Soutenance de Doctorat (LMD) en Mathématiques
Spécialité : Mathématiques Financières et Actuariat
Sujet

Estimation non paramétrique pour les données incomplètes et associées

Présentée par: Mme Farida HAMRANI

Encadrée par : Mme Zohra GUESSOUM
Université des Sciences et de la Technologie Houari Boumediene,
Laboratoire MSTD, Algérie

Plan

Plan

Plan

Plan

Plan

Introduction

Le modèle de régression

- Objectif de la régression : étudier les relations entre une variable à expliquer Y et une variable explicative X (unidimensionnelle ou multidimensionnelle).

- Modèle de la régression :

$$Y = r(X) + \epsilon,$$

où r $\longrightarrow$ la fonction de régression inconnue,
ϵ $\longrightarrow$ le terme d'errreur aléatoire.

La régression non paramétrique

Régression non paramétrique : aucune forme spécifique à r(.) à part certaines conditions de régularité.

On cherchera dans une famille fixée de fonctions, quelle est celle pour laquelle les Y sont les plus proche de r(X).

$$\mathbb{E}|r^*(X) - Y|^2 = \min_r \mathbb{E}|r(X) - Y|^2$$

Le minimum est donné par l'espérance conditionnelle.

$$m(x) = \mathbb{E}[Y|X = x]$$

La régression non paramétrique

Régression non paramétrique : aucune forme spécifique à r(.) à part certaines conditions de régularité.

On cherchera dans une famille fixée de fonctions, quelle est celle pour laquelle les Y sont les plus proche de r(X).

$$\mathbb{E}|r^*(X) - Y|^2 = \min_r \mathbb{E}|r(X) - Y|^2$$

Le minimum est donné par l'espérance conditionnelle.

$$m(x) = \mathbb{E}[Y|X = x]$$

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)

- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)
- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)

- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)

- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)
- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)

- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)
- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Historique (régression par la méthode du noyau)

- Données complètes :
  - i.i.d.
    - Nadaraya (1964), Watson (1964)
    - Devroye (1978), Stone (1980, 1982)
  - $\alpha-$ mélangeantes
    - Györfi et al. (1989)
    - Liebscher (2001)
  - associées
    - Oliveira (2012)

- Données tronquées :
  - i.i.d.
    - Ould Saïd et Lemdani (2006)
  - $\alpha-$ mélangeantes
    - Liang et al. (2009)
    - Liang (2011)
  - associées
    - Guessoum et Hamrani (2017)

Estimation sur données complètes et associées

Association

Définition : (Esary, Proschan et Walkup (1967))

Une suite finie de variables aléatoires $X_1, \ldots, X_N$ est dite associée si

$$\mathrm{Cov}(f(X_1, \ldots, X_N), g(X_1, \ldots, X_N)) \geq 0,$$

pour toute paire f, g de fonctions de $\mathbb{R}^N \to \mathbb{R}$ croissantes coordonnée par coordonnée et telles que cette covariance existe.

Une suite infinie de variables aléatoires est associée si toute sous suite finie est associée.

Notations

- $(\Omega, \mathcal{F}, \mathbb{P})$: un espace probabilisé.
- $\{(X_i, Y_i); 1 \leq i \leq N\}$: suite strictement stationnaire de N vecteurs aléatoires associés à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ $(d \geq 1)$.
- $f(.,.)$: densité conjointe du couple (X, Y).
- $v(.)$: densité marginale de X.
- $\theta_{i,j} := \sum_{k=1}^{d} \sum_{l=1}^{d} Cov(X_{i,k}, X_{j,l}) + 2 \sum_{k=1}^{d} Cov(X_{i,k}, Y_j) + Cov(Y_i, Y_j)$: coefficient de covariance.

Fonction de régression

$$m(x) := \mathbb{E}[Y|X = x] = \frac{\int_{\mathbb{R}} yf(x,y)dy}{\int_{\mathbb{R}} f(x,y)dy} =: \frac{\psi(x)}{v(x)}.$$

Notations

- $(\Omega, \mathcal{F}, \mathbb{P})$: un espace probabilisé.
- $\{(X_i, Y_i); 1 \leq i \leq N\}$: suite strictement stationnaire de N vecteurs aléatoires associés à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ $(d \geq 1)$.
- $f(.,.)$: densité conjointe du couple (X, Y).
- $v(.)$: densité marginale de X.
- $\theta_{i,j} := \sum_{k=1}^{d} \sum_{l=1}^{d} Cov(X_{i,k}, X_{j,l}) + 2 \sum_{k=1}^{d} Cov(X_{i,k}, Y_j) + Cov(Y_i, Y_j)$: coefficient de covariance.

Fonction de régression

$$m(x) := \mathbb{E}[Y|X = x] = \frac{\displaystyle\int_{\mathbb{R}} y f(x,y) dy}{\displaystyle\int_{\mathbb{R}} f(x,y) dy} =: \frac{\psi(x)}{v(x)}.$$

L'estimateur de la fonction de régression :

L'estimateur de Nadaraya-Watson

$$\hat{m}_N(x) := \frac{\sum_{i=1}^{N} Y_i K_d\left(\frac{x - X_i}{h_N}\right)}{\sum_{i=1}^{N} K_d\left(\frac{x - X_i}{h_N}\right)}$$

où $(h_N)_{N\geq 1}$ une suite de nombres réels positifs appelée fenêtre tendant vers 0 à ∞.
$K_d : \mathbb{R}^d \to \mathbb{R}$ est un noyau multivarié.

Remarque :

$$K_d(u) = K_d(u_1, \dots, u_d) = \prod_{j=1}^{d} K(u_j), \quad u \in \mathbb{R}^d.$$

Convergence uniforme presque sûre de $\hat{m}_N$:

Hypothèses pour la convergence uniforme presque sûre

Soit $\mathcal{U}$ un sous ensemble compact de $\dot{\mathcal{U}} = \{x \in \mathbb{R}^d | v(x) > \delta > 0\}$ pour un réel $\delta > 0$.

A1. $h_N \to 0$, $Nh_N^d \to \infty$ et $\dfrac{\log^5 N}{Nh_N^d} \to 0$ quand $N \to +\infty$.

A2. K_d est un noyau d'ordre 2 à support compact et hölderienne d'exposant $\beta > 0$. De plus $\int_{\mathbb{R}^d} K_d^2(z)dz < +\infty$ et $\displaystyle\int_{\mathbb{R}^d} |z_1 + \ldots + z_d| K_d^2(z)dz < +\infty$.

A3. Le terme de covariance définit par $\rho(s) := \sup_{|i-j| \geq s} \theta_{i,j}$, $s > 0$, vérifie $\rho(s) \leq \gamma_0 e^{-\gamma s}$, $\gamma_0 > 0$, $\gamma > 0$.

A4. La fonction $\psi(.)$ et la densité $v(.)$ sont bornées, deux fois différentiable et à dérivées partielles bornées.

A5. La fonction $\psi_1(x) := \int_{\mathbb{R}} y^2 f(x,y)dy$ est bornée, différentiable et à dérivées partielles bornées.

A6. La densité conjointe $v_{i,j}$ de (X_i, X_j) est bornée.

Convergence uniforme presque sûre de $\hat{m}_N$:

Théorème 1 :

Sous les hypothèses A1-A6, nous avons pour $N \to \infty$:

$$\sup_{x \in \mathcal{U}} |\hat{m}_N(x) - m(x)| = O\left(\sqrt{\frac{\log N}{N h_N^d}} \vee h_N^2\right) \quad \text{p.s.}$$

Estimation sur données tronquées à gauche

Le modèle aléatoire de troncature à gauche :

Troncature aléatoire à gauche :

Soit Y la variable d'intérêt et T une autre variable aléatoire, si Y et T sont observables uniquement si $Y \geq T$, et rien sinon, on dira que la variable Y est aléatoirement tronquée à gauche par la variable de troncature T.

- A partir d'un échantillon $(Y_1, T_1), \ldots, (Y_N, T_N)$ de taille N fixé mais inconnu de (Y, T), nous ne sommes capables d'observer que les n couples qui vérifient $Y_i \geq T_i$, $i = 1, \ldots, N$ avec $(n \leq N)$.
- $(Y_1, T_1), \ldots, (Y_n, T_n) \rightarrow$ L'échantillon observé.
- $\alpha := \mathbb{P}(Y \geq T) \rightarrow$ La probabilité de (non) troncature.

Le modèle aléatoire de troncature à gauche :

Remarque :

Comme N est inconnu et n est connu (aléatoire), nos résultats ne serons pas établi par rapport à la probabilité $\mathbb{P}$ (relative au N-échantillon) mais par rapport à la probabilité P (relative n-échantillon) définit par

$$P(.) = \mathbb{P}(.|Y \geq T).$$

Notations :

Modèle et notations

- $\{Y_i; i = 1, \ldots, N\}$: suite de N v.a.'s réelles, indépendantes et identiquement distribuées de la variable d'intérêt Y de f.d.r. F.
- $\{X_i; i = 1, \ldots, N\}$: suite de N vecteurs aléatoires, i.i.d., de vecteur aléatoire $X \in \mathbb{R}^d (d \geq 1)$ de covariables de densité v.
- $\{T_i; i = 1, \ldots, N\}$: suite de N v.a's réelles, i.i.d., de la variable de troncature T de f.d.r. G.
- T indépendante de (X, Y).
- $\{(X_i, Y_i, T_i); i = 1, \ldots, n\} \rightarrow$ l'échantillon observé (i.e. $Y_i \geq T_i$).
- $f(x, y)$: densité conjointe du couple (X, Y).

Estimateurs utilisés :

Estimateurs de Lynden-Bell (1971) :

$$F_n(y) := 1 - \prod_{i:Y_i \le y} \left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)}\right], \qquad G_n(t) := \prod_{i:T_i > t} \left[\frac{nC_n(T_i) - 1}{nC_n(T_i)}\right] \text{ où}$$

$C_n(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{T_i \le y \le Y_i\}}$ estime la fonction C définie par $C(y) = P\{T \le y \le Y\}$.

Estimateur de He et Yang (1998) :

$$\alpha_n := \frac{G_n(y)(1 - F_n(y))}{C_n(y)}$$

Estimateur de Lemdani et Ould Saïd (2006) :

$$\hat{v}_n(x) \quad := \quad \frac{\alpha_n}{nh_n^d}\sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right).$$

Estimateurs utilisés :

Estimateurs de Lynden-Bell (1971) :

$$F_n(y) := 1 - \prod_{i:Y_i \leq y} \left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)}\right], \qquad G_n(t) := \prod_{i:T_i > t} \left[\frac{nC_n(T_i) - 1}{nC_n(T_i)}\right] \text{ où}$$

$C_n(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{T_i \leq y \leq Y_i\}}$ estime la fonction C définie par $C(y) = P\{T \leq y \leq Y\}$.

Estimateur de He et Yang (1998) :

$$\alpha_n := \frac{G_n(y)(1 - F_n(y))}{C_n(y)}$$

Estimateur de Lemdani et Ould Saïd (2006) :

$$\hat{v}_n(x) \quad := \quad \frac{\alpha_n}{nh_n^d} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right).$$

Estimateurs utilisés :

Estimateurs de Lynden-Bell (1971) :

$$F_n(y) := 1 - \prod_{i:Y_i \le y} \left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)} \right], \qquad G_n(t) := \prod_{i:T_i > t} \left[\frac{nC_n(T_i) - 1}{nC_n(T_i)} \right] \text{ où}$$

$C_n(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{T_i \le y \le Y_i\}}$ estime la fonction C définie par $C(y) = P\{T \le y \le Y\}$.

Estimateur de He et Yang (1998) :

$$\alpha_n := \frac{G_n(y)(1 - F_n(y))}{C_n(y)}$$

Estimateur de Lemdani et Ould Saïd (2006) :

$$\hat{v}_n(x) \quad := \quad \frac{\alpha_n}{nh_n^d} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right).$$

Estimateur de la fonction de régression :

Estimateur de Lemdani et Ould Saïd (2006) :

$$\hat{m}_n(x) := \frac{\displaystyle\sum_{i=1}^{n} \frac{Y_i}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)}$$

Si on note

$$\hat{\psi}_n(x) := \frac{\alpha_n}{nh_n^d} \sum_{i=1}^{n} \frac{Y_i}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)$$

alors

$$\hat{m}_n(x) = \frac{\hat{\psi}_n(x)}{\hat{v}_n(x)}.$$

Estimateur de la fonction de régression :

Estimateur de Lemdani et Ould Saïd (2006) :

$$\hat{m}_n(x) := \frac{\displaystyle\sum_{i=1}^{n} \frac{Y_i}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)}$$

Si on note

$$\hat{\psi}_n(x) := \frac{\alpha_n}{nh_n^d} \sum_{i=1}^{n} \frac{Y_i}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)$$

alors

$$\hat{m}_n(x) = \frac{\hat{\psi}_n(x)}{\hat{v}_n(x)}.$$

Estimateur de la fonction de régression :

Propriétés asymptotiques de $\hat{m}_n$:

- Dans le cas d'une covariable réelle : Lemdani et Ould Saïd (2006) ont établi les propriétés asymptotiques (convergence uniforme presque sûre et normalité asymptotique) de $\hat{m}_n$.

- Nous étendons ces résultats au cas d'une covariable à valeurs dans $\mathbb{R}^d$.

Estimateur de la fonction de régression :

Propriétés asymptotiques de $\hat{m}_n$:

- Dans le cas d'une covariable réelle : Lemdani et Ould Saïd (2006) ont établi les propriétés asymptotiques (convergence uniforme presque sûre et normalité asymptotique) de $\hat{m}_n$.

- Nous étendons ces résultats au cas d'une covariable à valeurs dans $\mathbb{R}^d$.

Convergence uniforme presque sûre de $\hat{m}_n$:

- Pour toute f.d.r. L, nous noterons par a_L et b_L, respectivement, les bornes inférieure et supérieure du support de L définies respectivement par $a_L = \inf\{y : L(y) > 0\}$ et $b_L = \sup\{y : L(y) < 1\}$.

Hypothèses pour la convergence uniforme presque sûre

Soit $\mathcal{C}$ un sous ensemble compact de $\Omega_0 = \{x \in \mathbb{R}^d | v(x) > \eta > 0\}$.

B1. $a_G < a_F$ et $b_G \leq b_F$.

B2. $h_n \to 0$ et $\frac{\log n}{nh_n^d} \to 0$ quand $n \to +\infty$.

B3. K_d est un noyau d'ordre 2 à support compact et hölderienne d'exposant $\beta > 0$. De plus $\int_{\mathbb{R}^d} K_d^2(z)dz < +\infty$ et $\int_{\mathbb{R}^d} |z_1 + \ldots + z_d| K_d^2(z)dz < +\infty$.

B4. La fonction $\psi(.)$ et la densité $v(.)$ sont bornées, deux fois différentiable et à dérivées partielles bornées.

B5. La fonction $\psi_1(x) := \int_{\mathbb{R}} \frac{y^2}{G(y)} f(x,y)dy$ est bornée, différentiable et à dérivées partielles bornées.

Convergence uniforme presque sûre de $\hat{m}_n$:

- Pour toute f.d.r. L, nous noterons par a_L et b_L, respectivement, les bornes inférieure et supérieure du support de L définies respectivement par $a_L = \inf\{y : L(y) > 0\}$ et $b_L = \sup\{y : L(y) < 1\}$.

Hypothèses pour la convergence uniforme presque sûre

Soit $\mathcal{C}$ un sous ensemble compact de $\Omega_0 = \{x \in \mathbb{R}^d | v(x) > \eta > 0\}$.

B1. $a_G < a_F$ et $b_G \leq b_F$.

B2. $h_n \to 0$ et $\frac{\log n}{nh_n^d} \to 0$ quand $n \to +\infty$.

B3. K_d est un noyau d'ordre 2 à support compact et hölderienne d'exposant $\beta > 0$. De plus $\int_{\mathbb{R}^d} K_d^2(z)dz < +\infty$ et $\int_{\mathbb{R}^d} |z_1 + \ldots + z_d| K_d^2(z)dz < +\infty$.

B4. La fonction $\psi(.)$ et la densité $v(.)$ sont bornées, deux fois différentiable et à dérivées partielles bornées.

B5. La fonction $\psi_1(x) := \int_{\mathbb{R}} \frac{y^2}{G(y)} f(x,y)dy$ est bornée, différentiable et à dérivées partielles bornées.

Convergence uniforme presque sûre de $\hat{m}_n$:

Théorème 2 :

Supposons que les hypothèses B1-B5 sont vérifiées, nous avons

$$\sup_{x\in\mathcal{C}} |\hat{m}_n(x) - m(x)| = O\left(\sqrt{\frac{\log n}{nh_n^d}} \vee h_n^2\right) \text{ P-p.s,} \quad \text{lorsque } n \to +\infty.$$

Normalité asymptotique de $\hat{m}_n$:

Nous avons besoin des hypothèses B1-B4 précédentes et des hypothèses suivantes :

Hypothèses pour la normalité asymptotique

B6. $nh_n^{d+4} \to 0 \quad$ quand $n \to +\infty$.

B7. Il existe $\nu > 2$ tel que $\int_{\mathbb{R}^d}\int_{\mathbb{R}} y^\nu f(x,y)dxdy < +\infty$.

Normalité asymptotique de $\hat{m}_n$:

Théorème 3 :

Sous les hypothèses B1-B4 et B6-B7 et pour tout x tel que $v(x) > 0$, nous avons

$$\sqrt{nh_n^d}\left[\hat{m}_n(x) - m(x)\right] \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(x)\right),$$

avec $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi

$$\sigma^2(x) := \frac{\alpha\left[\Sigma_0(x)v^2(x) - 2\Sigma_1(x)\psi(x)v(x) + \Sigma_2(x)\psi^2(x)\right]}{v^4(x)}\kappa,$$

$$\Sigma_j(x) = \int_{\mathbb{R}} \frac{y^{2-j}}{G(y)} f(x,y)dy, \quad j = 0, 1, 2,$$

et

$$\kappa = \int_{\mathbb{R}^d} K_d^2(z)dz < +\infty.$$

Estimation sur données tronquées à gauche et associées

Modèle aléatoire tronqué à gauche et associé :

Modèle :

- $\{X_i; i = 1, \ldots, N\}$ et $\{Y_i; i = 1, \ldots, N\}$ sont strictement stationnaires et associées.
- $\{T_i; i = 1, \ldots, N\}$ est strictement stationnaire i.i.d.
- $\{T_i; i = 1, \ldots, N\}$ et $\{(X_i, Y_i); i = 1, \ldots, N\}$ sont indépendants.
- $\{(X_i, Y_i, T_i); i = 1, \ldots, n\} \rightarrow$ la suite observée.
- $\theta_{i,j} := \sum_{k=1}^{d} \sum_{l=1}^{d} \mathrm{Cov}(X_{i,k}, X_{j,l}) + 2 \sum_{k=1}^{d} \mathrm{Cov}(X_{i,k}, Y_j) + \mathrm{Cov}(Y_i, Y_j)$: coefficient de covariance.

Objectif :

Etudier les propriétés asymptotiques de l'estimateur à noyau de la fonction de régression $\hat{m}_n(.)$ introduit par Ould Saïd et Lemdani (2006).

Modèle aléatoire tronqué à gauche et associé :

Modèle :

- $\{X_i; i = 1, \ldots, N\}$ et $\{Y_i; i = 1, \ldots, N\}$ sont strictement stationnaires et associées.
- $\{T_i; i = 1, \ldots, N\}$ est strictement stationnaire i.i.d.
- $\{T_i; i = 1, \ldots, N\}$ et $\{(X_i, Y_i); i = 1, \ldots, N\}$ sont indépendants.
- $\{(X_i, Y_i, T_i); i = 1, \ldots, n\} \rightarrow$ la suite observée.
- $\theta_{i,j} := \sum_{k=1}^{d} \sum_{l=1}^{d} Cov(X_{i,k}, X_{j,l}) + 2 \sum_{k=1}^{d} Cov(X_{i,k}, Y_j) + Cov(Y_i, Y_j)$: coefficient de covariance.

Objectif :

Etudier les propriétés asymptotiques de l'estimateur à noyau de la fonction de régression $\hat{m}_n(.)$ introduit par Ould Saïd et Lemdani (2006).

Modèle aléatoire tronqué à gauche et associé :

Pseudo-estimateur

Ici on introduit un pseudo-estimateur de m(.) noté $\tilde{m}_n(.)$ et définit par

$$\tilde{m}_n(x) := \frac{\tilde{\psi}_n(x)}{\tilde{v}_n(x)}$$

où

$$\begin{aligned} \tilde{\psi}_n(x) &:= \frac{\alpha}{nh_n^d} \sum_{i=1}^{n} \frac{Y_i}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right) \\ \tilde{v}_n(x) &:= \frac{\alpha}{nh_n^d} \sum_{i=1}^{n} \frac{1}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right) \end{aligned}$$

Convergence uniforme presque sûre de $\hat{m}_n$:

Hypothèses pour la convergence uniforme presque sûre

Soit D un sous ensemble compact de $\Xi = \{x \in \mathbb{R}^d | v(x) > \delta > 0\}$.

C1. $a_G < a_F$ et $b_G \leq b_F$. C2. $\displaystyle\int_{\mathbb{R}} \frac{dF(z)}{G^2(z)} < +\infty$.

C3. $h_n \to 0$, $nh_n^d \to \infty$ et $\dfrac{\log^5 n}{nh_n^d} \to 0$ quand $n \to +\infty$.

C4. K_d est un noyau d'ordre 2 à support compact et hölderienne d'exposant $\beta > 0$. $\int_{\mathbb{R}^d} K_d^2(z)dz < +\infty$ et $\displaystyle\int_{\mathbb{R}^d} |z_1 + \ldots + z_d| K_d^2(z)dz < +\infty$.

C5. Le terme de covariance définit par $\rho(s) := \sup\limits_{|i-j| \geq s} \Delta_{i,j}, s > 0$, vérifie $\rho(s) \leq \gamma_0 e^{-\gamma s}$, $\gamma_0 > 0$, $\gamma > 0$.

C6. La fonction $\psi(.)$ et la densité $v(.)$ sont bornées, deux fois differentiable et à dérivées partielles bornées.

C7. La fonction $\psi_1(x) := \int_{\mathbb{R}} \dfrac{y^2}{G(y)} f(x,y)dy$ est bornée, différentiable et à dérivées partielles bornées.

C8. La densité conjointe $v_{i,j}^*$ de (X_i, X_j) est bornée.

Convergence uniforme presque sûre de $\hat{m}_n$:

- $\hat{\psi}_n(x) - \psi(x) = \left(\hat{\psi}_n(x) - \tilde{\psi}_n(x)\right) + \left(\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x))\right) + \left(E(\tilde{\psi}_n(x)) - \psi(x)\right).$
- $\hat{v}_n(x) - v(x) = (\hat{v}_n(x) - \tilde{v}_n(x)) + (\tilde{v}_n(x) - E(\tilde{v}_n(x))) + (E(\tilde{v}_n(x)) - v(x)$

Théorème 4 : (Guessoum et Hamrani (2017))

Sous les hypothèses C1 et C3-C8, nous avons pour $n \to \infty$:

$$\sup_{x \in D} \left|\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x))\right| = O\left(\sqrt{\frac{\log n}{n h_n^d}}\right) \quad P\text{- p.s.}$$

Théorème 5 : (Guessoum et Hamrani (2017))

Sous les hypothèses C1, C3-C6 et C8, nous avons pour $n \to \infty$:

$$\sup_{x \in D} |\tilde{v}_n(x) - E(\tilde{v}_n(x))| = O\left(\sqrt{\frac{\log n}{n h_n^d}}\right) \quad P\text{- p.s.}$$

Convergence uniforme presque sûre de $\hat{m}_n$:

- $\hat{\psi}_n(x) - \psi(x) = \left(\hat{\psi}_n(x) - \tilde{\psi}_n(x)\right) + \left(\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x))\right) + \left(E(\tilde{\psi}_n(x)) - \psi(x)\right).$
- $\hat{v}_n(x) - v(x) = (\hat{v}_n(x) - \tilde{v}_n(x)) + (\tilde{v}_n(x) - E(\tilde{v}_n(x))) + (E(\tilde{v}_n(x)) - v(x)$

Théorème 4 : (Guessoum et Hamrani (2017))

Sous les hypothèses C1 et C3-C8, nous avons pour $n \to \infty$:

$$\sup_{x \in D} \left|\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x))\right| = O\left(\sqrt{\frac{\log n}{n h_n^d}}\right) \quad P\text{- p.s.}$$

Théorème 5 : (Guessoum et Hamrani (2017))

Sous les hypothèses C1, C3-C6 et C8, nous avons pour $n \to \infty$:

$$\sup_{x \in D} |\tilde{v}_n(x) - E(\tilde{v}_n(x))| = O\left(\sqrt{\frac{\log n}{n h_n^d}}\right) \quad P\text{- p.s.}$$

Convergence uniforme presque sûre de $\hat{m}_n$:

- $\hat{\psi}_n(x) - \psi(x) = \left(\hat{\psi}_n(x) - \tilde{\psi}_n(x)\right) + \left(\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x))\right) + \left(E(\tilde{\psi}_n(x)) - \psi(x)\right).$
- $\hat{v}_n(x) - v(x) = (\hat{v}_n(x) - \tilde{v}_n(x)) + (\tilde{v}_n(x) - E(\tilde{v}_n(x))) + (E(\tilde{v}_n(x)) - v(x)$

Théorème 4 : (Guessoum et Hamrani (2017))

Sous les hypothèses C1 et C3-C8, nous avons pour $n \to \infty$:

$$\sup_{x \in D} \left|\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x))\right| = O\left(\sqrt{\frac{\log n}{n h_n^d}}\right) \quad P\text{- p.s.}$$

Théorème 5 : (Guessoum et Hamrani (2017))

Sous les hypothèses C1, C3-C6 et C8, nous avons pour $n \to \infty$:

$$\sup_{x \in D} |\tilde{v}_n(x) - E(\tilde{v}_n(x))| = O\left(\sqrt{\frac{\log n}{n h_n^d}}\right) \quad P\text{- p.s.}$$

Convergence uniforme presque sûre de $\hat{m}_n$:

- Résultat principal :

Théorème 6 : (Guessoum et Hamrani (2017))

Sous les hypothèses C1-C8, nous avons pour $n \to \infty$:

$$\sup_{x \in D} |\hat{m}_n(x) - m(x)| = O\left\{ \sqrt{\frac{\log n}{n h_n^d}} \vee \left(\frac{\log \log n}{n} \right)^{\theta} \vee h_n^2 \right\} \text{ P- p.s.}$$

avec $0 < \theta < \frac{\gamma}{2\gamma + 6 + 3\kappa/2}$ pour tout réel $\kappa > 0$.

Quelques élements de preuves :

Preuve du Théorème 4 :

- $\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x)) = \frac{1}{nh_n^d} \sum_{i=1}^{n} Z_i(x)$ avec
  $Z_i(x) := \frac{\alpha Y_i}{G(Y_i)} K_d\left(\frac{x-X_i}{h_n}\right) - E\left(\frac{\alpha Y_i}{G(Y_i)} K_d\left(\frac{x-X_i}{h_n}\right)\right).$
- On couvre le compact D par un nombre fini p_n de boules $B_k(x_k, a_n^d)$.
- On utilise la décomposition suivante :

$$\begin{aligned}
\sup_{x\in D} \left|\tilde{\psi}_n(x) - E(\tilde{\psi}_n(x))\right| &\leq \max_{1\leq k\leq p_n} \sup_{x\in B_k} \frac{1}{nh_n^d} \sum_{i=1}^{n} |Z_i(x) - Z_i(x_k)| \\
&\quad + \max_{1\leq k\leq p_n} \frac{1}{nh_n^d} \left|\sum_{i=1}^{n} Z_i(x_k)\right| \\
&=: S_1 + S_2.
\end{aligned}$$

- $S_1 = O\left(\frac{1}{\sqrt{nh_n^d}}\right)$ (K_d est hölderien).

Quelques élements de preuves :

Preuve du Théorème 4 :(siute)

- Pour évaluer S_2, on utilise l'inégalité exponentielle de Doukhan et Neumann (2007).

Lemme 1 :

Sous les hypothéses C4-C6, il existe des constantes $K, M, L_1, L_2 < +\infty, \mu, \lambda \geq 0$ tel que pour tous $(s_1, \ldots, s_u) \in \mathbb{N}^u$ et tous $(t_1, \ldots, t_v) \in \mathbb{N}^v$ avec $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v \leq n$, nous avons

a) $\mathrm{Cov}\left(Z_{s_1} \cdots Z_{s_u}, Z_{t_1} \cdots Z_{t_v}\right) \leq K^2 M^{u+v-2} ((u+v)!)^{\lambda} uv \left(\rho(t_1 - s_u)\right)^{\frac{d}{2d+2}}$,

b) $\displaystyle\sum_{s=0}^{\infty} (s+1)^k (\rho(s))^{d/(2d+2)} \leq L_1 L_2^k (k!)^{\mu}, \forall k \geq 0$,

c) $E\left(|Z_i|^k\right) \leq (k!)^{\lambda} M^k, \forall k \geq 0$.

Quelques élements de preuves :

Preuve du Théorème 4 :(siute)

- L'inégalité exponentielle :

$$P\left(\sum_{i=1}^{n} Z_i(x_k) \geq \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2/2}{A_n + B_n^{1/(\mu+\lambda+2)}\varepsilon^{(2\mu+2\lambda+3)/(\mu+\lambda+2)}}\right)$$

où $A_n \leq \sigma_n^2$ avec $\sigma_n^2 := \mathrm{Var}\left(\sum_{i=1}^{n} Z_i(x)\right)$ et

$$B_n = 2cL_2\left(\frac{2^{4+\mu+\lambda} n c h_n^d L_1}{A_n} \vee 1\right).$$

- $S_2 = O\left(\sqrt{\frac{\log n}{n h_n^d}}\right).$

Quelques élements de preuves :

Preuve du Théorème 6 :

La preuve est basée sur la décomposition suivante :

$$\begin{aligned}
\sup_{x\in D}|\hat{m}_n(x)-m(x)| \leq & \frac{1}{\delta-\sup_{x\in D}|\hat{v}_n(x)-v(x)|}\Big\{\sup_{x\in D}|\hat{\psi}_n(x)-\tilde{\psi}_n(x)| \\
& +\sup_{x\in D}|\tilde{\psi}_n(x)-E(\tilde{\psi}_n(x))|+\sup_{x\in D}|E(\tilde{\psi}_n(x))-\psi(x)| \\
& +\delta^{-1}\sup_{x\in D}\left|\psi(x)\right|\sup_{x\in D}|\hat{v}_n(x)-v(x)|\Big\}.
\end{aligned}$$

Les lemmes techniques :

Lemme 2 :

Sous les hypothèses C2, C4 et C5, nous avons

$$\sup_{x\in D}\left|\hat{\psi}_n(x)-\tilde{\psi}_n(x)\right| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \text{ P- p.s quand } n \to +\infty.$$

Lemme 3 :

Sous les hypothéses C3, C4 et C6, nous avons

$$\sup_{x\in D}\left|E(\tilde{\psi}_n(x))-\psi(x)\right| = O(h_n^2) \text{ p.s quand } n \to \infty$$

Lemme 4 :

Sous les hypothéses C2-C5 et C8, nous avons

$$\sup_{x\in D}\left|\hat{v}_n(x)-v(x)\right| = O\left\{\sqrt{\frac{\log n}{nh_n^d}} \vee \left(\frac{\log\log n}{n}\right)^{\theta} \vee h_n^2\right\} \text{P - p.s quand } n \to \infty$$

Les lemmes techniques :

Lemme 2 :

Sous les hypothèses C2, C4 et C5, nous avons

$$\sup_{x\in D}\left|\hat{\psi}_n(x)-\tilde{\psi}_n(x)\right| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \text{ P- p.s quand } n \to +\infty.$$

Lemme 3 :

Sous les hypothéses C3, C4 et C6, nous avons

$$\sup_{x\in D}\left|E(\tilde{\psi}_n(x))-\psi(x)\right| = O(h_n^2) \text{ p.s quand } n \to \infty$$

Lemme 4 :

Sous les hypothéses C2-C5 et C8, nous avons

$$\sup_{x\in D}\left|\hat{v}_n(x)-v(x)\right| = O\left\{\sqrt{\frac{\log n}{nh_n^d}} \vee \left(\frac{\log\log n}{n}\right)^{\theta} \vee h_n^2\right\} \text{P - p.s quand } n \to \infty$$

Les lemmes techniques :

Lemme 2 :

Sous les hypothèses C2, C4 et C5, nous avons

$$\sup_{x\in D}\left|\hat{\psi}_n(x)-\tilde{\psi}_n(x)\right| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \text{ P- p.s quand } n\to+\infty.$$

Lemme 3 :

Sous les hypothéses C3, C4 et C6, nous avons

$$\sup_{x\in D}\left|E(\tilde{\psi}_n(x))-\psi(x)\right| = O(h_n^2) \text{ p.s quand } n\to\infty$$

Lemme 4 :

Sous les hypothéses C2-C5 et C8, nous avons

$$\sup_{x\in D}\left|\hat{v}_n(x)-v(x)\right| = O\left\{\sqrt{\frac{\log n}{nh_n^d}}\vee\left(\frac{\log\log n}{n}\right)^{\theta}\vee h_n^2\right\} \text{P - p.s quand } n\to\infty$$

Les lemmes techniques :

Lemme 2 :

Sous les hypothèses C2, C4 et C5, nous avons

$$\sup_{x\in D}\left|\hat{\psi}_n(x)-\tilde{\psi}_n(x)\right| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \text{ P- p.s quand } n \to +\infty.$$

Lemme 3 :

Sous les hypothéses C3, C4 et C6, nous avons

$$\sup_{x\in D}|E(\tilde{\psi}_n(x))-\psi(x)| = O(h_n^2) \text{ p.s quand } n \to \infty$$

Lemme 4 :

Sous les hypothéses C2-C5 et C8, nous avons

$$\sup_{x\in D}|\hat{v}_n(x)-v(x)| = O\left\{\sqrt{\frac{\log n}{nh_n^d}} \vee \left(\frac{\log\log n}{n}\right)^{\theta} \vee h_n^2\right\} \text{P - p.s quand } n \to \infty$$

Simulation :

- Modèle linéaire associé :

$$Y_i = 2X_i + 1 + \varepsilon_i, \; i = 1, \ldots, N,$$

où

  - $X_i = \exp\left[\frac{1}{2}(W_{i-1} + W_{i-2})\right]$,
  - W_i; $i = -1, 0, \cdots, N-1$ sont $N+1$ v.a's iid $\sim \mathcal{N}(0,1)$,
  - ε_i; $i = 1, \cdots, N$ sont N v.a's iid $\sim \mathcal{N}(0, 0.2)$.

- N iid v.a's $T_i \sim \mathcal{N}(m, 1)$,
- Garder $\{(X_i, Y_i, T_i);\; i = 1, \cdots, n\}$ tel que $Y_i \geq T_i$,
- Calculer l'estimateur $\hat{m}_n(x)$ de $m(x) = 2x + 1$ pour $x \in [0, 2]$.

Simulation : α fixé et n variant

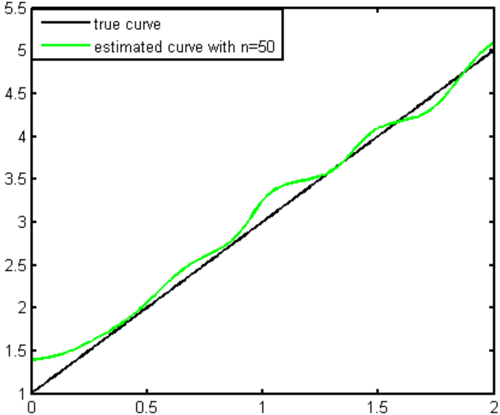

Figure: Comparaison entre m(x) (noir) et son estimée $m_n(x)$ pour $n = 50$ (vert) avec $\alpha \approx 80\%$.

Simulation : α fixé et n variant

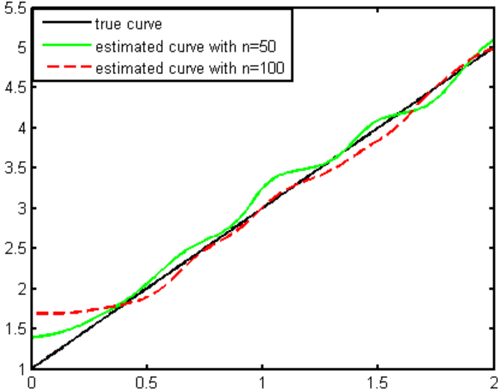

Figure: Comparaison entre m(x)(noir), son estimée $m_n(x)$ pour $n = 50$ (vert) et pour n=100 (rouge) avec $\alpha \approx 80\%$.

Simulation : α fixé et n variant

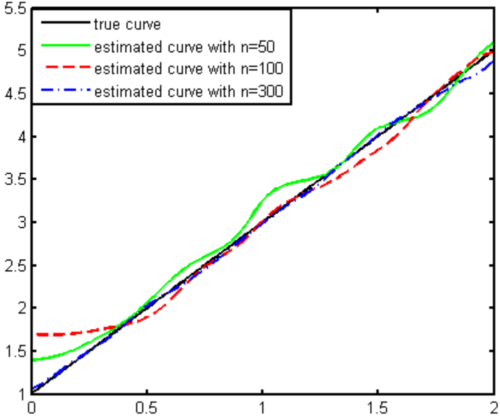

Figure: Comparaison entre m(x)(noir), son estimée $m_n(x)$ pour $n = 50$ (vert), pour n=100 (rouge) et pour n=300 (bleu) avec $\alpha \approx 80\%$.

Simulation : n fixé et α variant

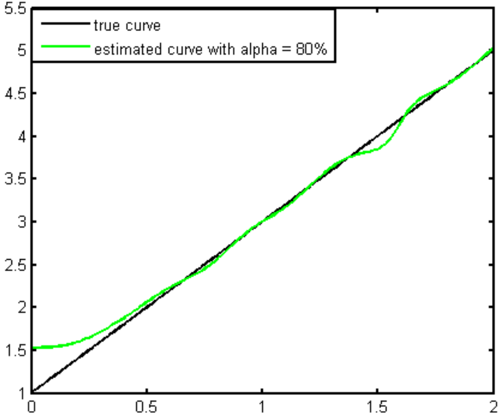

Figure: Comparaison entre m(x) (noir) et son estimée $m_n(x)$ pour $\alpha \approx 80\%$(vert)avec n=50.

Simulation : n fixé et α variant

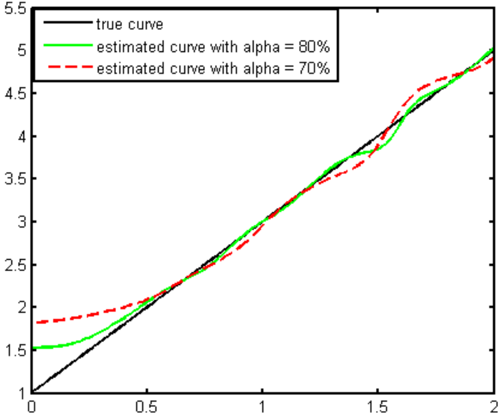

Figure: Comparaison entre m(x) (noir) et son estimée $m_n(x)$ pour $\alpha \approx 80\%$ (vert) et pour $\alpha \approx 70\%$ (rouge) avec $n = 50$.

Simulation : n fixé et α variant

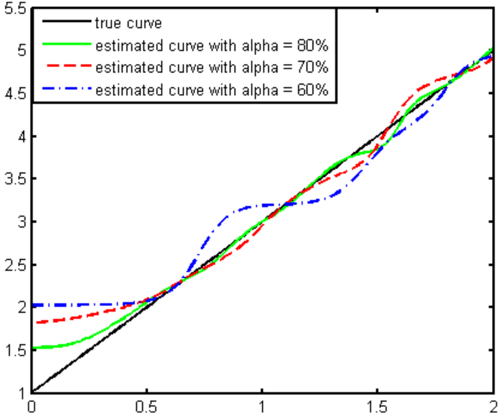

Figure: Comparaison entre m(x) (noir) et son estimée $m_n(x)$ pour $\alpha \approx 80\%$ (vert), pour $\alpha \approx 70\%$(rouge) et pour $\alpha \approx 60\%$ (blue) avec $n = 50$.

Simulation : n fixé et α variant

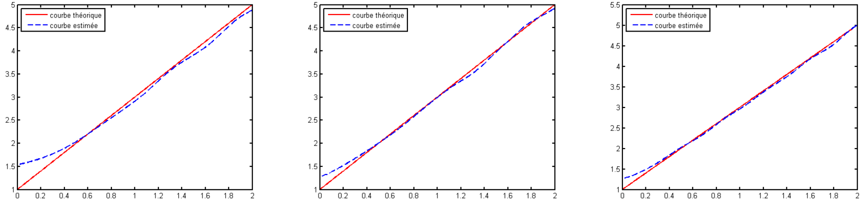

Figure: $m(.)$ and $\hat{m}_n(.)$ avec $n = 300$ et $\alpha \approx 60, 70$ et 80%.

Simulation :

La table suivante donne, dans chaque cas, la médiane des erreurs quadratiques moyenne (MSE) pour $x \in [0, 2]$ après 1000 réplications de l'estimateur.

Table: La mediane des erreurs quadratiques moyennes de $\hat{m}_n$.

$\alpha(\%)$	n=50	n=100	n=300
60	0.0033	0.0012	3.0539×10^{-4}
70	0.0028	0.0009	3.0283×10^{-4}
80	0.0027	0.0006	2.7858×10^{-4}

Simulation :

La table suivante donne, dans chaque cas, les valeurs de MISE pour $x \in [0, 2]$ après 1000 réplications de l'estimateur.

Table: les erreurs quadratiques moyennes intégrées de $\hat{m}_n$.

$\alpha(\%)$	n=50	n=100	n=300
60	0.0539	0.0305	0.0228
70	0.0341	0.0284	0.0128
80	0.0298	0.0198	0.0127

Simulation :

Modèle non linéaire :

$$
\begin{aligned}
Y_i &= \exp(X_i) + \varepsilon_i \\
Y_i &= \sin(\pi X_i + \frac{1}{2}) + \varepsilon_i \\
Y_i &= X_i^2 + X_i + 1 + \varepsilon_i
\end{aligned}
$$

Simulation :

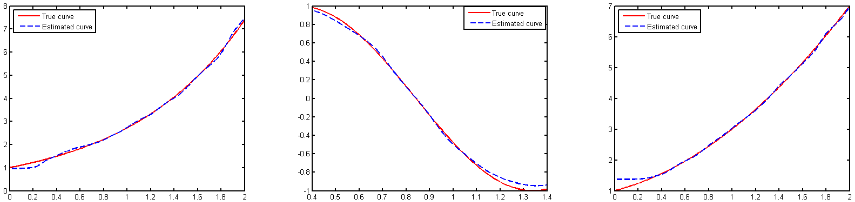

Figure: Le cas exponentiel, sinus et parabolique pour $n = 300$ et $\alpha \approx 80\%$

Normalité asymptotique de $\hat{m}_n$:

Hypothèses pour la normalité asymptotique :

D1. $a_G < a_F$ et $b_G \leq b_F$. D2. $\displaystyle\int_{\mathbb{R}} \frac{dF(z)}{G^2(z)} < +\infty$.

D3. $nh_n^{d+4} \to 0$ quand $n \to +\infty$,
$nh_n^{d/\tau}(\log\log n)^{1/\tau - 1} \to 0$ quand $n \to +\infty$, pour certain $0 < \tau < 1$

D4. K_d est un noyau d'ordre 2 à support compact et admet des dérivées partielles d'ordre 1 bornées .

D5. Le terme de covariance définit par
$\rho(s) := \sup_{|i-j| \geq s} \Delta_{i,j}, s > 0$, vérifie $\rho(s) \leq \gamma_0 e^{-\gamma s}$, $\gamma_0 > 0$, $\gamma > 0$.

D6. La fonction $\psi(.)$ et la densité $v(.)$ sont bornées, deux fois differentiable et à dérivées partielles bornées.

D7. La densité conjointe $v^*_{i,j}$ de (X_i, X_j) est bornée.

D8. Il existe des suites de nombres entiers $(p_n)_n$, $(q_n)_n$ et $(k_n)_n$ définie par $k_n := \left[\frac{n}{p_n + q_n}\right]$, tendant vers ∞ quand n tend vers ∞, telles que
$k_n(p_n + q_n) \leq n$ et $\frac{k_n(p_n+q_n)}{n} \to 1$ satisfaisant :
$\frac{p_n k_n}{n} \to 1$, $p_n h_n^d \to 0$, $\frac{p_n^2}{nh_n^d} \to 0$ et $\frac{e^{-\gamma q_n}}{h_n^{d+2}} \to 0$.

Normalité asymptotique de $\hat{m}_n$:

Théorème 7 : (Guessoum et Hamrani (soumis))

Sous les hypothèses D1-D8, nous avons pour tout x tel que $v(x) > 0$:

$$\sqrt{nh_n^d}\,[\hat{m}_n(x) - m(x)] \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(x)\right),$$

où

$$\sigma^2(x) := \frac{\alpha\left[\psi_0(x)v^2(x) - 2\psi_1(x)\psi(x)v(x) + \psi_2(x)\psi^2(x)\right]}{v^4(x)}\kappa,$$

$\kappa = \int_{\mathbb{R}^d} K_d^2(z)dz < +\infty$ et $\psi_j(x) = \int_{\mathbb{R}} \frac{y^{2-j}}{G(y)} f(x,y)dy, j = 0, 1, 2.$

Estimateur de $\sigma^2(x)$:

Remarque

Un estimateur de type plug-in $\hat{\sigma}_n^2(x)$ pour la variance asymptotique $\sigma^2(x)$ peut être obtenu en utilisant les estimateurs $\alpha_n, \hat{v}_n(.)$ et les estimateurs

$$\hat{\psi}_{j,n}(x) := \frac{\alpha_n}{nh_n^d} \sum_{i=1}^{n} \frac{Y_i^{2-j}}{G_n^2(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)$$

de $\psi_j(.), j = 0, 1, 2$.

Intervalle de confiance :

- Intervalle de confiance :

Corollaire

Sous les hypothèses de Théorème précédent, nous obtenons pour tout $\xi \in (0,1)$, l'intervalle de confiance suivant de niveau asymptotique $1-\xi$ pour $m(x)$

$$\left[\hat{m}_n(x) - \frac{u_{1-\xi/2}\hat{\sigma}_n(x)}{\sqrt{nh_n^d}}, \hat{m}_n(x) + \frac{u_{1-\xi/2}\hat{\sigma}_n(x)}{\sqrt{nh_n^d}}\right],$$

où $u_{1-\xi/2}$ dénote le quantile d'ordre $1-\xi/2$ de la loi normale centrée réduite.

Simulation :

- Modèle associé :

$$X_i = (W_{i-1} + W_{i-2})/2,$$
$$Y_i = 2X_i + 1 + \varepsilon_i, \ i = 1, \dots, N,$$

où

  - W_i; $i = -1, 0, \cdots, N-1$ sont $N+1$ v.a's iid $\sim \mathcal{E}xp(1)$,
  - ε_i; $i = 1, \cdots, N$ sont N v.a's iid $\sim \mathcal{N}(0, 0.2)$.
- N iid v.a's $T_i \sim \mathcal{E}xp(\lambda)$,
- Garder $\{(X_i, Y_i, T_i);\ i = 1, \cdots, n\}$ tel que $Y_i \geq T_i$.
- Calculer l'estimateur $\hat{m}_n(x)$ de $m(x) = 2x + 1$ et $\hat{\sigma}_n^2(x)$ pour $x = 0.5$.
- Calculer l'écart normalisé entre $\hat{m}_n(0.5)$ et $m(0.5)$:

$$\dot{m}_n = \dot{m}_n(0.5) := \frac{\sqrt{nh_n}}{\hat{\sigma}_n(0.5)}(\hat{m}_n(0.5) - m(0.5)) = \frac{\sqrt{nh_n}}{\hat{\sigma}_n(0.5)}(\hat{m}_n(0.5) - 2).$$

Simulation :

- Nous générons, en utilisant cette procédure, $\mathrm{B} = 1000$ suites observées de taille n.
  $\hookrightarrow$ Ceci donne une suite de v.a.'s i.i.d.

  $$\dot{\mathrm{m}}_1, \ldots, \dot{\mathrm{m}}_\mathrm{B}.$$

  Nous estimons ensuite sa densité en utilisant la méthode à noyau.

Simulation :

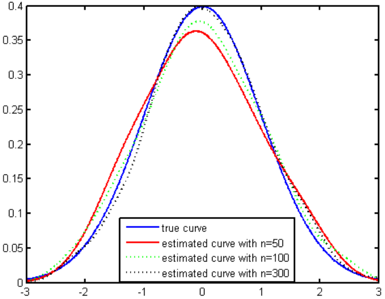

Figure : $\mathrm{n} = 50, 100, 300$ et $\alpha = 0.8$

Simulation :

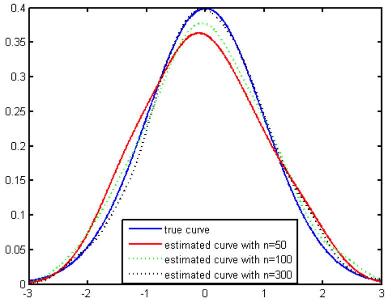

Figure : $\mathrm{n} = 50, 100, 300$ et $\alpha = 0.8$

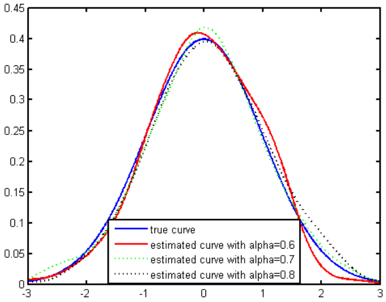

Figure : $\mathrm{n} = 200$ et $\alpha = 0.6, 0.7, 0.8$

Simulation :

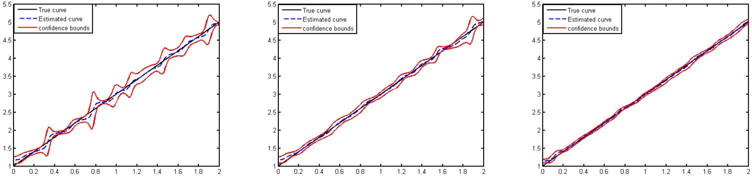

Figure: $\alpha \approx 70\%$, $\mathrm{n} = 50$, $\mathrm{n} = 100$ et $\mathrm{n} = 300$

Simulation :

Enfin, nous nous sommes intéressé à la probabilité de recouvrement (coverage probability) pour les intervalles de confiance de niveau asymptotique 95%. Nous avons pris $x \in [0, 2]$ et effectué 1000 replications de taille n. Nous avons obtenu les résultats suivants :

Table: Les probabilités de couverture de l'intervalle de confiance à 95%.

	n=50	n=100	n=300
$\alpha \approx 70\%$	0.9249	0.9311	0.9551
$\alpha \approx 90\%$	0.9307	0.9389	0.9556

Conclusion et Perspectives

Conclusion et perspectives

Ce qui est fait

- Nous avons donné une caractérisation de la vitesse de convergence uniforme presque sûre de l'estimateur à noyau de la fonction de régression dans le cas de données tronquées à gauche et associées.
- Nous avons établi sa normalité asymptotique.

Perspectives

- Étendre nos résultats au cas d'une covariable X fonctionnelle.
- Ètendre nos résultats aux données faiblement dépendantes.
- Établir un résultat de type Berry Esseen pour l'estimateur à noyau de la fonction de régression pour quantifier la vitesse avec laquelle s'effectue la convergence vers la loi normale.
- ....

Conclusion et perspectives

Ce qui est fait

- Nous avons donné une caractérisation de la vitesse de convergence uniforme presque sûre de l'estimateur à noyau de la fonction de régression dans le cas de données tronquées à gauche et associées.
- Nous avons établi sa normalité asymptotique.

Perspectives

- Étendre nos résultats au cas d'une covariable X fonctionnelle.
- Ètendre nos résultats aux données faiblement dépendantes.
- Établir un résultat de type Berry Esseen pour l'estimateur à noyau de la fonction de régression pour quantifier la vitesse avec laquelle s'effectue la convergence vers la loi normale.
- ....

Je vous remercie