h(y)$, on approche f par un polynôme f_θ où $\theta \in \Theta = [-M, M]^{D_b}$, où

$$f_\theta(x) = \sum_{\vec{p} \in \mathcal{P}_b} \theta_{\vec{p}} \left(\frac{x - y}{h} \right)^{\vec{p}}, \quad \mathcal{P}_b = \{ \vec{p} \in \mathbb{N}^d \mid |\vec{p}| \leq b \}$$

On estime les coefficients θ avec un critère $\tilde{\pi}_h(\cdot, \cdot)$, ainsi

$$\hat{\theta} = \arg \min_{t \in \Theta} \tilde{\pi}_h(Y, f_t), \quad h > 0,$$

est un *estimateur par polynômes locaux* (Katskovnik [1985]).

Méthode des polynômes locaux

Le but est d'estimer la fonction f en un point $y \in [0, 1]^d$.

Idée : Sur un voisinage de y de taille h , noté $V_h(y)$, on approche f par un polynôme f_θ où $\theta \in \Theta = [-M, M]^{D_b}$, où

$$f_\theta(x) = \sum_{\vec{p} \in \mathcal{P}_b} \theta_{\vec{p}} \left(\frac{x - y}{h} \right)^{\vec{p}}, \quad \mathcal{P}_b = \{ \vec{p} \in \mathbb{N}^d \mid |\vec{p}| \leq b \}$$

On estime les coefficients θ avec un critère $\tilde{\pi}_h(\cdot, \cdot)$, ainsi

$$\hat{\theta} = \arg \min_{t \in \Theta} \tilde{\pi}_h(Y, f_t), \quad h > 0,$$

est un *estimateur par polynômes locaux* ([Katkovnik \[1985\]](#)).

Estimateur bayésien

Définissons la *pseudo vraisemblance* pour tout $u \in \Theta$

$$L_h(u) = L_h(u, Y^{(n)}) = \prod_{X_i \in V_h(y)} g(Y_i, f_u(X_i)).$$

Soit $\hat{\theta}$ la solution du problème de minimization suivant :

$$\hat{\theta} = \arg \min_{t \in \Theta} \tilde{\pi}_h(t), \quad \tilde{\pi}_h(t) = \int_{\Theta} \|t - u\|_1 (L_h(u))^{1/m} du, \quad t \in \Theta$$

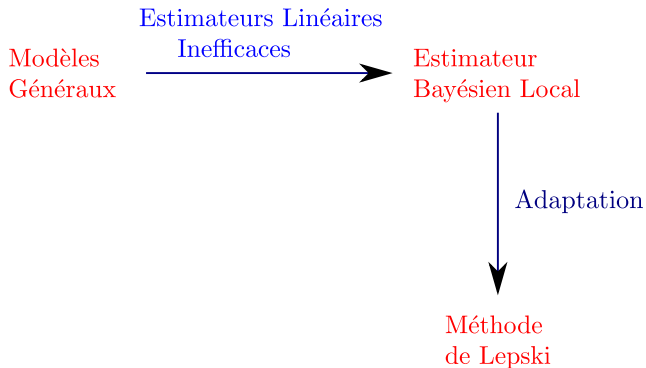
Pour tout h , l'estimateur bayésien est $\hat{f}^h(y) = f_{\hat{\theta}}(y)$.

Has'minskii and Ibragimov [1981] et Polzehl and Spokoiny [2006].

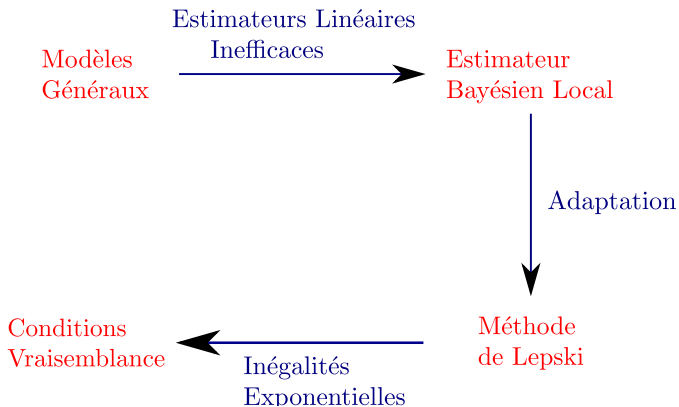
Hypothèses sur la vraisemblance L_h



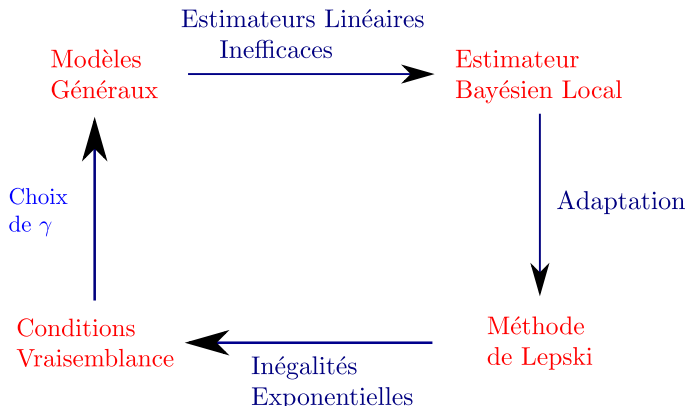
Hypothèses sur la vraisemblance L_h



Hypothèses sur la vraisemblance L_h



Hypothèses sur la vraisemblance L_h



Hypothèses sur la vraisemblance L_h

Soit f_θ , $\theta \in \Theta$ l'approximation inconnue de f . Soit *pseudo rapport de vraisemblance*,

$$Z_{n,\theta}(u) = \frac{L_h(\theta + u N_h^{-1})}{L_h(\theta)}, \quad u \in U_n = N_h [\Theta - \theta],$$

où la normalisation $N_h = (nh^d)^{1/\gamma}$, $\gamma > 0$.

2 Hypothèses sur la vraisemblance L_h

Supposons qu'il existe les constantes $\tau > 0$, $m \geq 1$ et $c_1, c_2, c_3, C_1, C_2, s_1, s_2 > 0$ telle que $\forall n > 1$, $f \in \mathbb{H}_d(\beta, L, M)$, et $h \in \mathcal{H}_n$,

$$\textcircled{1} \int_{[0, \delta]^{D_b}} \mathbb{E}_f \left[Z_{n, \theta}^{1/m}(0) - Z_{n, \theta}^{1/m}(v) \right]_+ dv \leq C_1 \delta^{D_b + \tau} e^{c_1 \mathcal{N}_\gamma(h)},$$

$$\textcircled{2} \mathbb{E}_f Z_{n, \theta}^{1/m}(u) \leq C_2 \exp \{c_2 \mathcal{N}_\gamma(h) - c_3 \|u\|_1^\gamma\},$$

pour tout $0 < \delta \leq s_1$ et $u \in U_n : \|u\|_1 \geq s_2 \mathcal{N}_\gamma(h)$, où $\mathcal{N}_\gamma(h) = (L d h^\beta \times N_h)^\gamma$.

Pour l'adaptation, il est nécessaire de connaître les constantes c_3, s_2, m, τ and γ .

2 Hypothèses sur la vraisemblance L_h

Supposons qu'il existe les constantes $\tau > 0$, $m \geq 1$ et $c_1, c_2, c_3, C_1, C_2, s_1, s_2 > 0$ telle que $\forall n > 1$, $f \in \mathbb{H}_d(\beta, L, M)$, et $h \in \mathcal{H}_n$,

$$\textcircled{1} \int_{[0, \delta]^{D_b}} \mathbb{E}_f \left[Z_{n, \theta}^{1/m}(0) - Z_{n, \theta}^{1/m}(v) \right]_+ dv \leq C_1 \delta^{D_b + \tau} e^{c_1 \mathcal{N}_\gamma(h)},$$

$$\textcircled{2} \mathbb{E}_f Z_{n, \theta}^{1/m}(u) \leq C_2 \exp \{c_2 \mathcal{N}_\gamma(h) - c_3 \|u\|_1^\gamma\},$$

pour tout $0 < \delta \leq s_1$ et $u \in U_n : \|u\|_1 \geq s_2 \mathcal{N}_\gamma(h)$, où $\mathcal{N}_\gamma(h) = (Ldh^\beta \times N_h)^\gamma$.

Pour l'adaptation, il est nécessaire de connaître les constantes c_3, s_2, m, τ and γ .

2 Hypothèses sur la vraisemblance L_h

Supposons qu'il existe les constantes $\tau > 0$, $m \geq 1$ et $c_1, c_2, c_3, C_1, C_2, s_1, s_2 > 0$ telle que $\forall n > 1$, $f \in \mathbb{H}_d(\beta, L, M)$, et $h \in \mathcal{H}_n$,

$$\textcircled{1} \int_{[0, \delta]^{D_b}} \mathbb{E}_f \left[Z_{n, \theta}^{1/m}(0) - Z_{n, \theta}^{1/m}(v) \right]_+ dv \leq C_1 \delta^{D_b + \tau} e^{c_1 \mathcal{N}_\gamma(h)},$$

$$\textcircled{2} \mathbb{E}_f Z_{n, \theta}^{1/m}(u) \leq C_2 \exp \{c_2 \mathcal{N}_\gamma(h) - c_3 \|u\|_1^\gamma\},$$

pour tout $0 < \delta \leq s_1$ et $u \in U_n : \|u\|_1 \geq s_2 \mathcal{N}_\gamma(h)$, où $\mathcal{N}_\gamma(h) = (Ldh^\beta \times N_h)^\gamma$.

Pour l'adaptation, il est nécessaire de connaître les constantes c_3, s_2, m, τ and γ .

2 Hypothèses sur la vraisemblance L_h

Supposons qu'il existe les constantes $\tau > 0$, $m \geq 1$ et $c_1, c_2, c_3, C_1, C_2, s_1, s_2 > 0$ telle que $\forall n > 1$, $f \in \mathbb{H}_d(\beta, L, M)$, et $h \in \mathcal{H}_n$,

$$\textcircled{1} \int_{[0, \delta]^{D_b}} \mathbb{E}_f \left[Z_{n, \theta}^{1/m}(0) - Z_{n, \theta}^{1/m}(v) \right]_+ dv \leq C_1 \delta^{D_b + \tau} e^{c_1 \mathcal{N}_\gamma(h)},$$

$$\textcircled{2} \mathbb{E}_f Z_{n, \theta}^{1/m}(u) \leq C_2 \exp \{ c_2 \mathcal{N}_\gamma(h) - c_3 \|u\|_1^\gamma \},$$

pour tout $0 < \delta \leq s_1$ et $u \in U_n : \|u\|_1 \geq s_2 \mathcal{N}_\gamma(h)$, où $\mathcal{N}_\gamma(h) = (Ldh^\beta \times N_h)^\gamma$.

Pour l'adaptation, il est nécessaire de connaître les constantes c_3, s_2, m, τ and γ .

Résultats minimax : bornes supérieures

Posons $\bar{h} = (L^\gamma n)^{-\frac{1}{\gamma\beta+d}}$.

Théorème

Soit $\beta, L, M > 0$ et $\hat{f}^{\bar{h}}$ l'estimateur bayésien,
Alors $\forall q \geq 1, \forall n \in \mathbb{N}^*$ tel que $N_{\bar{h}} \geq 1$, on a :

$$\sup_{f \in \mathbb{H}_d(\beta, L, M)} \mathbb{E}_f |\hat{f}^{\bar{h}}(y) - f(y)|^q \leq \bar{C} \varphi_{n, \gamma}^q(\beta).$$

où la constante $\bar{C} = \bar{C}(q, L, M, \beta, d, g)$ ne dépend pas de n .

Théorème

Soit $\hat{f}^h(y)$ l'estimateur choisi dans la famille $\{\hat{f}^h(y)\}_{h \in [h_{\min}, h_{\max}]}$ avec la règle de Lepski. Pour tout $\beta \in]0, b]$, $L, M > 0$, on a :

$$\limsup_{n \rightarrow \infty} \sup_{f \in \mathbb{H}_d(\beta, L, M)} \phi_{n, \gamma}^{-q}(\beta) \mathbb{E}_f |\hat{f}^h(y) - f(y)|^q < \infty,$$

$$\phi_{n, \gamma}(\beta) \simeq \left[n^{-1} (\ln n)^{\frac{1}{\gamma}} \right]^{\frac{\beta}{\gamma\beta+d}}, \quad \beta > 0, \gamma > 0.$$

Vitesses obtenues

Régression	Forme	Vitesse Adaptative
Gaussienne	$Y_i = f(X_i) + \xi_i$	$\phi_{n,2}$
Poisson	$Y_i \sim \mathcal{P}(f(X_i))$	$\phi_{n,2}$
$\exp\{- x ^\alpha\}$, $\alpha < 1/2$	$Y_i = f(X_i) + \epsilon_i$	$\phi_{n,1+2\alpha}$
Multiplicative Uniforme	$Y_i = f(X_i) \times U_i$	$\phi_{n,1}$

$$\phi_{n,\gamma}(\beta) \simeq \left[n^{-1} (\ln n)^{\frac{1}{\gamma}} \right]^{\frac{\beta}{\gamma\beta+d}}, \quad \beta > 0, \gamma > 0.$$

Idée de preuve

$$\begin{aligned} \mathbb{P}_f(N_h | \hat{f}^h(y) - f(y) | \geq \varepsilon) &\leq \mathbb{P}_f(N_h \| \hat{\theta} - \theta \|_1 \geq \varepsilon) \\ &\leq \mathbb{P}_f \left(\inf_{N_h \| t - \theta \|_1 \geq \varepsilon} \tilde{\pi}_h(t) \leq \tilde{\pi}_h(\theta) \right). \end{aligned}$$

Ainsi après quelques lignes, on obtient

$$\begin{aligned} &\mathbb{P}_f(N_h | \hat{f}^h(y) - f(y) | \geq \varepsilon) \\ &\leq \mathbb{P}_f \left\{ \int_{U_n} Z_{n,\theta}^{1/m}(v) dv < \nu/2 \right\} \\ &\quad + \mathbb{P}_f \left\{ \int_{U_n \cap \{\|u\|_1 > \varepsilon\}} \|u\|_1 Z_{n,\theta}^{1/m}(u) du > \nu \varepsilon/4 \right\}. \end{aligned}$$

Modèle

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, telles que

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \dots, n.$$

Ici $f : [0, 1]^d \rightarrow \mathbb{R}$ régulière et inconnue.

Les variables $(\xi_i)_{i=1, \dots, n}$ sont i.i.d. et de même loi que $\xi \sim g_\xi$.

Le design $(X_i)_{i=1, \dots, n}$ sont aléatoires i.i.d. de loi uniforme sur $[0, 1]^d$.

Condition sur g_ξ :

- g_ξ symétrique ,
- $g_\xi(0) \geq A > 0$ et g_ξ continue en 0.

Modèle

On observe les couples de variables $(X_i, Y_i)_{i=1, \dots, n}$, telles que

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \dots, n.$$

Ici $f : [0, 1]^d \rightarrow \mathbb{R}$ régulière et inconnue.

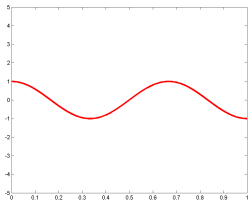
Les variables $(\xi_i)_{i=1, \dots, n}$ sont i.i.d. et de même loi que $\xi \sim g_\xi$.

Le design $(X_i)_{i=1, \dots, n}$ sont aléatoires i.i.d. de loi uniforme sur $[0, 1]^d$.

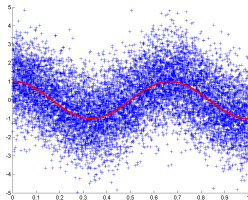
Condition sur g_ξ :

- g_ξ symétrique ,
- $g_\xi(0) \geq A > 0$ et g_ξ continue en 0.

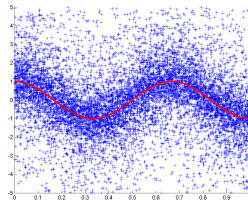
Exemples



Fonction de régression.



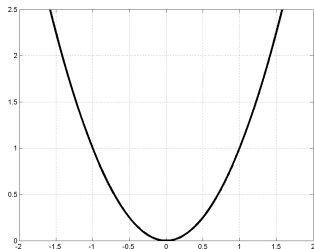
Régression Gaussienne



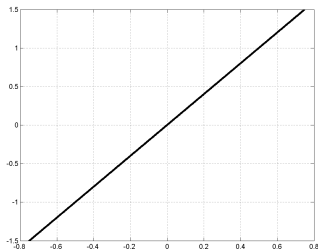
Régression de Cauchy.

Critère ℓ_2 (moindres carrés)

$$\tilde{\pi}_h(Y, f_t) = \sum_{i=1}^n (Y_i - f_t(X_i))^2 \mathbb{I}_{X_i \in V_h(y)}.$$



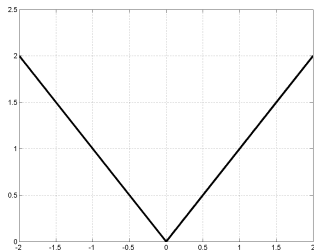
Fonction x^2 .



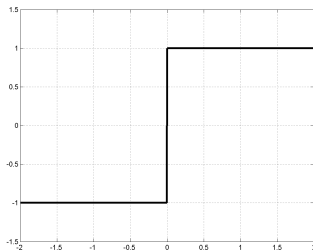
Dérivée

Critère ℓ_1

$$\tilde{\pi}_h(Y, f_t) = \frac{1}{nh^d} \sum_{i=1}^n |Y_i - f_t(X_i)| \mathbb{I}_{X_i \in V_h(y)}.$$



Fonction $|x|$.

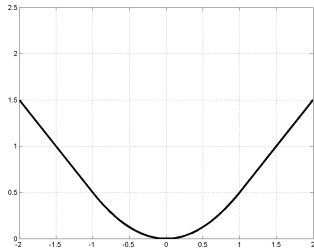


Dérivée

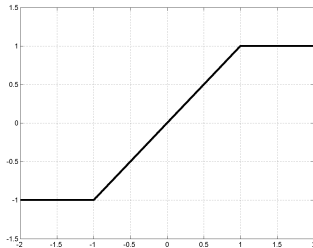
Critère de Huber

$$\tilde{\pi}_h(Y, f_t) = \frac{1}{nh^d} \sum_{i=1}^n Q[Y_i - f_t(X_i)] \mathbb{I}_{X_i \in V_h(y)},$$

où $Q(z) = \gamma (|z| - \frac{\gamma}{2}) \mathbb{I}_{|z| > \gamma} + \frac{z^2}{2} \mathbb{I}_{|z| \leq \gamma}$ (Voir [Huber \[1964\]](#)).



Fonction de Huber.



Dérivée

Estimateur de Huber

Ainsi,

$$\tilde{\theta} = \arg \min_{t \in \Theta} \frac{1}{nh^d} \sum_{i=1}^n Q[Y_i - f_t(X_i)] \mathbb{I}_{X_i \in V_h(y)}, \quad h > 0.$$

On peut alors construire une collection d'estimateurs de Huber

$$\left\{ \tilde{f}^h(y) = f_{\tilde{\theta}}(y) \right\}_{h>0}.$$

Adaptation : [Hall et Jones \[1990\]](#), [Hardle et Tsybakov \[1992\]](#)
et [Reiss et al. \[2010\]](#)

Théorème

Soit $\tilde{f}^{\hat{h}}(y)$ l'estimateur choisi par la procédure de Lepski.
Alors $\forall \beta \in]0, b]$, $L > 0$:

$$\limsup_{n \rightarrow \infty} \sup_{f \in \mathbb{H}(\beta, L, M)} \phi_n^{-r}(\beta) \mathbb{E}_f |\tilde{f}^{\hat{h}}(y) - f(y)|^r < \infty, \quad r \geq 1,$$

$$\phi_n(\beta) \simeq \left[n^{-1} \sqrt{\ln n} \right]^{\frac{\beta}{2\beta+d}}, \quad \beta > 0.$$

Inégalités exponentielles

Lemme

Pour tout $f \in \mathbb{H}(\beta, L, M)$, $n \in \mathbb{N}^*$, $h \in \mathcal{H}_n$,
Alors pour tout $\varepsilon > s \varepsilon_0(h)$, on a:

$$\mathbb{P}_f \left(|\tilde{f}^h(y) - f(y)| \geq \frac{\varepsilon}{\sqrt{nh^d}} \right) \leq \Sigma \exp \left\{ -c_1 \frac{(\varepsilon - c_2 \varepsilon_0(h))^2}{A + B\varepsilon/\sqrt{nh^d}} \right\},$$

où $\varepsilon_0(h) = Ldh^\beta \times \sqrt{nh^d}$.

Preuve : Principe de M-estimation.

Idée de preuve

Pour $\vec{p} \in \mathcal{P}_b$, soit

$$\tilde{D}_h^{\vec{p}}(t) = \frac{\partial \tilde{\pi}_h(t)}{\partial t_{\vec{p}}}(t) = -\frac{1}{nh} \sum_{i=1}^n Q'(Y_i - f_t(X_i)) \left(\frac{X_i - y}{h} \right)^{\vec{p}} \mathbb{I}_{[X_i \in V_h(y)]},$$

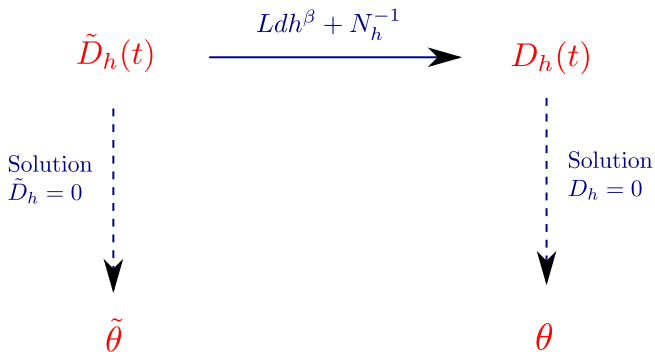
où $Q'(z) = \frac{z}{|z|} \mathbb{I}_{|z|>1} + z \mathbb{I}_{|z|\leq 1}$. On a

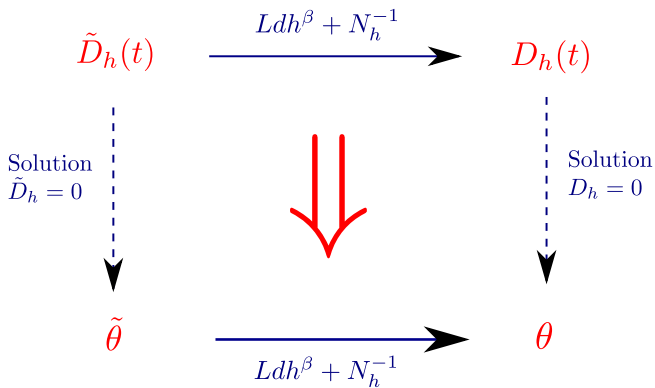
$$\tilde{\theta} = \arg \min_t \tilde{\pi}_h(t) \iff \tilde{D}_h(\tilde{\theta}) = 0.$$

Chaque dérivée partielle est une somme de v.a. bornées et indépendantes.

$$\tilde{D}_h(t) \xrightarrow{Ldh^\beta + N_h^{-1}} D_h(t)$$

$$\begin{array}{ccc} \tilde{D}_h(t) & \xrightarrow{Ldh^\beta + N_h^{-1}} & D_h(t) \\ \downarrow \text{Solution } \tilde{D}_h = 0 & & \\ \tilde{\theta} & & \end{array}$$





Majoration de la fonction de perte

On a :

$$\begin{aligned} & \mathbb{P}_f \left(N_h |\tilde{f}^h(y) - f(y)| \geq \varepsilon \right) \\ & \leq \mathbb{P}_f \left(N_h \frac{\sqrt{D_b}}{\sqrt{\lambda}} \left\| \tilde{D}_h(\tilde{\theta}) - D_h(\tilde{\theta}) \right\|_2 \geq \varepsilon \right) \\ & \leq \sum_{\vec{p} \in \mathcal{P}_b} \mathbb{P}_f \left(N_h \sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - D_h^{\vec{p}}(t) \right| \geq \frac{\varepsilon \sqrt{\lambda}}{D_b} \right). \end{aligned}$$

Contrôle de $\sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - D_h^{\vec{p}}(t) \right|$

$$\begin{aligned}
 & \sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - D_h^{\vec{p}}(t) \right| \\
 & \leq \sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - \mathbb{E}_f \tilde{D}_h^{\vec{p}}(t) \right| + \sup_{t \in \Theta} \left| \mathbb{E}_f \tilde{D}_h^{\vec{p}}(t) - D_h^{\vec{p}}(t) \right| \\
 & = \sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - \mathbb{E}_f \tilde{D}_h^{\vec{p}}(t) \right| + Ldh^\beta \\
 & = \sup_{t \in \Theta} \left| \sum_{i=1}^n \mathcal{W}_t(X_i, \xi_i) - \mathbb{E}_f \mathcal{W}_t(X_i, \xi_i) \right| + Ldh^\beta.
 \end{aligned}$$

- On utilise un argument de chaînage (continuité de la dérivée de Huber) pour contrôler le supremum,
- l'inégalité de Bernstein (Dérivée de Huber bornée).

Contrôle de $\sup_{t \in \Theta} |\tilde{D}_h^{\vec{\rho}}(t) - D_h^{\vec{\rho}}(t)|$

$$\begin{aligned}
 & \sup_{t \in \Theta} |\tilde{D}_h^{\vec{\rho}}(t) - D_h^{\vec{\rho}}(t)| \\
 & \leq \sup_{t \in \Theta} |\tilde{D}_h^{\vec{\rho}}(t) - \mathbb{E}_f \tilde{D}_h^{\vec{\rho}}(t)| + \sup_{t \in \Theta} |\mathbb{E}_f \tilde{D}_h^{\vec{\rho}}(t) - D_h^{\vec{\rho}}(t)| \\
 & = \sup_{t \in \Theta} |\tilde{D}_h^{\vec{\rho}}(t) - \mathbb{E}_f \tilde{D}_h^{\vec{\rho}}(t)| + Ldh^\beta \\
 & = \sup_{t \in \Theta} \left| \sum_{i=1}^n \mathcal{W}_t(X_i, \xi_i) - \mathbb{E}_f \mathcal{W}_t(X_i, \xi_i) \right| + Ldh^\beta.
 \end{aligned}$$

- On utilise un argument de chaînage (continuité de la dérivée de Huber) pour contrôler le supremum,
- l'inégalité de Bernstein (Dérivée de Huber bornée).

Contrôle de $\sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - D_h^{\vec{p}}(t) \right|$

$$\begin{aligned}
 & \sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - D_h^{\vec{p}}(t) \right| \\
 & \leq \sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - \mathbb{E}_f \tilde{D}_h^{\vec{p}}(t) \right| + \sup_{t \in \Theta} \left| \mathbb{E}_f \tilde{D}_h^{\vec{p}}(t) - D_h^{\vec{p}}(t) \right| \\
 & = \sup_{t \in \Theta} \left| \tilde{D}_h^{\vec{p}}(t) - \mathbb{E}_f \tilde{D}_h^{\vec{p}}(t) \right| + Ldh^\beta \\
 & = \sup_{t \in \Theta} \left| \sum_{i=1}^n \mathcal{W}_t(X_i, \xi_i) - \mathbb{E}_f \mathcal{W}_t(X_i, \xi_i) \right| + Ldh^\beta.
 \end{aligned}$$

- On utilise un argument de chaînage (continuité de la dérivée de Huber) pour contrôler le supremum,
- l'inégalité de Bernstein (Dérivée de Huber bornée).

Conclusions

- Recherche de la vitesse optimale en fonction de la vraisemblance,
- Estimateur de Huber qui ne dépend pas de la densité mais qui peut ne pas être optimal

Perspectives

1 Estimateur bayésien

- Conditions Vraisemblance \Rightarrow Distance d'Hellinger,
- $(X_i, Y_i)_i$ non i.i.d.
- Bornes inférieures pour l'adaptation,
- Approche bayésienne,
- Inégalités d'oracle.

2 Estimateur de Huber

- Hypthèses : g_ξ symétrique,
- Etude des modèles hétéroscedastiques?
- Trouver les conditions nécessaires sur la fonction de contrast,
- Etude des fonctions anisotropes et recherche d'inégalités oracle.

Perspectives

1 Estimateur bayésien

- Conditions Vraisemblance \Rightarrow Distance d'Hellinger,
- $(X_i, Y_i)_i$ non i.i.d.
- Bornes inférieures pour l'adaptation,
- Approche bayésienne,
- Inégalités d'oracle.

2 Estimateur de Huber

- Hypthèses : g_ξ symétrique,
- Etude des modèles hétéroscédastiques?
- Trouver les conditions nécessaires sur la fonction de contrast,
- Etude des fonctions anisotropes et recherche d'inégalités oracle.

Merci pour votre attention.