Je tiens à exprimer ma profonde reconnaissance à Monsieur Zaher MOHDEB , Professeur à l'université MENTOURI - Constantine d'avoir dirigé cette thèse. Il m'a apporté un grand soutien moral par sa grande disponibilité, ses encouragements et ses relations humaines. Pour tout cela, je le prie d'accepter mes sincères remerciements.

Je remercie très vivement Monsieur Nasser Eddine BENKAFADER, Professeur à l'université MENTOURI - Constantine, pour l'honneur qu'il me fait pour avoir accepter de présider ce jury.

Mes vifs remerciements vont aussi à Monsieur Hacène BOUTABIA, Professeur à l'université BADJI MOKHTAR - Annaba, Abdelhakim NECIR , Professeur à l'université MOHAMED KHIDER -Biskra , Lazhar Fouad RAHMANI , Maître de conférence à l'université MENTOURI- Constantine et Madame Nacera SEDDIK-AMER, Maître de conférence à l'université BADJI MOKHTAR - Annaba en participant à ce jury.

Je tiens à exprimer ma gratitude à Monsieur Paul DEHEUVELS , Professeur à l'université PIERRE ET MARIE CURIE - PARIS pour toute l'aide dont il a fait preuve à mon égard.

Mes remerciements vont aussi à Monsieur Djamel LOUANI , Professeur à l'université PIERRE ET MARIE CURIE - PARIS pour son aide.

Table des matières

INTRODUCTION

Considérons un modèle statistique $(E, \mathcal{A}, P)$, où E est un ensemble appelé espace des observations. $\mathcal{A}$ est une tribu sur E. P une famille de mesure de probabilité sur $(E, \mathcal{A})$ et qui peut toujours s'écrire $(P_\theta,\ \theta \in \Theta)$. Soit g une application de P dans un ensemble quelconque Θ' et une suite de variables aléatoires $Z = (Z_n)_{n\in\mathbb{N}}$ à valeurs dans E dont la loi de probabilité inconnue $\mathbf{P}$ appartenant à P. Estimer $g(\mathbf{P})$, c'est chercher à l'évaluer lorsque l'on observe les réalisations $\{Z_1, \ldots, Z_n\}$. La règle choisie pour procéder à cette estimation est une application de E dans Θ' ; on dira que c'est un estimateur du paramètre $g(\mathbf{P})$. Lorsque Θ est vaste, on dit que le modèle $(E, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$ est non paramétrique. Il n'y a pas de définition précise d'un modèle non paramétrique car la frontière entre paramétrique et non paramétrique est assez floue. Cependant, d'après Bosq et Lecoutre (1987), le modèle est non paramétrique si Θ contient un ensemble convexe de dimension infinie et qu'il est paramétrique quand Θ est un ouvert de $\mathbb{R}^s$.

Voici quelques exemples de paramètres fonctionnels : la fonction de répartition, la fonction caractéristique, pour $s = 1$, la fonction quantile, la densité et ses dérivées, les paramètres de régression, la densité spectrale, le taux de panne.

Sous des hypothèses de régularité et lorsque Θ est un compact de $\mathbb{R}^s$, le maximum de vraisemblance

$$\theta_n = Arg\max V(\theta; Z_1, \ldots, Z_n), \qquad V(\theta; Z_1, \ldots, Z_n) = \prod_{k=1}^{n} \frac{dP_\theta}{d\lambda}(Z_k)$$

satisfait le théorème central limite

$$\mathcal{L}\left[\sqrt{n}(\theta_n - \theta)\right] \longrightarrow \mathcal{N}(0, \sigma^2), \qquad n \to \infty .$$

De plus la variance σ^2 ainsi déterminée est minimale d'après le théorème de Cramer-Rao. Lorsque Θ est une partie d'un espace fonctionnel, la situation n'est pas aussi simple. Des résultats généraux montrent que pour une fonction de perte d donnée sur Θ, il existe une suite v_n et une constante $c > 0$ vérifiant

$$\frac{1}{c} \leq \inf_{\theta_n} \sup_{\theta} d\left(\frac{\theta}{v_n}, \frac{\theta_n}{v_n}\right) \leq c .$$

Ces vitesses de convergence sont considérées comme des vitesses étalon, elles nous permettent de dire que les vitesses de convergence obtenues sont satisfaisantes. La vitesse de convergence des estimateurs d'une fonction de θ dépend de la structure de l'espace Θ. Dans notre travail, nous considérons les cas suivants

$\theta = f$ est la densité par rapport à la mesure de Lebesgue de la loi marginale de la suite i.i.d. à valeurs réelles $X = (X_n)_{n \in \mathbb{N}}$.

$\theta = r$ est la régression $r(x) = E(Y_k/X_k = x)$ lorsque la suite i.i.d. à valeurs dans $\mathbb{R}^2$, et $Z_n = (X_n, Y_n)$, dans le cas où la suite $X = (X_n)_{n \in \mathbb{N}}$ est i.i.d. à valeurs réelles.

Grâce à de nombreux travaux, la théorie de l'estimation non paramétrique de la densité et de la régression a vu son domaine d'application croître considérablement et est devenue un domaine de recherche trés actif, citons quelques références, sans être exhaustif : Prakassa Rao (1983), Devroye et Györfi (1985), Silverman (1986), Devroye (1978), Nadaraya (1989), Roussas (1990), Härdle (1990), Scott (1992), Bosq et Lecoutre (1987), Wand et Jones (1995). Le comportement des estimateurs non paramétriques décrit dans ces travaux dépend fortement du paramètre de lissage. Plusieurs méthodes ont été proposées décrivant le choix de ce paramètre.

Récemment Deheuvels et Mason (2004) ont obtenu des bandes de confiance uniformes et non uniformes pour les fonctionnelles de la distribution. En s'inspirant de ce travail et pour un choix approprié de l'estimation de la fenêtre, nous obtenons des bandes de confiance asymptotiques basées sur les estimateurs à noyau de type Nadaraya-Watson pour la fonction de régression et de type Akaike-Parsen-Rosenblatt pour la densité de probabilité, dans le cas particulier de la loi normale centrée, d'écart type inconnu.

A partir du calcul de $h_n(x)$, $x \in I$, (window-width de l'estimateur à noyau) optimisant les critères d'erreurs, le but de notre travail consiste à construire leurs estimateurs $\widehat{h}_n(x)$ en fonction des estimateurs empiriques des paramètres inconnus, tels que

$$\frac{\acute{h}_n(x)}{h_n(x)} \xrightarrow{P} 1 \quad n \to \infty .$$

$\widehat{h}_n(x)$ ainsi choisi, est tel que les conditions imposées par Deheuvels et Mason (2004) soient vérifiées, nous déduisons ainsi des bandes de confiance asymptotiques de $f(x)$ (respectivement de $r(x)$) en fonction de $f_{X,n}\left(x;\widehat{h}_n(x)\right)$ (respectivement en fonction de $r_n(x;\widehat{h}_n(x))$).

Dans la deuxième partie, il s'agit de caractériser les distributions F telles que les espacements $S_{k,n}$ soient asymptotiquement indépendants. Beaucoup de recherches sur le k^{ieme} maximum de $S_{k,n}$ ont été faites (cf. Pyke (1965), Devroye (1982), Deheuvels (1982 ; 1983) et les références citées dans ces publications).

Le lien entre les deux parties concerne surtout l'influence des statistiques d'ordre dans l'estimation de la densité par la méthode de l'histogramme avec des partitions aléatoires.

Les deux premiers chapitres sont consacrés à la présentation et à la formulation du problème de la recherche optimale de l'estimation non paramétrique par la méthode du noyau et les propriétés asymptotiques du processus empirique permettant de dégager le principe des méthodes de l'estimation fonctionnelle.

Le troisième chapitre consiste à construire les fenêtres minimisant les critères d'erreurs dans le cas gaussien en fonction des estimateurs empiriques tels que les conditions imposées par Deheuvels et Mason (2004) soient vérifiées et nous donnons aussi des bandes de confiance asymptotiques pour les fonctionnelles de la distribution.

Le quatrième chapitre est consacré aux statistiques d'ordre. Nous exposons d'abord les premiers résultats concernant l'étude asymptotique de la variable extrême, la k^{iem} statistique d'ordre et les espacements. Nous nous basons sur les théorèmes de Hall (1978), Weisman (1978) et de Deheuvels (1986) pour établir que l'indépendance asymptotique des espacements associée à la compacité stochastique ou à l'existence du type limite de la variable extrême, caractérise les distributions attirées par la loi de Gumble. Nous démontrons aussi que sous certaines conditions, cette indépendance asymptotique ne caractérise pas nécessairement les lois dont les extrêmes sont attirées par Gumble.

Nous terminons cette partie par deux exemples illustrant le rôle des statistiques d'ordre dans l'estimation statistique.

Chapitre I

ESTIMATION DE LA DENSITÉ ET DE LA RÉGRESSION

1.1 Rappels

La théorie de l'estimation statistique est une branche fondamentale des statistiques mathématiques. Cette théorie est subdivisée en estimation paramétrique et non paramétrique. L'approche classique de l'estimation de la densité à partir d'un échantillon $X_1, X_2, \ldots, X_n$, est la méthode paramétrique. Elle consiste à supposer que la loi des observations appartient à un ensemble de lois indéxé par un ou plusieurs paramètres à partir des données. Cependant cette méthode peut, si le modèle choisi n'est pas raisonnable pour le phénomène étudié, conduire à des résultats erronés. Pour cela il est important de considérer la méthode non paramétrique. Ces dernières décennies, cette théorie a fait l'objet de plusieurs travaux, pour l'estimation de la densité et de la régression (voir Prakassa Rao (1983), Roussas (1990), Härdle (1990), Scott (1992), Bosq et Lecoutre (1987), Wand et Jones (1995) et les références citées dans ces publications).

Soient $(X_1, Y_1), (X_2, Y_2), \ldots$, une suite de vecteurs aléatoires indépendants, continus à valeurs dans $\mathbb{R}^2$ de même loi que le vecteur (X, Y) dont la densité est $f(x, y)$. Dans ce modèle, nous supposons que pour tout $x \in \mathbb{R}$, la densité marginale de X, la régression de Y sachant $X = x$ et la variance de Y sachant $X = x$ sont bien définies, et elles sont respectivement données par :

- $f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\, dx.$

- $r(x) = E(Y \mid X = x) = \int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_X(x)} dx.$

- $v(x) = Var(Y \mid X = x) = \frac{1}{f_X(x)} \int_{-\infty}^{+\infty} (y - r(x))^2 f(x, y)\, dy.$

1.2 Estimation de la densité

L'estimateur le plus étudié et le plus utilisé (voir Silverman (1986) pour plus de détails sur cet estimateur) est l'estimateur à noyau fixe, proposé par Rosenblatt (1956) et Parzen (1962) et défini par

$$f_{X,n}(x;\ h_n) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{1.1}$$

où

• h_n est une suite de réels positifs tendant vers zéro, quand $n \to \infty$

et

• K est une fonction mesurable vérifiant :

$(K1)$ $\int_{-\infty}^{+\infty} K(x)\,dx = 1;$

$(K2)$ $\sup_x K(x) \leq M < \infty;$

$(K3)$ $|x|\,K(x) \to 0$ quand $|x| \to \infty;$

$(K4)$ $\forall x \in \mathbb{R},\ K(x) = K(-x)\,;$

$(K5)$ $\int_{-\infty}^{+\infty} x^2 K(x)\,dx < \infty\,;$

$(K6)$ la fonction caractéristique de $K(.)$ est absolument intégrable.

Exemples : Les fonctions de densités de probabilité $K(.)$ citées ci dessous vérifient les conditions $(K.1 - K.6)$.

(i) $\qquad K_1(x) = \begin{cases} \frac{1}{2} & \text{si} \quad |x| \leq 1, \\ 0 & \text{si} \quad |x| > 1; \end{cases}$

(ii) $$K_2(x)=\begin{cases}1-|x| & \text{si} \quad |x|\leq 1,\\ 0 & \text{si} \quad |x|>1;\end{cases}$$

(iii) $$K_3(x)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}x^2\right), \quad x\in\mathbb{R};$$

(iv) $$K_4(x)=\left(\pi\left(1+x\right)^2\right)^{-1}, \quad x\in\mathbb{R};$$

(v) $$K_5(x)=\begin{cases}\frac{1}{2\pi}\left(\frac{\sin\left(\frac{x}{2}\right)}{\frac{x}{2}}\right)^2 & \text{si} \quad x\neq 0,\\ \frac{1}{2\pi} & \text{si} \quad x=0;\end{cases}$$

(vi) $$K_6(x)=\begin{cases}\frac{3}{4}\lambda^{-3}\left(\lambda^2-x^2\right) & \text{si} \quad x^2\leq\lambda^2,\\ 0 & \text{si} \quad x^2>\lambda^2, \ \lambda>0\end{cases}$$

Pour étudier les propriétés de cet estimateur, nous énonçons les définitions suivantes.

Définitions 1.1 *On dit que la suite des estimateurs de la densité*

1 $f_{X,n}(x;\ h_n)$ *est faiblement consistante si*

$$\forall x\in\mathbb{R},\quad f_{X,n}(x;h_n)\xrightarrow{P} f_X(x)\,,\ \text{quand } n\to\infty.$$

2 $f_{X,n}(x;\ h_n)$ *est faiblement et uniformément consistante si*

$$\sup_{x\in\mathbb{R}}|f_{X,n}(x;h_n)-f_X(x)|\xrightarrow{P}0,\ \text{quand } n\to\infty\ .$$

3 $f_{X,n}(x;\ h_n)$ *est fortement consistante si*

$$\forall x\in\mathbb{R},\quad f_{X,n}(x;h_n)\xrightarrow{P.S} f_X(x)\,,\ \text{quand } n\to\infty.$$

4 $f_{X,n}(x;\ h_n)$ *est fortement et uniformément consistante si*

$$\sup_{x\in\mathbb{R}} |\ f_{X,n}(x;h_n) - f_X(x)| \xrightarrow{P.S} 0, \text{ quand } n \to \infty.$$

5 $f_{X,n}(x;\ h_n)$ *est asymptotiquement sans biais si*

$$\forall x \in \mathbb{R},\ \lim_{n\to\infty} E(\ f_{X,n}(x;h_n)) = f_X(x).$$

6 $f_{X,n}(x;\ h_n)$ *est asymptotiquement et uniformément sans biais si*

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}} |E(f_{X,n}(x;h_n)) - f_X(x)| = 0.$$

L'application de ces définitions se résument en deux théorèmes dus à Parzen (1962) et Nadaraya (1964).

Parzen (1962) a établi la consistance faible de $f_{X,n}(x;h_n)$ en se basant sur le Théorème de Bochner (1955). Son travail est un outil important qui a permis à plusieurs chercheurs de développer le domaine de la consistance faible de l'estimateur $f_{X,n}(x;\ h_n)$ (cf Devroye et Györfi (1985), Silverman (1986), Izenman (1991), Scott (1992)).

Après un travail inédit, pionnier de Parthasarathy (1969) sur la consistance uniforme de $f_{X,n}(x;\ h_n)$, différentes approches ont été considérées pour ce problème, (cf Nadaraya (1976) et Schuster (1970), Bosq et Lecoutre (1987)).

Théorème 1.1. (Bochner (1955)). *Soient* $(\mathbb{R}^m,\ \beta_m)$ *un espace euclidien de dimension* m, *muni d'une tribu Borelienne;* $g,\ k : (\mathbb{R}^m,\ \beta_m) \longrightarrow (\mathbb{R},\ \beta_1)$ *des fonctions mesurables telles que,*

$\forall\ z \in \mathbb{R}^m$

$$|k(z)| \leq M, \int_{\mathbb{R}^m} |k(z)|\ dz < \infty, \int_{\mathbb{R}^m} |g(z)|\ dz < \infty. \tag{1.2}$$

$$\|z\|^m |k(z)| \longrightarrow 0, \text{ quand } \|z\| \longrightarrow \infty,$$

où $\|.\|$ est la norme euclidienne.

Si g est continue $0 < h_n \longrightarrow 0$, quand $n \longrightarrow \infty$, alors

$$\lim_{n\to\infty} \frac{1}{h_n^m} \int_{\mathbb{R}^m} k\left(\frac{z}{h_n}\right) g(x-z)\,dz = g(x) \int_{\mathbb{R}^m} k(z)\,dz \tag{1.3}$$

et si de plus g est uniformément continue alors la convergence dans (1.3) est uniforme.

Théorème 1.2. (Parzen (1962)). *Si $(K.1 - K.6)$ sont vérifiées et si $nh_n \to \infty$, $h_n \to 0$,* quand n$\to \infty$, alors

$$\forall x \in \mathbb{R}, \quad f_{X,n}(x; h_n) \xrightarrow{P} f_X(x).$$

Preuve

$$\begin{aligned} E\left(f_{X,n}(x; h_n)\right) &= E\left[\frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} E\left[K\left(\frac{x-X}{h_n}\right)\right] \\ &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right) f_X(t)\,dt . \end{aligned}$$

Comme f, K, h_n vérifient les conditions de Bochner (1955), on a alors,

$$\lim_{n\to\infty} E\left(f_{X,n}(x; h_n)\right) = f_X(x).$$

D'autre part, comme les X_i sont indépendantes et identiquement distribuées, il vient que

$$Var f_{X,n}(x; h_n) \sim \frac{1}{nh_n} f_X(x) \int_{-\infty}^{+\infty} K^2(t)\,dt.$$

Ce qui implique que

$$\lim_{n\to\infty} Var\left[f_{X,n}(x; h_n)\right] = 0.$$

Théorème 1.3. *Supposons que $K(.)$ est à variation bornée et que pour tout $\gamma > 0$, la série*

$\sum_{n=1}^{\infty} \exp\left(-\gamma n h_n^2\right)$ *converge. Alors*

$$v_n = \sup_{x \in \mathbb{R}} \left| f_{X,n}(x; h_n) - f_X(x) \right| \xrightarrow{P} 0, \text{ quand } n \to \infty,$$

si et seulement si la densité f_X est uniformément continue.

La démonstration de ce Théorème (voir Nadaraya (1964)) est basée sur le résultat de Dvoretzky (1956) et sur les lemmes suivants.

Lemme 1.1. $\forall n > 0$, $\forall \varepsilon_n > 0$, il existe $C > 0$ tel que

$$P\left[\sup_{-\infty < x < \infty} \left| F_n(x) - F(x) \right| > \varepsilon_n\right] \leq C \exp\left(-2n\varepsilon_n^2\right),$$

où F et F_n désignent respectivement la fonction de répartition, la distribution empirique de $X_1, X_2, \ldots, X_n$.

Lemme 1.2. *Pour toute fonction g, on a*

$$\lim_{n \to \infty} \sup_{x \in \mathbb{R}} \left| f_{X,n}(x; h_n) - g(x) \right| \xrightarrow{P} 0,$$

si et seulement si

$$\lim_{n \to \infty} \sup_{x \in \mathbb{R}} \left| E f_{X,n}(x; h_n) - g(x) \right| = 0.$$

Lemme 1.3. *Si*

$$\lim_{n \to \infty} \sup_{x \in \mathbb{R}} \left| f_{X,n}(x; h_n) - g(x) \right| = 0, \text{ P.S}$$

alors g est uniformément continue.

1.3 Estimation de la régression

L'estimateur non paramétrique de la régression est introduit et étudié par Nadaraya (1964) et Watson (1964). Cette estimateur a vu son domaine d'application croitre de plus en plus et beaucoup de résultats importants ont été obtenus en ex : U.R.S.S aussi bien qu'ailleurs (cf Nadaraya (1989), Wand et Jones (1995), Watson (1964)).

Nous rapellons que l'estimateur de $r(x)$ peut aussi utiliser les estimateurs de $f(x,y)$ et de $f_X(x)$.

Soient $J(x,y)$ une densité de probabilité sur $\mathbb{R}^2$, et $\forall x \in \mathbb{R}$, $J_1(x) = \int_{-\infty}^{+\infty} J(x,y)\,dy$, positive.

Si $h_n \to 0$, quand $n \to \infty$, alors

$$f_n(x,y) = \frac{1}{nh_n}\sum_{i=1}^{n} J\left(\frac{x-X_i}{h_n}, \frac{y-Y_i}{h_n}\right),$$

$$g_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} J_1\left(\frac{x-X_i}{h_n}\right),$$

peuvent être considérées respectivement, comme les estimateurs de $f(x,y)$, de $f_X(x)$ et

$$\begin{aligned} r_n(x) &= \frac{\int_{-\infty}^{+\infty} y f_n(x,y)\,dy}{g_n(x)} \\ &= \frac{h_n \sum_{i=1}^{n} m\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} J_1\left(\frac{x-X_i}{h_n}\right)} + \frac{\sum_{i=1}^{n} Y_i J_1\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} J_1\left(\frac{x-X_i}{h_n}\right)}, \end{aligned} \tag{1.4}$$

où

$$m(x) = \int_{-\infty}^{+\infty} yJ(x,y)\,dy,$$

peut être considéré, comme un estimateur de $r(x)$. Si

$$J(x,y) = J_1(x)\,L(y),$$

avec

$$\int_{\mathbb{R}} yL(y) = 0.$$

alors,

$m(x) = 0$ et l'estimateur défini dans la relation (1.4) se réduit à

$$r_n(x) = \frac{\sum_{i=1}^{n} Y_i J_1\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} J_1\left(\frac{x-X_i}{h_n}\right)}.$$

Donc l'estimateur à noyau de la régression est donné par

$$r_n(x;h_n) = \begin{cases} \sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right) / \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) = \Phi_n(x;h_n)/f_{X,n}(x;h_n) \\ \qquad\qquad \text{si} \quad \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \neq 0, \\ \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad\qquad \text{si} \quad \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) = 0, \end{cases}$$

où

K vérifie les conditions $(K.1 - K.6)$ et

$$\Phi_n(x;h_n) = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right).$$

La consistance de $r_n(x;h_n)$ dépend du comportement limite de h_n, Nadaraya (1964) s'est interessé et donna le résultat suivant.

Théorème 1.4. *Supposons que $E\left(Y^2\right) < \infty$, et que $f_X(x)$ est strictement positive.* Si *$h_n \to 0$, $nh_n \to +\infty$, quand $n \to \infty$, alors $r_n(x;h_n)$ est un estimateur consistant de $r(x)$.*

Preuve

Puisque $f_{X,n}(x;h_n)$ est un estimateur consistant de $f_X(x)$, il suffit donc de montrer que $\Phi_n(x)$ est un estimateur consistant de

$$\Phi(x) = \int_{-\infty}^{+\infty} y f(x,y)\, dy\ .$$

Nous avons

$$\begin{aligned} E\left(\Phi_n(x;h_n)\right) &= E\left[\frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} E\left[YK\left(\frac{x-X}{h_n}\right)\right] \\ &= \frac{1}{h_n}\int_{\mathbb{R}} E(Y \mid X=t)\, K\left(\frac{x-t}{h_n}\right) f_X(t)\, dt \\ &= \frac{1}{h_n}\int_{\mathbb{R}} r(t)\, K\left(\frac{x-t}{h_n}\right) f_X(t)\, dt \\ &= \frac{1}{h_n}\int_{\mathbb{R}} \Phi(t)\, K\left(\frac{x-t}{h_n}\right) dt. \end{aligned}$$

Comme Φ, K, h_n vérifient les conditions de Bochner (1955), on a alors,

$$\lim_{n\to\infty} E\left(\Phi_n\left(x;h_n\right)\right) = \Phi\left(x\right)$$

D'autre part, si $\Psi\left(x\right) = \int\limits_{-\infty}^{+\infty} y^2 f\left(x,\ y\right) dy$, comme les X_i sont indépendantes et identiquement distribuées, il s'ensuit quand $n \to \infty$,

$$Var\left[\Phi_n\left(x;h_n\right)\right] \sim \Psi\left(x\right) \int_{-\infty}^{+\infty} K^2\left(t\right) dt.$$

D'où

$$\lim_{n\to\infty} Var\left[\Phi_n\left(x;h_n\right)\right] = 0.$$

Ce qui achève la démonstration du théorème.

En outre, Nadaraya (1989) a établi le théorème suivant.

Théorème 1.5. *a) Si* $|Y| \leq C_1 < \infty$ *P.S et si* $nh_n \to \infty$, quand $n \to \infty$, *alors*

$$Er_n\left(x;h_n\right) = \frac{E\left[\Phi_n\left(x;h_n\right)\right]}{E\left[f_{X,n}\left(x;h_n\right)\right]} + O\left(\frac{1}{nh_n}\right).$$

b)Si $EY^2 < \infty$ *et si* $nh_n^2 \to \infty$, quand $n \to \infty$, *alors*

$$Er_n\left(x;h_n\right) = \frac{E\left[\Phi_n\left(x;h_n\right)\right]}{E\left[f_{X,n}\left(x;h_n\right)\right]} + O\left(n^{-\frac{1}{2}}h_n\right).$$

Les hypothèses a) et b) et le Théorème 1.4 impliquent que $r_n\left(x;h_n\right)$ est un estimateur asymptotiquement sans biais de $r\left(x\right)$.

Preuve

En utilisant l'identité suivante

$$\frac{1}{u} = \frac{1}{c} - c^{-2}\left(u-c\right) + c^{-2}u^{-1}\left(u-c\right)^2,\ c \neq 0,$$

où c est une constante positive. Nous obtenons ce qui suit

$$\begin{aligned} Er_n(x;h_n) &= \frac{E[\Phi_n(x;h_n)]}{Ef_{X,n}(x;h_n)} - (Ef_{X,n}(x;h_n))^{-2} E[\Phi_n(x;h_n) \\ &\quad - E\Phi_n(x;h_n)][f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)] \\ &\quad + E\left\{(f_{X,n}(x;h_n))^{-1}(Ef_{X,n}(x;h_n))^{-2}\Phi_n(x;h_n)\right. \\ &\quad \left.[f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)]^2\right\} \\ &= \frac{E\Phi_n(x;h_n)}{Ef_{X,n}(x;h_n)} + \left[c_n^1(x) + c_n^2(x)\right](Ef_{X,n}(x;h_n))^{-2}, \end{aligned} \tag{1.5}$$

où

$$\begin{aligned} c_n^1(x) &= E[\Phi_n(x;h_n) - E\Phi_n(x;h_n)][f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)] \ , \\ c_n^2(x) &= E\left\{(f_{X,n}(x;h_n))^{-1}\Phi_n(x;h_n)\ [f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)]^2\right\}. \end{aligned}$$

Comme $Var\ f_{X,n}(x;h_n) \sim \frac{1}{nh_n} f_X(x) \int_{-\infty}^{+\infty} K^2(t)\,dt$ et
$Var\ \Phi_n(x;h_n) \sim \frac{1}{nh_n} \int_{-\infty}^{+\infty} y^2 f(x,y)\,dy \int_{-\infty}^{+\infty} K^2(t)\,dt$, alors

$$\left|c_n^1(x)\right| \leq [Varf_{X,n}(x;h_n)]^{\frac{1}{2}}[Var\Phi_n(x;h_n)]^{\frac{1}{2}} = O\left(\frac{1}{nh_n}\right). \tag{1.6}$$

Nous constatons aussi que l'hypothèse a) implique l'inégalité suivante

$$\left|c_n^2(x)\right| \leq C_1 Var\ f_{X,n}(x;h_n) \sim \frac{1}{nh_n} f_X(x) \int_{-\infty}^{+\infty} K^2(t)\,dt = O\left(\frac{1}{nh_n}\right). \tag{1.7}$$

En combinant les relations (1.5), (1.6), (1.7) nous obtenons le résultat a).

Pour démontrer le cas b), il suffit de remarquer que la relation (1.6) est toujours valable mais la relation (1.7) devient

$$\begin{aligned} \left|c_n^2(x)\right| &\leq E\max\{|Y_1|,\ldots,\ |Y_n|\}\, E[f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)]^2 \\ &\leq \left[\sum_{i=1}^{n} EY_i^2\right]^{\frac{1}{2}} \left[E[f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)]^4\right]^{\frac{1}{2}} \\ &= \sqrt{n}\left(EY^2\right)^{\frac{1}{2}} O\left(n^{-1}h_n\right) = O\left(n^{-\frac{1}{2}}h_n\right). \end{aligned}$$

les relations, (1.6), (1.7), (1.8) donnent le résultat b).

Nadaraya (1989) a utilisé ce théorème comme support pour l'étude de la convergence uniforme et de la normalité asymptotique de $r_n(x;h_n)$, d'ailleurs il est le pilier de plusieurs recherches sur l'estimation de $r(x)$.

Théorème 1.6. *Soient* a, b des constantes $-\infty < a \leq b < \infty$, et γ un réel positif, tels que $\min_{a\leq x\leq b} f_X(x) > 0$ et $\sum_{n=1}^{\infty} \exp\left(-\gamma n h_n^2\right) < \infty$. *Si* $K(x)$ *vérifie les conditions* $(K.1 - K.6)$ *et à variation bornée sur* $[a,\ b]$, *alors*

$$\sup_{a\leq x\leq b} |r_n(x;h_n) - r(x)| \longrightarrow 0 \text{ P.S., quand } n \to \infty.$$

Nous ne donnons pas la démonstration de ce théorème, par contre nous énonçons les deux lemmes suivants qui facilitent cette dernière.

Lemme 1.4. *Si* $K(x)$ *vérifie les conditions* $(K.1 - K.6)$ *et si* $0 \leq Y \leq 1$ *P.S, alors ils existent* C_0, C_1, a_1, a_2 des constantes positives *telles que*

$$P\left\{\sup_{a\leq x\leq b} |\Phi_n(x;h_n) - E\ \Phi_n(x;h_n)| > \lambda\right\} \leq C_0 \exp\left(-a_1 n h_n^2\right) + C_1 \exp\left(-a_2 n h_n^2\right).$$

Devroye (1978*b*) a donné une extention de la consistance uniforme dans le cas multidimensionnel.

La normalité asymptotique de $r_n(x;h_n)$ a été traité d'abord par Schuster (1970) sur des points distincts de $\mathbb{R}$, en revanche Nadaraya (1989) l'a généralisé sur $\mathbb{R}$. Les Théorèmes suivants récapitulent nos dires.

Théorème 1.7. (Schuster (1970)) *Soient* $x_1, \ldots, x_k$ *des points distincts de* $\mathbb{R}$ *tels que* $f_X(x_i) > 0$, $1 \leq i \leq k$. Si $E\left(Y^3\right) < \infty$ *et si* f_X', Φ', Ψ', f_X'', Φ'' *existent et elles sont bornées, alors*

$$(nh_n)^{\frac{1}{2}} \left[r_n(x_1;h_n) - r(x_1), \ldots, r_n(x_k;h_n) - r(x_k)\right] \xrightarrow{\mathcal{L}} N(0,\ C),$$

où,

$C = [(C_{i\,j})]$ *est la matrice diagonale des covariances avec*

$$C_{i\,i} = \frac{Var(Y \mid X = x_i)}{f_X(x_i)} \int_{-\infty}^{+\infty} K^2(x)\,dx,\ 1 \leq i \leq k.$$

Théorème 1.8. *Soit* $E|Y|^{2+\delta} < \infty$ *pour* $\delta \in \left(\frac{1}{2},\ 1\right]$. *Si* $f_X(x)$, $\Phi(x)$, $\Psi(x)$ *possèdent des dérivées continues d'ordre inférieur où égale à 2, et si de plus* $n^\delta h_n^{2+\delta} \to \infty$, $nh_n^5 \to 0$, *quand* $n \to \infty$, alors

$$\sqrt{nh_n}\left(r_n\left(x;h_n\right)-r\left(x\right)\right)\xrightarrow{\mathcal{L}}N\left(0,\ \frac{Var\left(Y\mid X=x\right)}{f_X\left(x\right)}\int_{-\infty}^{+\infty}K^2\left(x\right)dx\right).$$

Chapitre II

ETUDE DES CRITÈRES D'ERREURS ET DE LA FENÊTRE

2.1 Introduction

Pour mesurer la perfomance des estimateurs automatiques de la densité et de la régression, plusieurs critères d'erreurs sont disponibles. Ces derniers se répartissent en deux classes, selon qu'on considère une estimation ponctuelle ou globale.
En ce qui concerne l'estimation ponctuelle de la densité, respectivement de la régression, le critère le plus utilisé est l'erreur quadratique moyenne de $f_{X,n}(x;h_n)$. et de $r_n(x;h_n)$ défini respectivement par

$$E\left(f_{X,n}(x;h_n) - f_X(x)\right)^2,$$

et

$$E\left(r_n(x;h_n) - r(x)\right)^2.$$

Pour ce qui est de l'estimation globale, le critère communément utilisé est l'erreur quadratique moyenne intégrée de $f_{X,n}(x;h_n)$ défini par

$$\int E\left(f_{X,n}(x;h_n) - f_X(x)\right)^2 dx.$$

Les critères considérés ci-dessus ont leur analogue dans L_1 (voir Devroye et Györfi (1985)) ainsi que des versions pondérées. Il existe d'autres possibilités comme la distance de d'Hellinger ou le nombre de Kullback-Leiber relié à la quantité d'entropie (voir le chapitre 1 de Devroye (1978) pour plus de détails).

Dans ce travail, nous nous intéressons surtout aux critères de mesure d'erreur de L_2 puisque ce sont les plus commodes à utiliser d'un point de vue théorique (voir Härdle (1990), Hall et Marron (1991) pour une discussion sur l'adaptation des critères cités ci-dessus comme choix correct), par contre les Théorèmes de Deheuvels et Mason (2004) que nous utilisons peuvent être orientés vers des déviations L_∞.

Dans ce chapitre nous supposons que :

- f_X a des dérivées d'ordre inférieur ou égal à trois sur $\mathbb{R}$,
- $f_{X,Y}(x,y)$ est trois fois continûment différentiable,
- $h_n \longrightarrow 0$ quand $n \longrightarrow \infty$,
- $EY^2 < \infty$,

et en plus des conditions $(K.1 - K.6)$, K vérifie :

$(K7)$ $[xK] = \int_{-\infty}^{+\infty} xK(x)\,dx = 0,$

$(K8)$ $[x^2K] = \int_{-\infty}^{+\infty} x^2K(x)\,dx < \infty.$

2.2 Etude de l'erreur quadratique moyenne de $f_{X,n}(x;h_n)$

C'est courant de décomposer la différence de $f_{X,n}(x;h_n) - f_X(x)$ en deux composantes : le terme aléatoire $f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)$;

et le terme biais $Ef_{X,n}(x;h_n) - f_X(x)$.

En se basant sur cette décomposition, l'erreur quadratique moyenne peut s'écrire

$$\begin{aligned} E(f_{X,n}(x;h_n) - f_X(x))^2 &= E(f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n))^2 \\ &\quad +2E(f_{X,n}(x;h_n) - Ef_{X,n}(x;h_n)) \\ &\quad \times (Ef_{X,n}(x;h_n) - f_X(x)) \\ &\quad + (Ef_{X,n}(x;h_n) - f_X(x))^2. \end{aligned}$$

Comme le double produit est nul, donc on a

$$E\left(f_{X,n}(x;h_n)-f_X(x)\right)^2 = Var\ f_{X,n}(x;h_n)+\left(Ef_{X,n}(x;h_n)-f_X(x)\right)^2.$$

Essayons de trouver des expressions simples pour ces deux derniers termes.

On sait que

$$\begin{aligned} Ef_{X,n}(x;h_n) &= E\left(\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)\right) \\ &= \frac{1}{h_n}EK\left(\frac{x-X}{h_n}\right) \\ &= \frac{1}{h_n}\int_{\mathbb{R}}K\left(\frac{x-u}{h_n}\right)f_X(u)\,du. \end{aligned}$$

En posant $\frac{x-u}{h_n}=\alpha$, la relation précédente s'écrit asymptotiquement,

$$\begin{aligned} Ef_{X,n}(x;h_n) &= \int_{\mathbb{R}}K(\alpha)f_X(x-h_n\alpha)\,d\alpha \\ &= \int_{\mathbb{R}}K(\alpha)\left\{f_X(x)-h_n\alpha f_X'(x)+\frac{(h_n\alpha)^2}{2}f_X''(x)\right\}d\alpha + \\ &\quad \frac{h_n^2}{2}o(1). \end{aligned}$$

En utilisant les conditions $(K.7-K.8)$, on a asymptotiquement,

$$Ef_{X,n}(x;h_n)=f_X(x)+\frac{h_n^2}{2}\left\{f_X''(x)\left[t^2K\right]+o(1)\right\}.$$

Finalement on a asymptotiquement,

$$E\left(f_{X,n}(x;h_n)-f_X(x)\right)^2=\frac{h_n^4}{4}\left\{\left(f_X''(x)\right)^2\left[t^2K\right]^2+o(1)\right\}.$$

On a par définition

$$\begin{aligned} Var\ f_{X,n}(x;h_n) &= Var\ \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \\ &= \frac{1}{nh_n^2} Var\ K\left(\frac{x-X}{h_n}\right) \\ &= \frac{1}{nh_n^2} EK^2\left(\frac{x-X}{h_n}\right) - \frac{1}{nh_n^2}\left(EK\left(\frac{x-X}{h_n}\right)\right)^2 \end{aligned}$$

En procédant de la même façon que précédemment et en utilisant les conditions $(K.7 - K.8)$, on a asymptotiquement,

$$\frac{1}{nh_n^2} EK^2\left(\frac{x-X}{h_n}\right) = \frac{1}{nh_n} f_X(x)\left[K^2\right](1+o(1))$$

et

$$\begin{aligned} \frac{1}{nh_n^2}\left(EK\left(\frac{x-X}{h_n}\right)\right)^2 &= \frac{1}{n} f_X^2(x)\left[K^2\right]^2 \\ &= o(1). \end{aligned}$$

Ainsi

$$Var f_{X,n}(x;h_n) = \frac{1}{nh_n} f_X(x)\left[K^2\right](1+o(1)).$$

Par conséquent, quand $n \to \infty$

$$\begin{aligned} E\left(f_{X,n}(x;h_n) - f_X(x)\right)^2 &= \frac{h_n^4}{4}\left\{\left(f_X''(x)\right)^2\left[x^2K\right]^2 + o(1)\right\} \\ &\quad + \frac{1}{nh_n} f_X(x)\left[K^2\right](1+o(1)). \end{aligned}$$

Quand $f_X''(x) \neq 0$, la fenêtre

$$h_n = n^{\frac{-1}{5}}\left[\frac{f_X(x)\left[K^2\right]}{\left(f_X''(x)\right)^2\left[t^2K\right]^2}\right]^{\frac{1}{5}}, \tag{2.1}$$

minimise asympotiquement l'erreur quadratique moyenne de $f_{X,n}(x;h_n)$.

2.3 Etude de l'erreur quadratique moyenne intégrée de $f_{X,n}(x;h_n)$

L'erreur quadratique moyenne intégrée est donnée par

$$\begin{aligned}\int_{\mathbb{R}} E\left(f_{X,n}(x;h_n) - f_X(x)\right)^2 dx &= \int_{\mathbb{R}} \frac{h_n^4}{4}\left(f_X''(x)\right)^2 \left[t^2K\right]^2 dx \\ &+ \frac{\left[K^2\right]}{nh_n}(1+o(1)) \int_{\mathbb{R}} f_X(x)\,dx \\ &= \int_{\mathbb{R}} \frac{h_n^4}{4}\left(f_X''(x)\right)^2 \left[t^2K\right]^2 dx \\ &+ \frac{\left[K^2\right]}{nh_n}(1+o(1)).\end{aligned}$$

Cette quantité est asympotiquement minimisée par

$$h_n = n^{\frac{-1}{5}} \left[\frac{\left[K^2\right]}{\left[t^2K\right]^2 \int_{\mathbb{R}} \left[f_X''(x)\right]^2 dx}\right]^{\frac{1}{5}}.$$

2.4 Etude de l'erreur quadratique moyenne de $r_n(x;h_n)$

En tenant compte de la décomposition de l'erreur quadratique moyenne de $f_{X,n}(x;h_n)$, l'erreur quadratique moyenne de $r_n(x;h_n)$ peut s'écrire

$$E\left(r_n(x;h_n) - r(x)\right)^2 = Var\ r_n(x;h_n) + \left(Er_n(x;h_n) - r(x)\right)^2.$$

Essayons de donner une écriture asymptotique des deux derniers termes.

La condition $EY^2 < \infty$, implique le Théorème 1.5, c'est à dire on doit avoir asymptotiquement

$$Er_n(x;h_n) \sim \frac{E\left[\Phi_n(x;h_n)\right]}{E\left[f_{X,n}(x;h_n)\right]},$$

où

$$\Phi_n(x;h_n) = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right).$$

Donc pour donner une expression asymptotique de $Er_n(x;h_n)$, il suffit d'expliciter asymptotiquement chaque terme de cette fraction.

On a

$$\begin{aligned}
E\left[\Phi_n(x;h_n)\right] &= E\frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right) \\
&= \frac{1}{h_n} E\left(E(Y \mid X) K\left(\frac{x-X}{h_n}\right)\right) \\
&= \frac{1}{h_n}\int_{-\infty}^{+\infty} r(u) K\left(\frac{x-u}{h_n}\right) f_X(u)\, du.
\end{aligned}$$

En posant $\frac{x-u}{h_n} = \alpha$ et en appliquant les conditions $(K.7 - K.8)$, la relation précédente, s'écrit asymptotiquement

$$\begin{aligned}
E\left[\Phi_n(x;h_n)\right] &= \int_{\mathbb{R}} K(\alpha)\, r(x-h_n\alpha)\, f_X(x-h_n\alpha)\, d\alpha \\
&= \int_{\mathbb{R}} K(\alpha)\left\{r(x) - h_n\alpha r'(x) + \frac{(h_n\alpha)^2}{2} r''(x)\right\} \\
&\quad \times \left\{f_X(x) - h_n\alpha f_X'(x) + \frac{(h_n\alpha)^2}{2} f_X''(x)\right\} d\alpha + \frac{h_n^2}{2} o(1) \\
&= \int_{\mathbb{R}} K(\alpha)\left\{r(x) f_X(x) - h_n\alpha r'(x) f_X(x)\right. \\
&\quad \left. - h_n\alpha r(x) f_X'(x) + \frac{\alpha^2 h_n^2}{2}\left[r(x) f_X(x)\right]''\right\} d\alpha + \frac{h_n^2}{2} o(1) \\
&= r(x) f_X(x) + \frac{h_n^2}{2}\left[r(x) f_X(x)\right]''\left[t^2K\right] + \frac{h_n^2}{2} o(1).
\end{aligned}$$

On a aussi d'après précédemment l'expression asymptotique de $Ef_{X,n}(x;h_n)$, c'est à dire

$$Ef_{X,n}(x;h_n) = f_X(x) + \frac{h_n^2}{2}\left\{f_X''(x)\left[t^2K\right] + o(1)\right\}.$$

Par suite, asymptotiquement

$$\begin{aligned} Er_n(x;h_n) &= \frac{r(x) f_X(x) + \frac{h_n^2}{2}\left[r(x) f_X(x)\right]''\left[t^2K\right]}{f_X(x)\left\{1+\frac{h_n^2}{2f_X(x)} f_X''(x)\left[t^2K\right]\right\}} + \frac{h_n^2}{2} o(1) \\ &= r(x) + \frac{h_n^2}{2}\left\{\left[r''(x) + 2r'(x)\frac{f_X'(x)}{f_X(x)}\right]\left[t^2K\right] + o(1)\right\}. \end{aligned}$$

Ces mêmes relations combinées avec l'expression de la variance asymptotique de $f_{X,n}(x;h_n)$, nous mènent à donner la formule de la variance asymptotique de $r_n(x;h_n)$,

$$Varr_n(x;h_n) = \frac{1}{nh_n}\frac{\sigma_I^2}{f_X(x)}\left[K^2\right](1+o(1)),$$

où

$$\sigma_I^2 = Var(Y \mid X = x).$$

Finalement, on a

$$\begin{aligned} E\left(r_n(x;h_n) - r(x)\right)^2 &= \frac{h_n^4}{4}\left\{\left[r''(x) + 2r'(x)\frac{f_X'(x)}{f_X(x)}\right]\left[t^2K\right] + \right. \\ &\quad \left. + o(1)\right\}^2 + \frac{1}{nh_n}\frac{\sigma_I^2}{f_X(x)}\left[K^2\right](1+o(1)). \end{aligned}$$

D'où le h_n minimisant l'erreur quadratique moyenne de $r_n(x;h_n)$ est donnée par

$$h_n = n^{\frac{-1}{5}}\left\{\frac{\frac{\sigma_I^2}{f_X(x)}\left[K^2\right]}{\left\{\frac{h_n^2}{2}\left[r''(x) + 2r'(x)\frac{f_X'(x)}{f_X(x)}\right]\left[t^2K\right]\right\}^2}\right\}^{\frac{-1}{5}}.$$

Il est évident que tous ces arguments sont aussi applicables pour les noyaux d'ordre supérieur (voir Bartlett (1963) ; Watson (1964) et Leadbetter (1963) ; Epanechnikov (1969) ; Deheuvels (1990) ; Schucany (1989) ; Scott (1992) ; Berlinet (1993)), où $[t^pK] = 0$ pour $p = 1, \ldots, q-1$ et $[t^qK] \neq 0$.

Nous remarquons que l'expression de h_n optimal, minimisant asymptotiquement les trois critères d'erreurs cités précédemment a la forme $Cn^{\frac{-1}{5}}$, où la constante C est en fonction de distributions et de termes aléatoires inconnus. Pour couvrir cette difficulté, beaucoup de résultats ont été proposés. Nous, nous limitons à quelques travaux sur l'estimation de la densité.

Une théorie semblable sur la régression a été traité par Härdle (1990) ; Wand et Jones (1995).

2.5 *Quelques méthodes d'optimisation de h_n*

Ces méthodes d'optimisation de h_n ont fait l'objet de trés nombreux travaux de recherche durant cette dernière décennie. Par conséquent, nous ne pouvons pas énoncer la totalité de ces élaborations ni tous les auteurs qui y ont coopéré, nous nous citons que quelques uns.

2.5.1 La méthode de pseudo-maximum de vraisemblance (Pseudo Likelihood Cross-Validation)

La méthode de pseudo-maximum de vraisemblance appelée aussi la validation croisée de Kullback-Leiber (Kullback-Leiber Cross-Validation) ou tout simplement la méthode de la validation croisée, a été étudiée par Habbema, Hermans et Van Der Broek (1974) et Duin (1976). Elle est fondée sur une combinaison du principe de la validation croisée et des idées de vraisemblance, elle consiste à maximiser l'expression suivante

$$L(h_n) = \prod_{i=1}^{n} f_{X,n}(X_i; h_n). \tag{2.2}$$

Cette dernière quantité a un maximum trivial en $h_n = 0$. Pour éviter ce problème de dégénérescence, Habbema, Hermans, Van Der Broek (1974) et Duin (1976) ont eu recours au principe de validation croisée en remplaçant dans la relation (2.2) chaque facteur $f_{X,n}(X_i; h_n)$ par $\hat{f}_{X,n}(X_i; h_n)$, l'estimateur de la densité calculé sur le sous-échantillon $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, i.e.,

$$\hat{f}_{X,n}(X_i; h_n) = \frac{1}{(n-1)\,h_n} \sum_{j=1, j\neq i}^{n} K\left(\frac{X_j - X_i}{h_n}\right)$$

Schuster et Gregory (1981) ont montré que dans le cas où le noyau est à support compact , l'estimateur obtenu en remplaçant h_n par sa valeur optimale, notée h_{PLCV}, est sévèrement affecté par les lois de probabilité ayant des queues lourdes, i.e., des lois qui contiennent relativement moins de probabilité dans ses régions centrales que dans ses queues.

Pour illuster le problème rencontré, nous supposons que le support du noyau K est l'intervalle $[-1,\ 1]$. Soit $X_{1,n},\ldots,X_{n,n}$ la statistique d'ordre de $X_1,\ldots,X_n$, si

$$h_n < X_{n,n} - X_{n-1,n},$$

nous avons

$$1 < \frac{X_{n,n} - X_{n-1,n}}{h_n} \leq \frac{X_{n,n} - X_{j,n}}{h_n} \quad \text{pour tout } 1 \leq j \leq n-1.$$

Ceci implique que $\hat{f}_{X,n}\left(X_{n,n}; h_n\right) = 0$ et donc $L\left(h_n\right) = 0$. D'où le paramètre h_n qui maximise $L\left(h_n\right)$, i.e., h_{PLCV} doit satisfaire :

$$h_{PLCV} \geq X_{n,n} - X_{n-1,n} \quad \text{pour tout } n.$$

Une condition nécessaire pour la convergence de l'estimateur à noyau vers la densité f_X est $h_n \to 0$ dans un certain mode de convergence. Le problème qui se pose avec des lois à queues lourdes est que l'espacement des statistiques d'ordre extrêmes ne converge pas vers zéro dans un certain mode de convergence (voir la deuxième partie de la thèse). Nous profitons de cette occasion pour citer les travaux de Deheuvels et Devroye (1983), qui montrent que la convergence dans L_1 de l'estimateur correspondant à la fenêtre maximisant $L\left(h_n\right)$ est caractérisé par la stabilité des statistiques d'ordre extrêmes de f_X.

La consistance des estimateurs par la méthodes de la validation croisée a été fourni par Chow, Geman et Wu (1983). Ils ont été les premiers à donner une description théorique dans le cas continu, ils ont montré que si le noyau K et la densité f_X sont à support compact alors l'estimateur correspondant est consistant.

Pour remédier aux problèmes des lois à queues lourdes, Hall (1978) a tronqué la densité de l'échantillon observé à un intervalle borné où f_X est supposée strictement positive, il a montré que sa procédure pouvait conduire à de mauvais résultats. En fait, si f_X est concave sur l'intervalle d'estimation, la fenêtre maximisant $L\left(h_n\right)$, était de l'ordre de $n^{\frac{-1}{3}}$, et non

de $n^{\frac{-1}{5}}$. Cependant, Marron (1985) a proposé une modification de la fonction de pseudo-maximum de vraisemblance à validation croisée, la fenêtre choisie par cette méthode ayant les mêmes propriétés asymptotiques qu'un choix déterministe de h_n.
Finalement, nous citons les travaux de Hall (1982) où il est démontré que la méthode de pseudo-maximum de vraisemblance est influencée par l'interaction existant entre les propriétés de queues du noyau et de la densité à estimer. Aussi, une caractérisation en est donnée en montrant qu'elle vise la minimisation du nombre de Kullback-Leiber et qu'elle est plutôt adaptée aux problèmes de discrimination.

2.5.2 Méthode de la validation croisée des moindres carrés

L'erreur quadratique intégrée est un autre moyen d'estimation globale, on l'a pas citée auparavant car elle n'est pas utilisée dans notre travail, mais nous verrons dans la suite qu'il est important de connaitre sa fenêtre optimale.
L'erreur quadratique intégrée est donnée par

$$ISE(h_n) = \int_{\mathbb{R}} \left(f_{X,n}(x;h_n) - f_X(x)\right)^2 dx.$$

Elle peut s'écrire

$$ISE(h_n) = \int_{\mathbb{R}} \left(f_{X,n}(x;h_n)\right)^2 dx - 2\int_{\mathbb{R}} f_{X,n}(x;h_n) f_X(x)\, dx + \int_{\mathbb{R}} \left(f_X(x)\right)^2 dx. \qquad (2.3)$$

Nous constatons que le premier terme est connu par le statisticien alors que le dernier terme est indépendant de h_n. L'idée de Rumedo (1982) et de Bowman (1984) consistait à trouver un estimateur noté $LSCV(h_n)$ à partir des données disponibles et qui en moyenne coïncide avec les deux premiers termes trouvés en (2.3). Le choix de la fenêtre, notée h_{LSCV} ($LSCV$: Least Squares Cross-Validation), s'effectue alors en minimisant $LSCV(h_n)$ par rapport à h_n. L'expression de $LSCV(h_n)$ est donnée par :

$$LSCV(h_n) = \int_{\mathbb{R}} \left(f_{X,n}(x;h_n)\right)^2 dx - \frac{2}{n}\sum_{j=1}^{n} \hat{f}_{X,n}(X_j;h_n), \qquad (2.4)$$

où

$$\hat{f}_{X,n}(X_j;h_n) = \frac{1}{(n-1)h_n} \sum_{j=1, j\neq i}^{n} K\left(\frac{X_j - X_i}{h_n}\right).$$

Une autre écriture intéressante de cette même expression est donnée par :

$$LSCV(h_n) = \frac{[K^2]}{nh_n} + \frac{2}{n^2h_n} \sum\sum_{i<j} \gamma\left(\frac{X_i - X_j}{h_n}\right),$$

où

$$\gamma(u) = \int K(t) K(t+u) \, dt - 2K(u).$$

Nous avons $E(LSCV(h_n)) = E(ISE(h_n)) - \int f_X^2$, ce qui montre que $LSCV(h_n) + \int f_X^2$ est un estimateur sans biais de $MISE(h_n)$ (l'erreur quadratique moyenne intégrée). Minimiser $E(LSCV(h_n))$ revient exactement à minimiser l'erreur quadratique moyenne intégrée $MISE(h_n)$ puisque le terme $\int f_X^2$ est indépendant de h_n. Si nous supposons que les fenêtres minimisant $E(LSCV(h_n))$ et $LSCV(h_n)$ sont proches l'une de l'autre alors la minimisation de $LSCV(h_n)$ peut conduire à de bons choix de la fenêtre.

La méthode de la validation croisée des moindres carrés est une des méthodes qui s'adresse plutôt au critère ISE qu'au critère $MISE$. En fait, le critère ISE a l'avantage de produire une fenêtre aléatoire qui rapproche $f_{X,n}(x; h_n)$ le plus de $f_X(x)$ pour un échantillon donné, au lieu de le faire par rapport à tous les échantillons comme c'est le cas du critère de $MISE$.

Le fait que la fenêtre choisie par cette méthode est asymptotiquement correcte a été démontré par plusieurs auteurs ; nous citons, entre autres, les articles de Hall (1983) et Stone (1984).

Hall a été le premier à démontrer des résultats de ce type. Il a montré que sous certaines conditions sur $f_X(x)$ et K,

$$\frac{h_{LSCV}}{h_{AMISE}} \longrightarrow 1$$

en probabilité quand $n \to \infty$, où h_{AMISE} représente la fenêtre minimisant l'erreur quadratique moyenne intégrée asymptotique donnée précédemment.

Hall et Maron (1991) ont montré que sous les hypothèses suivantes :

•K est un noyau à support compact, symétrique et possède une première dérivée Höldérienne.

•f_X est bornée et deux fois dérivable, f_X' et f_X'' sont bornées et intégrables, et $f_X^{(3)}$ uniformément continue.

Si nous posons

$$C_0 = [K^2]\,,\ C_1 = \frac{[t^2K]^2}{4}\int f_X''(x)^2\,dx,\ C_2 = \left(\frac{C_0}{4C_1}\right)^{\frac{1}{5}},$$

alors,

Théorème 2.1. (Hall-Marron (1991)) *Sous les hypothèses précédentes,*

$$n^{\frac{3}{10}}\left(h_{LSCV} - h_{ISE}\right) \xrightarrow{L} N\left(0,\ \left(\frac{\sigma_1}{C_3}\right)^2\right)$$

et

$$n^{\frac{3}{10}}\left(h_{ISE} - h_{MISE}\right) \xrightarrow{L} N\left(0,\ \left(\frac{\sigma_2}{C_3}\right)^2\right),$$

où

$$C_3 = \frac{2C_0}{C_2^3} + 12C_1C_2^2,$$

$$\begin{aligned}\sigma_1^2 &= \left(\frac{2}{C_3}\right)^3\left(\int f_X^2(x)\,dx\right)\int\left[\int K(y+z)\left\{zK(z)\right\}'dz\right]^2dy + \left(2C_2\left[t^2K\right]\right)^2\\ &\times\left\{\int\left(f_X''(x)\right)^2 f_X(x)\,dx - \left(\int f_X''(x)\,f_X(x)\,dx\right)^2\right\}\end{aligned}$$

et

$$\begin{aligned}\sigma_2^2 &= \left(\frac{2}{C_3}\right)^3\left(\int f_X^2(x)\,dx\right)\left[t^2\left(K'\right)^2\right] + \left(2C_2\left[t^2K\right]\right)^2\\ &\times\left\{\int\left(f_X''(x)\right)^2 f_X(x)\,dx - \left(\int f_X''(x)\,f_X(x)\,dx\right)^2\right\}.\end{aligned}$$

Il est plus intéressant de représenter ces résultats sous cette forme

$$n^{\frac{3}{10}}\left(\frac{h_{LSCV}}{h_{ISE}} - 1\right) \xrightarrow{L} N\left(0,\ \sigma_3^2\right), \tag{2.5}$$

$$n^{\frac{3}{10}}\left(\frac{h_{ISE}}{h_{MISE}} - 1\right) \xrightarrow{L} N\left(0,\ \sigma_4^2\right), \tag{2.6}$$

où

$$\sigma_3 = \frac{\sigma_1}{C_2C_3} \quad \text{et} \quad \sigma_4 = \frac{\sigma_2}{C_2C_3}.$$

Pour montrer (2.5) et (2.6), il suffit de remarquer que ;

$$n^{\frac{1}{10}}\left(\frac{h_{LSCV}}{h_{ISE}} - 1\right) = \frac{n^{\frac{3}{10}}\left(h_{LSCV} - h_{ISE}\right)}{n^{\frac{1}{5}}h_{ISE}},$$

mais d'après le Théorème de Hall -Marron (1991), *nous avons*

$$n^{\frac{1}{5}} h_{ISE} = n^{\frac{1}{5}} h_{MISE} + O_p\left(n^{\frac{-1}{10}}\right) = C_2 + o_p(1).$$

Les Théorèmes de Hall, Marron et Slutsky (voir Serfling (1979) p19) achèvent la démonstration).

Le Théorème 4.1 de Scott et de Terrell (1987) montre que sous certaines conditions, nous avons :

$$n^{\frac{1}{10}} \left(\frac{h_{LSCV}}{h_{MISE}} - 1\right) \xrightarrow{L} N\left(0,\ \sigma^2_{LSCV}\right), \tag{2.7}$$

où

$$\sigma^2_{LSCV} = \frac{2 \int_{\mathbb{R}} g^2(u)\, du \int_{\mathbb{R}} f^2(x)\, dx}{25 \left[t^2 K\right]^{\frac{2}{5}} \left(\int_{\mathbb{R}} \left(f_X''\right)^2(x)\, dx\right)^{\frac{1}{5}} \left(\left[K(t)\right]\right)^{\frac{9}{5}}},$$

et

$$g(u) = u \int K(t)\, K'(t+u)\, dt - 2uK'(u).$$

Nous constatons que la vitesse de convergence de h_{LSCV} vers h_{ISE} est assez lente.

Cette dernière remarque laisse à supposer que la méthode des moindres carrés à validation croisée n'est pas aussi performante qu'on le croyait, et l'expression asymptotiquement correcte doit être interprétée avec précaution. Cependant, la vitesse de convergence de h_{ISE} vers h_{MISE} est du même ordre de grandeur, ce qui est rassurant dans un certain sens et justifie l'utilisation de cette méthode. Puisque les vitesses sont si lentes, les constantes σ_3^2 et σ_4^2 ont un rôle considérable à jouer. Dans ce cas l'inégalité de Cauchy-Shwartz permet d'affirmer que $\sigma_3^2 \geq \sigma_4^2$.

Le fait que $n^{\frac{3}{10}}$ est présent dans les deux relations données par (2.5) et (2.6) laissent à suspecter que cette vitesse de convergence est la meilleure possible. En particulier, Hall et Marron (1991) ont montré sous certaines hypothèses, que pour toute fonction mesurable $h_n = h(X_1, \ldots, X_n)$,

$$\lim_{\varepsilon \to 0} \liminf_{n \to \infty} \sup_{f \in \sum} P_f\left(\left|\frac{h_n - h_{ISE}}{h_{ISE}}\right| > \varepsilon n^{\frac{-1}{10}}\right) = 1,$$

où

$$\sum = \left\{f :\ f \text{ densité deux fois dérivable telle que : } \left|f''(x)\right| \leq B,\ B > 0\right\}.$$

La conséquence de ce résultat est, bien que la méthode de la validation croisée des moindres carrés donne une fenêtre qui n'est pas très proche de l'optimum, qu'elle reste la meilleure méthode automatique de sélection d'une fenêtre si ce sont les mêmes hypothèses sur f_X qui sont posées.

Toutefois, un des inconvénients de cette méthode est que la fonction $LSCV\left(h_n\right)$ semble (voir Scott et Terrell (1987)) posséder plusieurs minima locaux. Ce fait a été analysé théoriquement par Hall et Marron (1991). En modélisant cette dernière fonction sous la forme d'un processus aléatoire Gaussien, Hall et Marron (1991) ont montré que : asymptotiquement, le nombre de maxima locaux de $LSCV\left(h_n\right)$ était une fonction croissante d'un rapport de deux quantités mesurant les irrégularités de f_X. Une solution à ce problème est de choisir le plus grand parmi ces minima.

Nous terminons ce paragraphe par citer un second inconvenient qui est le temps consommé par cette méthode du point de vue calcul, la méthode de la transformée de Fourrier rapide FFT (Fast Fourier Transform) peut alors être appliquée pour pallier à ce dernier problème (voir Silverman (1986) p. 61).

2.5.3 Méthode d'injection (Plug-In methods)

On sait que le choix optimal asymptotique de la fenêtre minimisant l'erreur quadratique moyenne intégrée de $f_{X,n}\left(x;h_n\right)$, dépend de la densité, du choix du noyau et de la taille de l'échantillon et elle est donnée par :

$$h_{AMISE} = \alpha\left(K\right)\beta\left(f_X\right)n^{\frac{-1}{5}}, \tag{2.8}$$

où

$$\alpha\left(K\right) = \left\{\left[K^2\right]\left[t^2K\right]^{-2}\right\}^{\frac{1}{5}}$$

et

$$\beta\left(f_X\right) = \left\{\int\limits_{\mathbb{R}}\left[f''_X\left(x\right)\right]^2 dx\right\}^{\frac{-1}{5}}.$$

L'idée essentielle de ces méthodes consiste à remplacer la seule partie inconnue dans (2.8), $\beta\left(f_X\right)$, par un estimateur consistant. L'idée de base semble remonter aux premiers travaux de Woodroofe (1970). parmi les auteurs qui s'y sont également intéressés, nous citons Sheather (1992) et Park et Marron (1990).

Nous profitons de ce paragraphe pour décrire la méthode appelée communément règle du pouce (Rule of Thumb), elle est à rapprocher des méthodes d'injection et consiste à remplacer f_X dans h_{AMISE} par une densité normale de paramètre d'échelle λ et ensuite à l'estimer. Cette approche revient à Deheuvels (1977) qui fut le premier à la proposer. La fenêtre donnée par la régle du pouce a pour expression

$$h_{RT} = 1.364\hat{\lambda}\left\{\left[K^2\right]\left[t^2K\right]^{-2}\right\}^{\frac{1}{5}} n^{\frac{-1}{5}}. \tag{2.9}$$

où $\hat{\lambda}$ est un estimateur robuste de λ, c'est à dire, vérifiant $\hat{\lambda} - \lambda = O_p\left(n^{\frac{-1}{2}}\right)$.

Nous n'avons pas voulu citer cette méthode en détail car elle n'est pas entièrement automatique puisqu'elle utilise une information à priori sur f_X.

2.5.4 Méthode de la validation croisée biaisée

Scott et Terrell (1987) ont proposé et étudié cette méthode qui est hybride, combinant les méthodes dites d'injection et la méthode de la validation croisée des moindres carrés. Nous savons, après étude de l'erreur quadratique moyenne intégrée de $f_{X,n}(x; h_n)$, sa représentation asymptotique est donnée par

$$AMISE(h_n) = \frac{\left[K^2\right]}{nh_n} + \left[t^2K\right]^2 \frac{h_n^4}{4} \int_{\mathbb{R}} \left(f_X''(x)\right)^2 dx.$$

La fenêtre obtenue par cette méthode, notée h_{BCV}, est le paramètre minimisant l'estimateur de $AMISE(h_n)$, obtenu en remplaçant $\int_{\mathbb{R}} \left(f_X''(x)\right)^2 dx$ par

$$\int_{\mathbb{R}} \left(f_{X,n}''(x, X_i)\right)^2 dx - \frac{1}{nh_n}\int_{\mathbb{R}} \left\{K''(t)\right\}^2 dt.$$

Plus clairement, h_{BCV} minimise la fonction donnée par :

$$BCV(h_n) = \frac{\left[K^2\right]}{nh_n} + \frac{\left[t^2K\right]^2}{2n h_n^2} \sum\sum_{i<j} \phi\left(\frac{X_i - X_j}{h_n}\right), \tag{2.10}$$

où

$$\phi(u) = \int_{\mathbb{R}} K''(t)\, K''(t+u)\, dt.$$

(On notera la similitude des fonctions $LSCV(h_n)$ et $BCV(h_n)$ données respectivement par (2.4) et (2.10)).

A la différence des méthodes d'injection, le paramètre utilisé dans l'estimateur de $\int (f_X^{"})^2$ est le même que celui présent dans la représentation asymptotique de l'erreur quadratique moyenne intégrée. Dans les méthodes d'injection, on considère que le problème de l'estimateur $\int (f_X^{"})^2$ est différent du problème de l'estimation de la densité, par conséquent, l'utilisation d'une autre fenêtre différente de celle présente dans $AMISE$ est nécessaire.

Pour la méthode de la validation croisée biaisée, Scott et Terrell (1987) ont établi des résultats analogues à ceux du paragraphe précédent. Nous allons en donner une extension faite par Park et Marron (1990).

Théorème 2.2 *Si f est régulier d'ordre p (voir Définition 4.1), nous avons quand $2 < p \leq 2.5$,*

$$\frac{h_{BCV}}{h_{MISE}} - 1 = O_P\left(n^{\frac{-p}{5}}\right),$$

et quand $p > 2.5$,

$$n^{\frac{1}{10}}\left(\frac{h_{BCV}}{h_{MISE}} - 1\right) \xrightarrow{L} N\left(0,\ \sigma^2_{BCV}\right), \tag{2.11}$$

où

$$\sigma^2_{BCV} = \frac{\left[t^2K\right]^{\frac{18}{5}} \int_{\mathbb{R}} \Theta^2(x)\,dx \int_{\mathbb{R}} f_X^2(x)\,dx}{200\left[\int_{\mathbb{R}} \left(f_X^{"}(x)\right)^2 dx\right]^{\frac{1}{5}} \left[K^2\right]^{\frac{9}{5}}},$$

avec

$$\Theta(x) = \int_{\mathbb{R}} K^{"}(t)\,K^{(3)}(t+x)\,dt$$

La comparaison des résultats trouvés en (2.11) et (2.7) montre que le facteur $n^{\frac{-1}{10}}$ est le même dans les deux quantités. Comme cette vitesse est assez lente, les constantes σ^2_{BCV} et σ^2_{LSCV} sont déterminantes. Scott et Terrell (1987) ont montré que la constante σ^2_{BCV} est inférieure à la constante σ^2_{LSCV}. De plus, ils ont remarqué que pour des échantillons dont la taille est de l'ordre de 25, la fonction $BCV(h_n)$ pouvait ne pas présenter de minima locaux et que ce problème disparaissait en augmentant la taille de l'échantillon.

2.5.5 Sur-estimation (oversmoothing)

Nous savons que la représentation asymptotique de h_{AMISE} minimisant l'erreur quadratique moyenne intégrée de $f_{X,n}(x;h_n)$ est

$$h_{AMISE} = n^{\frac{-1}{5}} \left[\frac{[K^2]}{[t^2K]^2 \int\limits_{\mathbb{R}} \left[f_X''(x)\right]^2 dx} \right]^{\frac{1}{5}}. \tag{2.12}$$

Si nous pouvons trouver une borne inférieure à la seule quantité inconnue dans h_{AMISE} qui est $\int (f_X'')^2$, sa substitution dans (2.12), donne une fenêtre qui sera plus grande asymptotiquement que toute autre fenêtre minimisant le critère d'erreur quadratique moyenne intégrée de $f_{X,n}(x;h_n)$ $(MISE)$. Cette idée a été proposée par Devroye et Györfi (1985) dans le cadre de L_1 et par Scott et Terrell (1987) dans celui de L_2. Terrell (1990) a montré que si la densité f_X possède une variance σ^2, alors la fonction $\int (f_X'')^2$ est l'élément de la famille donnée par :

$$h(x) = \begin{cases} \frac{35}{32}\left(1-x^2\right)^3, & -1 \leq x \leq 1 \\ 0, & \text{sinon,} \end{cases}$$

de variance σ^2. Plus précisément,

$$h(x) = \begin{cases} \frac{1}{3\sigma}\frac{35}{32}\left(1-\frac{x^2}{9\sigma^2}\right)^3, & -3\sigma \leq x \leq 3\sigma \\ 0, & \text{sinon.} \end{cases}$$

Plus généralement, la famille des lois $\beta(k+2,\ k+2)$ minimise la quantité $\int \left(f_X^{(k)}\right)^2$ correspondant aux noyaux d'ordres supérieurs. Des résultats similaires existent aussi pour l'histogramme.

Toute fenêtre h_{AMISE} minimisant l'erreur quadratique moyenne intégrée asymptotique vérifie la relation suivante :

$$h_{AMISE} \leq 3\sigma \frac{\left([K^2]\right)^{\frac{1}{5}}}{\left([t^2K]\right)^{\frac{2}{5}}} (35n)^{\frac{-1}{5}}.$$

La fenêtre obtenue par la méthode sur-estimation, notée h_{OS}, est donnée par :

$$h_{OS} = 3\hat{\sigma} \frac{\left([K^2]\right)^{\frac{1}{5}}}{\left([t^2K]\right)^{\frac{2}{5}}} (35n)^{\frac{-1}{5}},$$

où $\hat{\sigma}$ désigne un estimateure robuste de l'écart-type σ obtenu à partir des données. L'un des intérêts de cette méthode est que l'estimateur obtenu est stable puisque les données n'interviennent que dans l'écart-type, et que l'on connait des estimateurs de cette dernière quantité qui ont une vitesse de convergence assez rapide.

2.5.6 L'estimateur proposé par Kappenman

L'estimateur de Kappenman (1958) consiste à considérer deux estimateurs différents de $E(f_X(X))$.

Un premier estimateur de $E(f_X(X))$ est :

$$\int f_{X,n}(x;\ h_n)\,dx = \frac{1}{n^2 h_n}\sum_{j=1}^{n}\sum_{i=1}^{n} K^*\left(\frac{X_j - X_i}{h_n}\right),$$

où K^* désigne la convolée de K par lui même, i.e., $K^*(u) = K*K(u)$. Le second estimateur est :

$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{(n-1)\,h_n}\sum_{i=1,j\neq i}^{n} K\left(\frac{X_j - X_i}{h_n}\right).$$

Le paramètre de Kappenman, noté h_K, est la solution de l'équation,

$$\frac{1}{n^2 h_n}\sum_{j=1}^{n}\sum_{i=1}^{n} K^*\left(\frac{X_j - X_i}{h_n}\right) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{(n-1)\,h_n}\sum_{i=1,j\neq i}^{n} K\left(\frac{X_j - X_i}{h_n}\right),$$

qui, si le noyau est normal, après simplification devient,

$$\sum_{i=1,j\neq i}^{n}\exp\left\{-\frac{(X_j - X_i)^2}{2h_n^2}\right\} - \frac{n-1}{n\sqrt{2}}\left[n + \sum_{i=1,j\neq i}^{n}\exp\left\{-\frac{(X_j - X_i)^2}{4h_n^2}\right\}\right] = 0. \qquad (2.13)$$

Son idée était basée sur le fait qu'une valeur raisonnable du paramètre h_n, est celle qui à partir de deux estimateurs convenables d'une même quantité, produit le même estimateur. Mielniczuk, Sarda et Vieu (1989) ont prouvé que la fenêtre choisie par cette méthode est asymptotiquement sous-optimale (voir Marron (1992)). L'idée intuitive de Kappenman et les résultats pessimistes de Mielniczuk, Sarda et Vieu (1989) ont conduit Marron (1992) à modifier cet estimateur afin de réduire son biais. Il a aussi montré que h_K a une distribution asymptotique normale et que,

$$\frac{h_K}{h_{EK}} - 1 = O_P\left(n^{\frac{-1}{6}}\right),$$

où h_{EK} est la solution de la moyenne de la quantité donnée en (2.18).
Nous remarquons que cette correction a permis une nouvelle caractérisation de la fenêtre h_{LSCV} discutée antérieurement.

2.5.7 L'estimation des k points les plus proches

Nous achevons ce paragraphe par un dernier estimateur des k points les plus proches, cette méthode a été proposée initialement par Fix et Hodges (1951) et introduite explicitement et étudiée par Loftsgaarden et Quesenberry (1965). L'estimateur correspondant est défini de la façon suivante :

$$f_n(x) = \frac{k-1}{2nd_k(x)},$$

où $d_k(x)$ représente la distance de x à son k^{ieme} voisin le plus proche parmi $X_1, \ldots, X_n$. Cet estimateur a été étudié par plusieurs auteurs, nous citons particulièrement Moore et Yackel (1977), Mack et Rosenblatt (1979), une étude détaillée se trouve dans Bosq et Lecoutre (1987).
Si on veut estimer une densité en un point particilier x, il est naturel de construire un estimateur basé sur les observations situées dans un voisinage de ce point. La différence essentielle entre l'estimateur des k points les plus proches et les estimateurs à noyau discutés précédemment est que le nombre d'observations est spécifié à l'avance et c'est la mesure du voisinage qui est aléatoire alors pour l'estimateur naïf, par exemple, le nombre d'observations situées dans le voisinage de x est aléatoire alors que la mesure de ce dernier est fixe.
Dans les queues d'une distribution, le nombre d'observations peut être très faible et l'estimateur des k points les plus proches se trouve alors plus adapté à ce genre de situation que l'estimateur à noyau fixe.
Notons que cet estimateur n'est pas une densité et qu'il est plutôt réservé à l'estimation ponctuelle de la densité.

Jusqu'à présent nous avons discuté surtout sur la fenêtre optimale minimisant les critères d'erreur de $f_{X,n}(x;\ h_n)$, une théorie très semblable a été établie sur $r_n(x; h_n)$ (voir Härdle (1990), Wand et Jones (1995)).

où h_{EK} est la solution de la moyenne de la quantité donnée en (2.10).
Nous remarquons que cette correction a permis une nouvelle caractérisation de la fenêtre h_{LSCV} discutée antérieurement.

2.5.7 L'estimation des k points les plus proches

Nous achevons ce paragraphe par un dernier estimateur des k points les plus proches, cette méthode a été proposée initialement par Fix et Hodges (1951) et introduite explicitement et étudiée par Loftsgaarden et Quesenberry (1965). L'estimateur correspondant est défini de la façon suivante :

$$f_n(x) = \frac{k-1}{2nd_k(x)},$$

où $d_k(x)$ représente la distance de x à son k^{ieme} voisin le plus proche parmi $X_1, \ldots, X_n$. Cet estimateur a été étudié par plusieurs auteurs, nous citons particulièrement Moore et Yackel (1977), Mack et Rosenblatt (1979), une étude détaillée se trouve dans Bosq et Lecoutre (1987).
Si on veut estimer une densité en un point particilier x, il est naturel de construire un estimateur basé sur les observations situées dans un voisinage de ce point. La différence essentielle entre l'estimateur des k points les plus proches et les estimateurs à noyau discutés précédemment est que le nombre d'observations est spécifié à l'avance et c'est la mesure du voisinage qui est aléatoire alors pour l'estimateur naïf, par exemple, le nombre d'observations situées dans le voisinage de x est aléatoire alors que la mesure de ce dernier est fixe. Dans les queues d'une distribution, le nombre d'observations peut être très faible et l'estimateur des k points les plus proches se trouve alors plus adapté à ce genre de situation que l'estimateur à noyau fixe.
Notons que cet estimateur n'est pas une densité et qu'il est plutôt réservé à l'estimation ponctuelle de la densité.

Jusqu'à présent nous avons discuté surtout sur la fenêtre optimale minimisant les critères d'erreur de $f_{X,n}(x;\ h_n)$, une théorie très semblable a été établie sur $r_n(x; h_n)$ (voir Härdle (1990), Wand et Jones (1995)).

Discutons maintenant sur l'évaluation de la vitesse de convergence des estimateurs.

2.6 Evaluation de la vitesse de convergence des estimateurs

Soient $U_1, \ldots, U_n$ une suite de variables aléatoires indépendantes, de loi uniforme sur $[0,\ 1]$, $G_n(t)$ la distribution empirique définie par

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{U_i \le t\}}, \qquad -\infty < t < +\infty,$$

et $\varpi_n(h)$ le module de l'oscillation du processus empirique uniforme donné par :

$\forall h,\ 0 < h < 1$

$$\varpi_n(h) = \sup\left\{\left|\sqrt{n}\left\{G_n(t+s) - G_n(t) - s\right\}\right| : 0 \le t,\ t+s \le 1,\ 0 \le s \le h\right\}.$$

Stute (1986a) est l'initiateur de la convergence presque sûre du module de l'oscillation du processus empirique uniforme, il a montré que si (h_n) est une suite de nombres réels positifs vérifiant les conditions suivantes :

(i) $h_n \to 0$, $nh_n \to \infty$ quand $n \to \infty$;

(ii) $\frac{nh_n}{\log \frac{1}{h_n}} \to \infty$, quand $n \to \infty$;

(iii) $\frac{\log \frac{1}{h_n}}{\log \log n} \to \infty$, quand $n \to \infty$,

alors

$$\lim_{n\to\infty} \frac{\varpi_n(h_n)}{\sqrt{2h_n \log \frac{1}{h_n}}} = 1 \qquad P.S. \tag{1.14}$$

Cette quantité est une source importante de plusieurs résultats sur la vitesse de la consistance uniforme des estimateurs.

A partir de la relation (1.14) Stute (1986b) a établi la vitesse exacte de la consistance forte uniforme du noyau $f_{X,n}(x;\ h_n)$ sur des intervalles compacts de $\mathbb{R}$; à savoir, sous certaines conditions sur f_X,

$$\lim_{n\to\infty}\sqrt{nh_n}\sup_{z\in J}\left\{\frac{|f_{X,n}(z;\ h_n)-Ef_{X,n}(z;\ h_n)|}{\sqrt{2\left[K^2\right]f_X(z)\log\frac{1}{h_n}}}\right\}=1\qquad P.S\ ,\tag{1.15}$$

où J est un intervalle compact de $\mathbb{R}$.

Deheuvels et Mason (1992) ont étendu le résultat (1.14) aux fonctions aléatoires $\zeta_n(t,\ .)$ définies sur $[0,\ 1]$ par :

$$\zeta_n(t,\ s)=\sqrt{\frac{n}{h_n}}\left\{G_n(t+h_ns)-G_n(t)-h_ns\right\},$$

où $0\le s\le 1,\ \ 0\le t\le 1-h_n$.

Cette extension leur a permis d'obtenir une vitesse exacte de la consistance forte pour un nombre d'estimateurs non paramétriques de la densité (voir le corollaire 3 de Deheuvels et Mason (1992)). Motivés par ces travaux , Einhmal et Mason (2000) ont développé des techniques pour la théorie des processus empiriques qu'ils ont combinées avec les méthodes Deheuvels et Mason (1992) pour aboutir au résultat suivant :

Soient $I=[a,b]$ un intervalle de $\mathbb{R}$, p une constante réelle superieur à 2 ,

$F=\left\{g; g \text{ une fonction mesurable telle que } \sup_{y\in\mathbb{R}}|g(y)|<\infty,\right.$

$\left.\text{et } \sup_{z\in I}E\left(g(Y)^p\mid X=z\right)<\infty\right\},$

$W_n(z,\ g)=\sum_{i=1}^{n}\left(c(z)g(Y_i)+d(z)\right)K\left(\frac{X_i-z}{h_n}\right)-nE\left[\left(c(z)g(Y)+d(z)\right)\times K\left(\frac{X-z}{h_n}\right)\right],$

$\sigma^2=\sup_{g\in F}\sup_{z\in I}E\left[\left(c(z)g(Y)+d(z)\right)^2\mid X=z\right]f_X(z)\left[K^2\right]$,

où c et d sont deux fonctions continues sur I .

Théorème 2.3. *Si K est une fonction continue sur* $\left(\frac{-1}{2},\ \frac{1}{2}\right]$, *vérifiant* $(K.1)$ et $(K.5)$, *et* si (h_n) est un paramètre réel positif vérifiant les conditions suivantes :

(i) $\quad h_n \to 0,\ nh_n \to \infty$ *quand* $n \to \infty$;

(ii) $\quad \frac{|\log h_n|}{\log\log n} \to \infty$ *quand* $n \to \infty$;

(iii) $\quad \frac{nh_n}{\log n} \to \infty$ *quand* $n \to \infty$,

$(iiii)$ $\quad h_n^{-1} \leq \left(\frac{n}{\log h_n}\right)^{1-2p}$ *pour* $p > 2$,

alors

$$\frac{\sup_{g\in F}\sup_{z\in I}|W_n(z,\ g)|}{\sqrt{2nh_n|\log h_n|}} \xrightarrow[n\to\infty]{P} \sigma.$$

En combinant ce théorème avec les méthodes de Deheuvels et Mason (1992), Einhmal et Mason (2000) ont déduit une vitesse exacte de la consistance forte sur des intervalles compactes pour un certain type d'estimateurs non paramétriques.

Soient

$$r_n(z,g) = \frac{\sum\limits_{i=1}^{n} g(Y_i)\, K\left(\frac{z-X_i}{h_n}\right)}{\sum\limits_{i=1}^{n} K\left(\frac{z-X_i}{h_n}\right)},$$

$$m_n(z,g) = \frac{1}{nh_n} E\sum_{i=1}^{n} g(Y_i)\, K\left(\frac{z-X_i}{h_n}\right),$$

et

$$\tau_{a,b} = \sup_{g\in F}\sup_{z\in I}\frac{Var(g(Y)\mid X=z)\left[K^2\right]}{\sqrt{f_X(z)}}.$$

Corollaire 2.1. (Einhmal et Mason (2000)). *Si les conditions* $(i-iiii)$ $(K.1)$ et $(K.5)$ *sont vérifiées, alors*

$$\sup_{z\in J}\sqrt{nh_n}\left\{\frac{|f_{X,n}(z;\ h_n) - Ef_{X,n}(z;\ h_n)|}{\sqrt{2\left[K^2\right] f_X(z)\,|\log h_n|}}\right\} \xrightarrow[n\to\infty]{P} 1 \tag{1.16}$$

et

$$\frac{\sup_{g\in F}\sup_{z\in I}|r_n(z,g) - (E\, m_n(z,g)/\, Ef_{X,n}(z;\ h_n))|}{\sqrt{2|\log h_n|}} \xrightarrow[n\to\infty]{P} \tau_{a,b},$$

Nous constatons que ces résultats sont meilleurs que ceux de Härdle, Jansen et Serfling (1988) qui ont obtenu seulement des vitesses approximatives.

La meilleure évolution de la vitesse est traitée par Deheuvels et Mason (2004), elle leur a permis de construire des bandes de confiance asymptotiques basées sur le résultat de la consistance uniforme des estimateurs sur des intervalles fermés de $\mathbb{R}$.

2.7 Bandes de confiance

Deheuvels et Mason (2004) ont établi des bandes de confiance simultanées asymptotiques uniformes et non uniformes pour les fonctionnelles de la distribution $(X,\ Y)$, ils s'intéressent en particulier à la régression de $\Psi(Y)$ sachant $X = x$, notée par

$$r_{\Psi}(x) = E(\Psi(Y) \mid X = x),$$

où Ψ est une fonction mesurable à valeur dans $\mathbb{R}$ et bornée sur chaque compact de $\mathbb{R}$.

Ils imposent certaines conditions sur la distribution du couple $(X,\ Y)$ fournies par les hypothèses $(F.1 - F.4)$ ci-dessous

Notons par $I = [a,\ b]$ et $J = \left[a',\ b'\right]$ deux intervalles de $\mathbb{R}$ tels que

$$-\infty < a' < a < b < b' < +\infty.$$

$(F.1)$ pour chaque $x \in J$, $\lim\limits_{z \to x;\ z \in J} f_{X,Y}(z,\ y) = f_{X,Y}(x,\ y)$ pour presque tout $y \in \mathbb{R}$;

$(F.2)$ $f_X(.)$ est continue et strictement positive sur J;.

$(F.3)$ $Y 1_{\{X \in J\}}$ est borné;

$(F.4)$ $\beta_{M(\Psi)} := \sup_{x \in J} E(M(|\Psi(Y)|) \mid X = x) < \infty$;

où

$M(x) = x^p$ pour $p > 2$; ou bien $M(x) = \exp(sx)$ pour $s > 0$.

Leurs estimateurs sont basés sur les estimateurs à noyaux non paramétriques de Akaike (1954), Rosenblatt (1956), Parzen (1962), Nadaraya (1964) et Watson (1964). Par noyau, ils entendent une fonction $K(.)$ mesurable à variation $|K|_v$ bornée sur $\mathbb{R}$, tel que

$$0 < |K|_v = \int_{\mathbb{R}} |dK(t)| < \infty.$$

De plus, le noyau $K(.)$ vérifie

$(K.1)$ (i) $K = K_1 - K_2$ où K_1 et K_2 sont non-décroissantes sur $\mathbb{R}$;

(ii) K_1 et K_2 sont continus à droite sur $\mathbb{R}$;

(iii) $K_j(s) = 0$, $j = 1, 2$, pour $s \notin \left[-\frac{\zeta}{2}, \frac{\zeta}{2}\right)$, pour un certain $\zeta \in]0, \infty[$;

$(K.2)$ $\int_{\mathbb{R}} K(t)\, dt = 1.$

Pour la réalisation de leurs travaux, ils ont choisi une fenêtre aléatoire $H_n(x)$ et une fonction aléatoire $\Theta_n(x)$, telles que :

$H_n(x) = h_n C_n(x)$, $x \in I$;

$(B.1)$ $P\left(c_1 h_n \leq \inf_{x\in I} H_n(x) \leq \sup_{x\in I} H_n(x) \leq c_2 h_n\right) \longrightarrow 1$, quand $n \to \infty$;

$(B.2)$ $\forall \varepsilon > 0$, $P\left(\sup_{x\in I}\left|\frac{H_n(x)}{h_n} - C(x)\right| \geq \varepsilon\right) \to 0$, quand $n \to \infty$;

$(\Theta.1)$ $\forall \varepsilon > 0$, $P\left(\sup_{x\in I}\left|\frac{\Theta_n(x)}{\Theta(x)} - 1\right| > \varepsilon\right) \to 0$, quand $n \to \infty$.

où $\Theta(x)$ et $C(x)$ sont des fonctions spécifiées positives et continues sur I, c_1 et c_2 sont des constantes strictement positives et h_n est un paramètre réel positif vérifiant les conditions suivantes :

$(H.1)$ $h_n \to 0$, quand $n \to \infty$;

$(H.2)$ $\frac{nh_n}{\log n} \to \infty$, quand $n \to \infty$;

$(H.3)$ $nh_n \left\{\frac{\log n}{M^{inv}(n)^2}\right\} \to \infty$, quand $n \to \infty$.

Notons que pour tout $x \in J$, les conditions $(F.1 - F.4)$ impliquent que la fonction de régression de $\Psi(Y)$ sachant $X = x$ et la variance conditionnelle de $\Psi(Y)$ sachant $X = x$ sont bien définies et elles sont données par

$$r_\Psi(x) = E(\Psi(Y) \mid X = x) = \frac{1}{f_X(x)} \int_{\mathbb{R}} \Psi(y) f_{X,Y}(x, y)\, dy = \frac{m_\Psi(x)}{f_X(x)},$$

où

$$m_\Psi(x) = \int_{\mathbb{R}} \Psi(y) f_{X,Y}(x, y)\, dy$$

et

$$
\begin{aligned}
\sigma_{\Psi}^{2} &= Var\left(\Psi\left(Y\right) \mid X=x\right) \\
&= \frac{1}{f_{X}\left(x\right)} \int_{\mathbb{R}} \left(\Psi\left(y\right)-r_{\Psi}\left(x\right)\right)^{2} f_{X,Y}\left(x,\ y\right) dy\ .
\end{aligned}
$$

Posons

$$
m_{\Psi;n}\left(x;h_{n}\right)=\frac{1}{nh_{n}}\sum_{i=1}^{n}\Psi\left(Y_{i}\right)K\left(\frac{x-X_{i}}{h_{n}}\right),
$$

$$
r_{\Psi;n}(x;h_n)=\begin{cases}\dfrac{\sum_{i=1}^{n}\Psi(Y_i)K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)} & \text{si } f_{X,n}\left(x;\ h_n\right)\neq 0,\\[2ex] \frac{1}{n}\sum_{i=1}^{n}\Psi\left(Y_i\right) & \text{si } f_{X,n}\left(x;\ h_n\right)=0,\end{cases}
$$

$$
\sigma_{\Psi;n}^{2}(x;n)=\begin{cases}\dfrac{\sum_{i=1}^{n}(\Psi(Y_i)-r_{\Psi.n}(x;h_n))^2K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)} & \text{si } f_{X,n}\left(x;\ h_n\right)\neq 0,\\[2ex] \frac{1}{n}\sum_{i=1}^{n}(\Psi\left(Y_i\right)-\frac{1}{n}\sum_{j=1}^{n}\Psi\left(Y_j\right))^2 & \text{si } f_{X,n}\left(x;\ h_n\right)=0,\end{cases}
$$

$$
\hat{E}r_{\Psi;n}(x;h_n)=\begin{cases}\dfrac{E(\Psi(Y)K(\frac{x-X}{h_n}))}{E\left(K(\frac{x-X}{h_n})\right)} & \text{si } Ef_{X,n}\left(x;\ h_n\right)\neq 0,\\[2ex] E(\Psi\left(Y\right)) & \text{si } Ef_{X,n}\left(x;\ h_n\right)=0,\end{cases}
$$

alors,

Théorème 2.4. (Deheuvels et Mason (2004)). Sous les hypothèses $(F.1)$, $(F.2)$, $(F.4)$, $(K.1-K.2)$, $(H.1-H.3)$, $(B.1)$, (Θ_1) *et quand* $n\rightarrow\infty$,

$$
\begin{aligned}
&\sup_{x\in I}\pm\left\{\frac{nH_{n}\left(x\right)}{2\log_{\theta,K}\left(\left|I\right|/\ H_{n}\left(x\right)\right)}\right\}^{\frac{1}{2}}\left\{\Theta_{n}\left(x\right)r_{\Psi;n}(x;H_{n}\left(x\right))\right. \qquad (2.17)\\
&\left.-\hat{E}\ r_{\Psi;n}(x;H_{n}\left(x\right))\right\}\xrightarrow{P}\left[\sup_{x\in I}\left\{\frac{\Theta^{2}\left(x\right)\sigma_{\Psi}^{2}\left(x\right)}{f_{X}\left(x\right)}\right\}\int_{\mathbb{R}}K^{2}\left(t\right)dt\right]^{\frac{1}{2}}.
\end{aligned}
$$

Si, de plus la condition $(B.2)$ est vérifiée,

$$\left\{\frac{nh_n}{2\log_{\theta,K}\left(|I|/\,h_n\right)}\right\}^{\frac{1}{2}}\sup_{x\in I}\pm\Theta_n\left(x\right)\left\{r_{\Psi;n}(x;H_n\left(x\right))\right. \tag{2.18}$$

$$\left.-\hat{E}\ r_{\Psi;n}(x;H_n\left(x\right))\right\}\xrightarrow{P}\left[\sup_{x\in I}\left\{\frac{\Theta^2\left(x\right)\sigma_\Psi^2\left(x\right)}{f_X\left(x\right)C\left(x\right)}\right\}\int_{\mathbb{R}}K^2\left(t\right)dt\right]^{\frac{1}{2}}.$$

Théorème 2.5. (Deheuvels et Mason (2004)). *Sous les hypothèses* $(F.1-F.2)$, $(K.1-K.2)$, $(H.1-H.3)$, $(B.1)$, (Θ_1) *et quand* $n\to\infty$,

$$\sup_{x\in I}\pm\left\{\frac{nH_n\left(x\right)}{2\log_{\theta,K}\left(|I|/\,H_n\left(x\right)\right)}\right\}^{\frac{1}{2}}\left\{\Theta_n\left(x\right)f_{X;n}(x;H_n\left(x\right))\right. \tag{2.19}$$

$$\left.-E\ f_{X;n}(x;H_n\left(x\right))\right\}\xrightarrow{P}\left[\sup_{x\in I}\left\{\Theta^2\left(x\right)f_X\left(x\right)\right\}\int_{\mathbb{R}}K^2\left(t\right)dt\right]^{\frac{1}{2}}.$$

Si de plus la condition $(B.2)$ *est vérifiée,*

$$\left\{\frac{nh_n}{2\log_{\theta,K}\left(|I|/\,h_n\right)}\right\}^{\frac{1}{2}}\sup_{x\in I}\pm\Theta_n\left(x\right)\left\{f_{X;n}(x;H_n\left(x\right))\right. \tag{2.20}$$

$$\left.-E\ f_{X;n}(x;H_n\left(x\right))\right\}\xrightarrow{P}\left[\sup_{x\in I}\left\{\frac{\Theta^2\left(x\right)f_X\left(x\right)}{C\left(x\right)}\right\}\int_{\mathbb{R}}K^2\left(t\right)dt\right]^{\frac{1}{2}},$$

où $|I|=b-a$ et $\log_{\theta,K}\left(|I|/\,h_n\right)=\log\left(\theta\vee u\left\{\int_R K^2\left(t\right)dt\right\}\right)$, $\theta=7$ (voir la remarque (2.4)).

Remarque 2.1. *Pour un choix convenable de la fonction aléatoire* $L_n\left(x\right)=L_n\left(X_1,\ldots,X_n;x\right)$, *les relations du Théorème 2.5 peuvent s'écrire sous la forme*

$$\sup_{x\in I}\pm\left\{\frac{1}{L_n\left(x\right)}\right\}\left\{f_{X;n}(x;H_n\left(x\right))-E\ f_{X;n}(x;H_n\left(x\right))\right\}\xrightarrow{P}1 \tag{2.21}$$

et si de plus $H_n\left(x\right)$ *est choisi de telle façon que le biais de* $f_X\left(x\right)$ *soit négligeable, c'est à dire*

$$\sup_{x\in I}\pm\left\{\frac{1}{L_n\left(x\right)}\right\}\left\{E\ f_{X;n}(x;H_n\left(x\right))-f_X\left(x\right)\right\}\xrightarrow{P}0, \tag{2.22}$$

alors ces deux relations combinées ensemble donnent, asymptotiquement

$\forall\varepsilon>0$

$$P\left(f_X\left(x\right)\in\left[f_{X;n}(x;H_n\left(x\right))-\left(1+\varepsilon\right)L_n\left(x\right)\right.,\right. \tag{2.23}$$

$$\left.\left.f_{X;n}(x;H_n\left(x\right))+\left(1+\varepsilon\right)L_n\left(x\right)\right]\ ,\forall x\in I\right)\longrightarrow 1,$$

$$P\left(f_X(x) \in \left[f_{X;n}(x; H_n(x)) - (1-\varepsilon) L_n(x)\right.,\right. \tag{2.24}$$
$$\left.\left. f_{X;n}(x; H_n(x)) + (1-\varepsilon) L_n(x)\right], \forall x \in I\right) \longrightarrow 0.$$

Comme les relations (2.23) et (2.24) sont vérifiées simultanément $\forall \varepsilon > 0$, *nous disons que :* $\forall x \in I$,

$$[A_n(x), \; B_n(x)] = [f_{X;n}(x; H_n(x)) - L_n(x), \; f_{X;n}(x; H_n(x)) + L_n(x)],$$

constituent des bandes de confiance optimales simultanées asymptotiques pour $f_X(x)$. *D'une manière analogue nous obtenons des bandes de confiance pour* $r(x)$.

En pratique $\forall x \in I$, *la surface entre* $A_n(x)$ *et* $B_n(x)$ *fournie une information utile sur les valeurs inconnues de* $f_X(x)$ *(voir Derzko et Deheuvels (2002) pour l'exemple en biomédical).*

Remarque 2.2. *Dans le Théorème 2.4 la condition* (H_3) *est une condition nécessaire et suffisante (voir la proposition A1 de Deheuvels et Mason(2004)).*

Remarque 2.3. *Comme* $\theta > 1$, $\log_{\theta,K}(|I|/h_n)$ *peut être remplaçé par* $\log(|I|/h_n)$, *d'ailleurs les démonstrations de ces théorèmes sont formulées avec le dernier terme.*

Remarque 2.4. *Supposons que le but est d'obtenir un intervalle de confiance de* $f_X(x_0)$ *(*x_0 *un point de* I*) de niveau de confiance égal à* 95%. *Supposant de plus que le comportement limite de* h_n *assure que le biais de* $f_X(x_0)$ *tend vers zéro, dans ce cas d'après la section 4.2 de Prakasa Rao (1983), on a*

$$P\left(f_X(x_0) \in \left[f_{X;n}(x_0; h_n) - (1.96) \times \left\{\frac{f_X(x_0)}{nh_n} \int_{\mathbb{R}} K^2(t)\,dt\right\}^{\frac{1}{2}},\right.\right.$$
$$\left.\left. f_{X;n}(x_0; h_n) + (1.96) \times \left\{\frac{f_X(x_0)}{nh_n} \int_{\mathbb{R}} K^2(t)\,dt\right\}^{\frac{1}{2}}\right]\right) \approx 95\%, \quad \text{quand } n \to \infty, \tag{2.25}$$

où 1.96 *est la valeur du quantile d'ordre* 2,5% *de la loi normale centrée et réduite. En comparant ce résultat avec les bandes de confiance simultanées sur* I, *avec un niveau de* *confiance asymptotique* 100%, *à savoir pour chaque de* $\varepsilon > 0$,

$$P\left(f_X(x) \in \left[f_{X;n}(x;h_n) - (1+\varepsilon) \times \sqrt{2\log_{\theta,K}(|I|/h_n)} \times \sup_{x\in I}\left\{\frac{f_X(x)}{nh_n}\int_{\mathbb{R}} K^2(t)\,dt\right\}^{\frac{1}{2}},\right.\right.$$

$$\left.\left. f_{X;n}(x;h_n) + (1+\varepsilon) \times \sqrt{2\log_{\theta,K}(|I|/h_n)} \sup_{x\in I}\left\{\frac{f_X(x)}{nh_n}\int_{\mathbb{R}} K^2(t)\,dt\right\}^{\frac{1}{2}}\right]\right) \approx 100\% \quad (2.26)$$

quand $n \to \infty$,

nous constatons alors que les bandes de confiance de (2.26) contiennent les intervalles de confiance de la relation (2.25) et comme $(2\log 7)^{\frac{1}{2}} \simeq 1.97$, *il est préférable donc de choisir* $\theta \geq 7$.

Remarque 2.5. *Supposons que* $f_X(x) = \frac{\exp\left(\frac{-x^2}{2}\right)}{\sqrt{2\pi}}$, $K(t) = (1-t^2)\,1_{\{|t|\leq 1\}}$ *le noyau de Bartlett-Epanechnikov (Bartlett (1963), Epanechnikov (1969)) et* $I = [-2, 2]$. *Dans ce cas, le choix de* h_n *est*

$$h_n = \left(\frac{15e\sqrt{2\pi}}{8n}\right)^{\frac{1}{5}} = (1.6644) \times n^{\frac{-1}{5}}. \quad (2.27)$$

(Voir p.18 dans Berlinet et Devroye (1994)). h_n *ainsi choisi donne,*

pour $\theta = 7$, $\frac{|I|}{h_n}\left\{\int_R K^2(t)\,dt\right\} \geq 7 \Rightarrow h_n \leq \frac{12}{35} \Rightarrow n \geq 2697$,

pour $\theta = 15$, $\frac{|I|}{h_n}\left\{\int_R K^2(t)\,dt\right\} \geq 15 \Rightarrow h_n \leq \frac{4}{25} \Rightarrow n \geq 121839$.

Une telle taille d'un échantillon est rarement rencontrée, donc le meilleur choix de θ *est* 7.

La situation la plus fréquente de la fenêtre optimale $H_n(x)$ est quand elle est aléatoire et ne dépend pas de x. Dans ce cas constant de la fenêtre, si nous posons $H_n = H_n(X_1, \ldots, X_n)$ (indépendant de x), la condition (B_2) s'écrit :

(B_3) $\left\{\frac{H_n}{h_n} - c\right\} = o_p(1)$, où c est une constante positive.

Les h_n de toutes les méthodes d'optimisation citées antérieurement vérifient la condition (B_3), par conséquent le corollaire suivant des Théorèmes 2.4 et 2.5 peut s'appliquer à ces derniers.

Corollaire 2.1. *Supposons que les conditions* $(F.1)$, $(F2)$, $(F4)$, $(K.1 - K.3)$, $(H.1 - H.3)$,

(Θ_1), (B_3) *sont vérifiées et si de plus*

$$P\left(cn^{-\delta} \leq H_n \leq dn^{-\delta}\right),$$

où c, d *et* δ *sont des constantes telles que* $0 < c < d < \infty$ *et* $\frac{1}{5} \leq \delta < 1$,

alors

$$\left\{\frac{nH_n}{2\log_{\theta,K}\left(|I|/H_n\right)}\right\}^{\frac{1}{2}} \sup_{x\in I} \pm\Theta_n(x)\left\{f_{X;n}(x;H_n)\right. \tag{2.28}$$

$$\left. -\hat{E}\, f_{X;n}(x;H_n)\right\} \xrightarrow{P} \left[\sup_{x\in I}\left\{\Theta^2(x)\, f_X(x)\right\}\int_{\mathbb{R}} K^2(t)\,dt\right]^{\frac{1}{2}}$$

et

$$\left\{\frac{nH_n}{2\log_{\theta,K}\left(|I|/H_n\right)}\right\}^{\frac{1}{2}} \sup_{x\in I} \pm\left\{\Theta_n(x)\, r_{\Psi;n}(x;H_n)\right. \tag{2.29}$$

$$\left. -\hat{E} r_{\Psi;n}(x;H_n)\right\} \xrightarrow{P} \left[\sup_{x\in I}\left\{\frac{\Theta^2(x)\,\sigma_\Psi^2(x)}{f_X(x)}\right\}\int_{\mathbb{R}} K^2(t)\,dt\right]^{\frac{1}{2}}.$$

Nous n'exposons pas la preuve de ces théorèmes, mais nous donnons le théorème général sur les lois limites des fonctionnelles, établi par Deheuvels et Mason (2004), dont les corollaires conduisent aux démonstrations recherchées.

Soient $c(.)$, $d(.)$ deux fonctions continues sur J et λ, λ_1, λ_2 des constantes telles que $0 < \lambda_1 \leq 1 \leq \lambda_2 < \infty$ et soit le processus suivant

$$\begin{aligned} W_{\lambda,n}(x,\Psi) &= \sum_{j=1}^{n}\left[c(x)\,\Psi(Y_j) + d(x)\right] K\left(\frac{x-X_j}{\lambda h_n}\right) \\ &\quad - nE\left\{\left(c(x)\,\Psi(Y) + d(x)\right) K\left(\frac{x-X}{\lambda h_n}\right)\right\}, \end{aligned}$$

alors,

Théorème 2.6. *Si les conditions* $(F.1-F.4)$, $(K.1)$, $(H.1-H.3)$ *sont vérifiées, nous avons*

$$\sup_{\lambda_1\leq\lambda\leq\lambda_2}\left|\left\{2n\lambda\log\left(\frac{1}{h_n}\right)\right\}^{-\frac{1}{2}}\sup_{x\in I}\left\{\pm W_{\lambda,n}(x,\Psi)\right\} - \sigma(\Psi)\right| = o_p(1),$$

$$\textit{quand } n \to \infty,$$

où

$$\sigma^2(\Psi) = \sup_{x\in I} E\left\{\left(c(x)\,\Psi(Y) + d(x)\right)^2 \mid X = x\right\} f_X(x)\int_{\mathbb{R}} K^2(t)\,dt.$$

Corollaire 2.2. *Si les conditions* $(F.1-F.4)$, $(K.1)$, $(H.1-H.3)$ *sont vérifiées, nous avons quand* $n\to\infty$,

$$\sup_{\lambda_1\le\lambda\le\lambda_2}\left|\sup_{x\in I}\left\{\frac{n\lambda h_n}{2\log(1/h_n)}\right\}^{\frac{1}{2}}\pm\left\{m_{\Psi;n}(x;\lambda h_n)-Em_{\Psi;n}(x;\lambda h_n)\right\}-\sigma_{a,b}\right|=o_p(1),$$

où

$$\sigma_{a,b}=\sup_{x\in I}\left(\sigma_\Psi(x)\left\{f_X(x)\int_{\mathbb{R}}K^2(t)\,dt\right\}^{\frac{1}{2}}\right).$$

Corollaire 2.3. *Si les conditions* $(F.2)$, $(K.1-K.2)$, $(H.1-H2)$ *sont vérifiées, nous avons,*

$$\sup_{\lambda_1\le\lambda\le\lambda_2}\left|\sup_{x\in I}\left\{\frac{n\lambda h_n}{2\log(1/h_n)}\right\}^{\frac{1}{2}}\pm\left\{f_{X;n}(x;\lambda h_n)-Ef_{X;n}(x;\lambda h_n)\right\}-f_{a,b}\right|=o_p(1),$$

quand $n\to\infty$,

où

$$f_{a,b}=\sup_{x\in I}\left\{f_X(x)\int_{\mathbb{R}}K^2(t)\,dt\right\}^{\frac{1}{2}}.$$

Corollaire 2.4. *Si les conditions* $(F.1)$, $(F.2)$, $(F.4)$, $(K.1-K.2)$, $(H.1-H.3)$ *sont vérifiées, alors nous avons , quand* $n\to\infty$,

$$\sup_{\lambda_1\le\lambda\le\lambda_2}\left|\left\{\frac{n\lambda h_n}{2\log(1/h_n)}\right\}^{\frac{1}{2}}\sup_{x\in I}\pm\left\{r_{\Psi;n}(x;\lambda h_n)-\hat{E}\,r_{\Psi;n}(x;\lambda h_n)\right\}-\tau_{a,b}\right|=o_p(1),$$

où

$$\tau_{a,b}=\sup_{x\in I}\left\{\frac{\sigma_\Psi^2}{f_X(x)}\int_{\mathbb{R}}K^2(t)\,dt\right\}^{\frac{1}{2}}.$$

Les Corollaires 2.2 et 2.3 sont des cas particuliers du Théorème 2.6. Il est clair que les Théorèmes 2.4 et 2.5 sont des variétés de ces corollaires et ils sont prouvés avec les mêmes arguments que les Corollaires 2.3 et 2.4 (voir p. 249.de Deheuvels et Mason (2004)).

Chapitre III

ASYMPTOTIC CONFIDENCE BANDS FOR NONPARAMETRIC FUNCTION ESTIMATES IN THE GAUSSIAN CASE

Abstract. In this paper, we obtain asymptotic confidence bands for both the density and regression functions in the framework of nonparametric estimation. Before hand, the asymptotic behaviours in probability of the kernel estimator of the density and the Nadaraya-Watson estimator of the regression function are described while local and global optimal smoothing parameters are investigated.

Key Words : Confidence bounds, density estimation, kernel estimation, nonparametric estimation, regression estimation.

3.1 Introduction

Let $(X_1, Y_1), (X_2, Y_2), \ldots,$ be independent and identically distributed random replicates of the random vector $(X, Y) \in \mathbb{R}^2$. Let us assume that the marginal distributions of X and Y are centered normal with standard deviations $\sigma_X > 0$ and $\sigma_Y > 0$ respectively, and an unknown linear correlation coefficient ρ. In the sequel, the following notations and assumptions will be adopted.

$I = [a, b]$ and $J = [a', b']$ denote two fixed intervals of $\mathbb{R}$ such that

$-\infty < a' < a < b < b' < +\infty$. $|I|$ denotes the Lebesgue measure of I.

K is a real kernel weight function satisfying the folllowing conditions

$(K.1)$ For every $x \in \mathbb{R}$, $K(x) = K(-x)$;

$(K.2)$ $\int_{\mathbb{R}} K(t)\, dt = 1$;

$(K.3)$ $[t^2K] = \int_{\mathbb{R}} t^2 K(t) dt < \infty$;

$(K.4)\quad [K^2] = \int_{\mathbb{R}} K^2(t)dt < \infty;$

$(K.5)\quad K(u) = 0$ for $u \notin \left[-\frac{\alpha}{2}, \frac{\alpha}{2}\right)$, where $\alpha \in]0, \infty[$.

For every $u \in \mathbb{R}$, set $\log_{\theta,K}(u) = \log\left(\theta \vee u\left\{\int K^2(t)dt\right\}\right)$, where $\theta > 1$ is a specified constant.

The regression function and the conditional variance of Y given $X = x$ are defined for $x \in J$ by

$$r(x) = E(Y \mid X = x) = \int_{\mathbb{R}} y \frac{f_{X,Y}(x,y)}{f_X(x)} dy = \rho \frac{\sigma_X}{\sigma_Y} x,$$

$$v^2(x) = \mathrm{Var}(Y \mid X = x) = \int_{\mathbb{R}} [y - E(Y \mid X = x)]^2 \frac{f_{X,Y}(x,y)}{f_X(x)} dy = (1-\rho^2)\sigma_Y^2,$$

where $f_{X,Y}(x,y)$ is the joint density of (X,Y) and $f_X(x)$ is the marginal density of the random variable X.

Let $(h_n)_{n\geq 1}$ be a sequence of real numbers. Following Rosenblatt (1956), Parzen (1962), Nadaraya (1964) and Watson (1964), the kernel estimators of the functions $f_X(x)$, $r(x)$ and $v^2(x)$ are defined respectively by

$$f_{X,n}(x; h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

$$r_n(x; h_n) = \begin{cases} \dfrac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} & \text{if} \quad \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \neq 0, \\[2ex] \dfrac{1}{n}\sum_{i=1}^{n} Y_i & \text{if} \quad \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) = 0, \end{cases}$$

and

$$v_n^2(x; h_n) = \begin{cases} \dfrac{\sum_{i=1}^{n} [Y_i - r_n(x;h_n)]^2 K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} & \text{if} \quad \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \neq 0, \\[2ex] \dfrac{1}{n}\sum_{i=1}^{n} \left(Y_i - \dfrac{1}{n}\sum_{j=1}^{n} Y_j\right)^2 & \text{if} \quad \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) = 0. \end{cases}$$

Introduce now the following centering factors

$$Ef_{X,n}(x;h_n) = \frac{1}{h_n} E\left[K\left(\frac{x-X}{h_n}\right)\right],$$

$$\widehat{E}r_n(x;h_n) = \begin{cases} \dfrac{E\left[YK\left(\frac{x-X}{h_n}\right)\right]}{E\left[K\left(\frac{x-X}{h_n}\right)\right]} & \text{if} \quad E\left[K\left(\frac{x-X}{h_n}\right)\right] \neq 0, \\ \\ E(Y) & \text{if} \quad E\left[K\left(\frac{x-X}{h_n}\right)\right] = 0, \end{cases}$$

which are needed hereafter in our study.

Recall that whenever $\lim_{n\to\infty} h_n = 0$, we have

$$\lim_{n\to\infty} \left\{ E\left[f_{X,n}(x;h_n)\right] - f(x) \right\} = 0 .$$

If, in addition $\lim_{n\to\infty} nh_n = \infty$, then we have

$$\lim_{n\to\infty} \left\{ \widehat{E}\left[r_n(x;h_n)\right] - E\left[r_n(x;h_n)\right] \right\} = 0 .$$

From the seminal works by Rosenblatt (1956), Parzen (1962), Nadaraya (1964) and Watson (1964), nonparametric function estimation has been widely investigated. Theoretical apsects and properties have been described and numerical implementations have been performed. For an overview on the question, we refer to Prakasa Rao (1983), Devroye & Györfi (1985), Wand & Jones (1985), Bosq & Lecoutre (1987), Silverman (1986), Nadaraya (1989), Scott (1992) and the references therein.

The behaviour of nonparametric estimators depends strongly upon the smoothing parameter. Several procedures have been proposed describing how to choose this parameter. We quote the cross-validation and the plug-in methods related essentially to the mean square error criterion. Asymptotic properties of estimators need also conditions upon the smoothing parameter. The minimal conditions to set are $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. It is well known that these conditions are necessary and sufficient for the pointwise weak consistency of the kernel estimator of f_X and the Nadaraya-Watson estimator of the regression function r. Moreover, these conditions allow to state a rate of convergence of $\widehat{E}\Big(r_n(x,h_n)\Big)$

towards $r(x)$, i.e. we have

$$\widehat{E}\Big(r_n(x,h_n)\Big) - r(x) = 0\left(\frac{1}{nh_n}\right) \quad \text{as} \quad n \to \infty .$$

Recently, Deheuvels & Mason (2004) obtained the rate of uniform convergence in probability for the kernel density estimator. Their result is stated in the following terms. If

$$h_n \to 0 \quad \text{and} \quad \frac{nh_n}{\log n} \to \infty , \quad \text{as } n \to \infty ,$$

then

$$\sqrt{\frac{nh_n}{2\log(\frac{1}{h_n})}} \sup_{x\in I} \left| \frac{f_{X,n}(x,h_n) - E\left[f_{X,n}(x,h_n)\right]}{\sqrt{2\left[K^2\right]^2 f_X(x)}} \right| \xrightarrow{P} 1 , \quad \text{as } n \to \infty ,$$

where I is a compact interval in $\mathbb{R}$. Deheuvels & Mason (2004) have also obtained the optimal rate of convergence in probability for the estimate $r_n(x,h_n)$ of the regression function r. These results allow them to construct asymptotic confidence bands for the functions f_X and r.

Consider now the problem related to the optimal choices $h_n(x)$ and h_n of the smoothing parameter with respect to both the mean square error and the integrated mean square error criteria. Our aim is to construct the estimates $\widehat{h}_n(x)$ and $\widehat{h}_n$ of the optimal parameter for these criteria in terms of the empirical estimators $\widehat{\sigma}_X, \widehat{\sigma}_Y, \widehat{\rho}$, such that

$$\frac{\widehat{h}_n(x)}{h_n(x)} \xrightarrow{P} 1 \quad \text{and} \quad \frac{\widehat{h}_n}{h_n} \xrightarrow{P} 1 , \text{ as } n \to \infty .$$

Confidence bounds for $f_X(x)$ (respectively $r(x)$) in terms of $f_{X,n}(x;\widehat{h}_n(x))$ (respectively $r_n(x;\widehat{h}_n(x))$) are then deduced.

The remainder of the paper is organized as follows. In section 2, we present the results. The section 3 is devoted to the proofs.

3.2 Results

3.2.1 The mean square error of $f_{X,n}(x;h_n)$

According to Wand and Jones (1995), the mean square error of $f_{X,n}(x;h_n)$ is given by

$$E\left[f_{X,n}(x;h_n)-f_X(x)\right]^2=\mathrm{Var} f_{X,n}(x;h_n)+\left[E\left(f_{X,n}(x;h_n)\right)-f_X(x)\right]^2.$$

This quantity is minimized with respect to h_n, by

$$\begin{aligned}
H_{n,1}(x) &= n^{-1/5}\left(\frac{f_X(x)\left[K^2\right]}{\left(f_X''(x)\left[t^2K\right]\right)^2}\right)^{1/5}\\
&= n^{-1/5}\,\sigma_X\left(\frac{\sqrt{2\pi}\left[K^2\right]\exp\frac{1}{2}\left(\frac{x}{\sigma_X}\right)^2}{\left[t^2K\right]^2\left(\left(\frac{x}{\sigma_X}\right)^2-1\right)^2}\right)^{1/5}.
\end{aligned}$$

As σ_X^2 is unknown, we replace it by its empirical estimator

$$\widehat{\sigma}_X^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2,$$

with $\bar{X}$ the sample mean. Therefore, we obtain

$$\widehat{H}_{n,1}(x)=n^{-1/5}\widehat{\sigma}_X\left(\frac{\sqrt{2\pi}\left[K^2\right]\exp\frac{1}{2}\left(\frac{x}{\widehat{\sigma}_X}\right)^2}{\left[t^2K\right]^2\left[\left(\frac{x}{\widehat{\sigma}_X}\right)^2-1\right]^2}\right)^{1/5} \tag{3.1}$$

as an estimator of $h_n(x)$.

Set now

$$\Theta_{n,1}(x)=\sqrt{\frac{\widehat{\sigma}_X}{\left[K^2\right]}}\left(\frac{\sqrt{(2\pi)^3}\left[K^2\right]^{1/2}\exp\frac{3}{2}\left(\frac{x}{\widehat{\sigma}_X}\right)^2}{\left[t^2K\right]\left|\left(\frac{x}{\widehat{\sigma}_X}\right)^2-1\right|}\right)^{1/5}. \tag{3.2}$$

Following Deheuvels & Mason (2004), one can see that both $\widehat{h}_{n,1}(x)=\widehat{H}_{n,1}(x)/\widehat{\sigma}_X$, $h_{n,1}(x)=H_{n,1}(x)/\sigma_X$ and $\Theta_{n,1}(x)$ satisfy the following conditions

$(B.1)$ for any $\varepsilon>0$,

$$P\left(\inf_{x\in I}h_{n,1}(x)-\varepsilon\, n^{-1/5}\leq\inf_{x\in I}\widehat{h}_{n,1}(x)\leq\sup_{x\in I}\widehat{h}_{n,1}(x)\leq\sup_{x\in I}h_{n,1}(x)+\varepsilon\, n^{-1/5}\right)\xrightarrow{P}1,$$

as $n \longrightarrow \infty$,

$(B.2)$ for any $\varepsilon > 0$,

$$P\left(\sup_{x\in I}\left|\frac{\widehat{h}_{n,1}(x)}{h_n} - \frac{h_{n,1}(x)}{h_n}\right| \geq \varepsilon\right) \longrightarrow 0 \quad \text{as } n \longrightarrow \infty,$$

$(\Theta.1)$ for any $\varepsilon > 0$,

$$P\left(\sup_{x\in I}\left|\frac{\Theta_{n,1}(x)}{\Theta_1(x)} - 1\right| > \varepsilon\right) \longrightarrow 0 \quad \text{as } n \longrightarrow \infty,$$

where

$$\Theta_1(x) = \sqrt{\frac{\sigma_X}{[K^2]}}\left(\frac{\sqrt{(2\pi)^3}\,[K^2]^{1/2}\exp\frac{3}{2}\left(\frac{x}{\sigma_X}\right)^2}{[t^2K]\left|\left(\frac{x}{\sigma_X}\right)^2 - 1\right|}\right)^{1/5}. \tag{3.3}$$

3.2.2 The integrated mean square error of $f_{X,n}(x;h_n)$

Wand et Jones (1995) stated that the integrated mean square error relative to $f_{X,n}(x;h_n)$ is given by

$$\int_{\mathbb{R}} E\left[f_{X,n}(x;h_n) - f_X(x)\right]^2 dx = \frac{1}{n}\left[t^2K\right] + \int_{\mathbb{R}}\left\{f_X''(x)\left[t^2K\right]\right\}^2 dx + o(h_n^2).$$

The minimum of this quantity, with respect to h_n, is

$$\begin{aligned} h_{n,2} &= n^{-1/5}\left(\frac{[K^2]}{[t^2K]^2\int_{\mathbb{R}}\left[f_X''(x)\right]^2 dx}\right)^{1/5} \\ &= n^{-1/5}\sigma_X\left(\frac{8\sqrt{\pi}\,[K^2]}{3\,[t^2K]^2}\right)^{1/5}. \end{aligned}$$

Define

$$\widehat{h}_{n,2} = n^{-1/5}\,\widehat{\sigma}_X\left(\frac{8\sqrt{\pi}\,[K^2]}{3\,[t^2K]^2}\right)^{1/5} \tag{3.4}$$

and

$$\Theta_{n,2}(x) = \left\{\frac{[K^2]}{\sqrt{2\pi}\widehat{\sigma}_X}\exp -\frac{1}{2}\left(\frac{x}{\widehat{\sigma}_X}\right)^2\right\}^{-1/2}. \tag{3.5}$$

As in the corollary of Deheuvels & Mason (2004), one can see that $\widehat{h}_{n,2}$ satisfies the condition

$$P\left(h_{n,2}-\varepsilon n^{-1/5}<\widehat{h}_{n,2}<h_{n,2}+\varepsilon n^{-1/5}\right)\longrightarrow 1 \; as \;\; n\to\infty, \forall\varepsilon>0. \tag{3.6}$$

Moreover, taking $\Theta_2(x)=\frac{1}{\sqrt{f_X(x)[K^2]}}$, we have

$$P\left(\sup_{x\in I}\left|\frac{\Theta_{n,2}(x)}{\Theta_2(x)}-1\right|>\varepsilon\right)\longrightarrow 0, \qquad \text{as } n\to\infty \;\; \forall\varepsilon>0\,.$$

3.2.3 The mean square error of $r_n(x;h_n)$

In a similar way, according to Wand and Jones (1995), the value of $h_{n,3}$ that minimizes the mean square error of $r_n(x;h_n)$ given by

$$E\left[r_n(x;h_n)-r(x)\right]^2=\operatorname{Var}\left[r_n(x;h_n)\right]+\left\{E\left[r_n(x;h_n)\right]-r(x)\right\}^2$$

is

$$\begin{aligned} H_{n,3}(x) &= n^{-1/5}\left(\frac{\frac{v(x)}{f_X(x)}\left[K^2\right]}{\left[r''(x)+2\,r'(x)\frac{f_X'(x)}{f_X(x)}\left[t^2K\right]\right]^2}\right)^{1/5} = \\ & n^{-1/5}\,\sigma_X\left(\frac{\sqrt{2\pi}\left[K^2\right]\left(1-\rho^2\right)\sigma_X^2\exp\frac{1}{2}\left(\frac{x}{\sigma_X}\right)^2}{4\left[t^2K\right]^2\rho^2\,x^2}\right)^{1/5}. \end{aligned}$$

As σ_X and ρ are unknown parameters, we replace them by their empirical estimators, i.e., $\widehat{\sigma}_X$ and

$$\widehat{\rho}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\widehat{\sigma}_X\widehat{\sigma}_Y},$$

respectively.

Therefore, $h_{n,3}(x)$ is estimated by

$$\widehat{H}_{n,3}(x)=n^{-1/5}\,\widehat{\sigma}_X\left(\frac{\sqrt{2\pi}\left[K^2\right]\left(1-\widehat{\rho}^2\right)\widehat{\sigma}_X^2\exp\frac{1}{2}\left(\frac{x}{\widehat{\sigma}_X}\right)^2}{4\left[t^2K\right]^2\widehat{\rho}^2\,x^2}\right)^{1/5}. \tag{3.7}$$

Set now

$$\Theta_{n,3}(x) = \left(\frac{\left(1-\widehat{\rho}^2\right)^{-2} |\widehat{\rho}\, x|^{-1} \exp - \left(\frac{x}{\widehat{\sigma}_X}\right)^2}{4\pi\, \widehat{\sigma}_X^{3/2}\, \widehat{\sigma}_Y^5\, [K^2]^2\, [t^2 K]} \right)^{1/5}, \tag{3.8}$$

as the estimator of the quantity

$$\Theta_3(x) = \left(\frac{\left(1-\rho^2\right)^{-2} |\rho x|^{-1} \exp - \left(\frac{x}{\sigma_X}\right)^2}{4\pi\, \sigma_X^{3/2}\, \sigma_Y^5\, [K^2]^2\, [t^2 K]} \right)^{1/5}. \tag{3.9}$$

Similarly, as above, one may show that $\widehat{h}_{n,3}(x) = \widehat{H}_{n,3}(x)/\widehat{\sigma}_X$, $h_{n,3}(x) = H_{n,3}(x)/\sigma_X$ and $\Theta_{n,3}(x)$ satisfy the conditions $(B.1)$, $(B.2)$, $(\Theta.1)$.

The asymptotic behaviours of the estimates of the functions f_X and r are described in the following theorems.

Theorem 3.1 *Let $\widehat{H}_{n,1}(x)$, $\Theta_{n,1}(x)$ be the estimates given in the statements (3.1) and (3.2) respectively. Then, we have*

$$\left(\frac{n^{4/5}\, \widehat{\sigma}_X}{2 \log_{\theta,K}\left(\frac{|I|}{n^{-1/5}\, \widehat{\sigma}_X}\right)} \right)^{1/2} \sup_{x\in I} \pm\, \Theta_{n,1}(x) \left(f_{X,n}(x; \widehat{H}_{n,1}(x)) - f_X(x) \right) \xrightarrow{P} 1,$$

as $n \to \infty$.

Remark.3.1 *Note from Theorem 3.1 that, for any $0 < \varepsilon < 1$, as $n \to \infty$, we have for any $x \in I$,*

$$\lim_{n\to\infty} P\Big(f_X(x) \in \Big[f_{X,n}(x; \widehat{H}_{n,1}(x)) - (1+\varepsilon)\Delta_{n,1}(x)\ , \\ f_{X,n}(x; \widehat{H}_{n,1}(x)) + (1+\varepsilon)\Delta_{n,1}(x) \Big] \Big) = 1$$

and

$$\lim_{n\to\infty} P\Big(f_X(x) \in \Big[f_{X,n}(x; \widehat{H}_{n,1}(x)) - (1-\varepsilon)\Delta_{n,1}(x)\ , \\ f_{X,n}(x; \widehat{H}_{n,1}(x)) + (1-\varepsilon)\Delta_{n,1}(x) \Big] \Big) = 0,$$

where

$$\Delta_{n,1}(x) = \frac{1}{\Theta_{n,1}(x)} \left(\frac{2 \log_{\theta,K}\left(\frac{|I|}{n^{-1/5}\, \widehat{\sigma}_X}\right)}{n^{4/5}\, \widehat{\sigma}_X} \right)^{1/2}.$$

Therefore, the interval

$$\Big[f_{X,n}(x;\widehat{H}_{n,1}(x)) - \Delta_{n,1}(x)\ ,\ f_{X,n}(x;\widehat{H}_{n,1}(x)) + \Delta_{n,1}(x)\ \Big]$$

stands as an asymptotic confidence domain for $f_X(x)$.

Theorem 3.2 *Let* $\widehat{h}_{n,2}$ *and* $\Theta_{n,2}(x)$ *be as given in the statements (3.4) and (3.5) respectively. Then, we have*

$$\left(\frac{n\widehat{h}_{n,2}}{2\log_{\theta,K}\left(\frac{|I|}{\widehat{h}_{n,2}}\right)}\right)^{1/2} \sup_{x\in I} \pm\,\Theta_{n,2}(x)\left(f_{X,n}(x;\widehat{h}_{n,2}) - f_X(x)\right) \xrightarrow{P} 1\,,$$

as $n \longrightarrow \infty$.

Theorem 3.3 *Let* $\widehat{H}_{n,3}(x)$ *and* $\Theta_{n,3}(x)$ *be as given in the statements (3.7) and (3.8) respectively. Then, we have*

$$\left(\frac{n^{4/5}\,\widehat{\sigma}_X}{2\log_{\theta,K}\left(\frac{|I|}{n^{-1/5}\,\widehat{\sigma}_X}\right)}\right)^{1/2} \sup_{x\in I} \pm\,\Theta_{n,3}(x)\Big\{r_n(x;\widehat{H}_{n,3}(x)) - r(x)\Big\} \xrightarrow{P} 1\,,$$

as $n \longrightarrow \infty$.

Remark.3.2 *Note from Theorem 3.3 that, for any* $0 < \varepsilon < 1$, *we have for any* $x \in I$

$$\lim_{n\longrightarrow\infty} P\Big(r(x) \in \Big[r_n(x;\widehat{H}_{n,3}(x)) - (1+\varepsilon)\Delta_{n,2}(x)\,,\\ r_n(x;\widehat{H}_{n,3}(x)) + (1+\varepsilon)\Delta_{n,3}(x)\Big]\ \Big) = 1$$

and

$$\lim_{n\longrightarrow\infty} P\Big(r(x) \in \Big[r_n(x;\widehat{H}_{n,3}\,(x)) - (1-\varepsilon)\Delta_{n,2}(x)\,,\\ r_n(x;\widehat{H}_{n,3}\,(x)) + (1-\varepsilon)\Delta_{n,3}(x)\Big]\ \Big) = 0\,,$$

where

$$\Delta_{n,2}\,(x) = \frac{1}{\Theta_{n,3}(x)}\left(\frac{2\log_{\theta,K}\left(\frac{|I|}{n^{-1/5}\,\widehat{\sigma}_X}\right)}{n^{4/5}\,\widehat{\sigma}_X}\right)^{1/2}.$$

Therefore, we obtain for any $x \in I$, *the following asymptotic confidence bounds for the regression function*

$$\Big[r_n(x;\widehat{H}_{n,3}(x)) - \Delta_{n,2}(x)\,,\ \ r_n(x;\widehat{H}_{n,3}(x)) + \Delta_{n,2}(x)\ \Big]\,.$$

3.3 Proofs

Proof of Theorem 3.1. Taking $h_n = n^{-1/5}$, it follows that the conditions $(B.1)$, $(B.2)$, $(\Theta.1)$ are satisfied, Theorem 2.5 implies

$$\left(\frac{nh_n}{2\log_{\theta,K}\left(\frac{|I|}{h_n}\right)}\right)^{1/2} \sup_{x\in I} \pm\, \Theta_{n,1}(x)\left\{Ef_{X,n}\left(x;\widehat{h}_{n,1}(x)\right) - f_{X,n}\left(x;\widehat{h}_{n,1}(x)\right)\right\}$$
$$\xrightarrow{P} \left\{\sup_{x\in I}\frac{\Theta_1^2(x)\, f_X(x)}{C_1(x)}[K^2]\right\}^{1/2},$$

as $n \to \infty$, where

$$C_1(x) = \left\{\frac{\sqrt{2\pi}\,[K^2]\exp\frac{1}{2}\left(\frac{x}{\sigma_X}\right)^2}{[t^2K]^2\left[\left(\frac{x}{\sigma_X}\right)^2 - 1\right]^2}\right\}^{1/5}. \qquad (3.10)$$

Taking into account the expressions $\Theta_1(x)$ and $C_1(x)$ given by (3.3) and (3.10) respectively, we obtain

$$\left(\frac{nh_n}{2\log_{\theta,K}\left(\frac{|I|}{h_n}\right)}\right)^{1/2} \sup_{x\in I} \pm\, \Theta_{n,1}(x)\left\{Ef_{X,n}\left(x;\widehat{h}_{n,1}(x)\right) - f_{X,n}\left(x;\widehat{h}_{n,1}(x)\right)\right\} \xrightarrow{P} 1, \qquad (3.12)$$

as $n \to \infty$.

Now, with this choice of the parameter h_n, we have (see Nadaraya (1989))

$$\left(\frac{nh_n}{2\log_{\theta,K}\left(\frac{|I|}{h_n}\right)}\right)^{1/2} \sup_{x\in I} \Theta_{n,1}(x)\left|Ef_{X,n}(x;h_n) - f_X(x)\right| \xrightarrow{P} 0, \qquad (3.13)$$

as $n \to \infty$.

According to Corollary 2.3, for every fixed λ_1 and λ_2 such that $0 < \lambda_1 \le 1 \le \lambda_2 < +\infty$, we have, as $n \to \infty$,

$$\sup_{\lambda_1\le\lambda\le\lambda_2}\left|\left\{\frac{n\lambda h_n}{2\log(1/h_n)}\right\}^{1/2}\sup_{x\in I}\pm\left\{Ef_{X,n}(x,\lambda h_n) - f_{X,n}(x,\lambda h_n)\right\} - \sup_{x\in I}\left\{f_X(x)\int_{\mathbb{R}}K^2(t)\,dt\right\}^{1/2}\right| = o_P(1) \qquad (3.14)$$

Moreover, we have $\widehat{\sigma}_X \xrightarrow{P} \sigma_X \in]0,\infty[$. Therefore, using the statement (3.14), the expressions (3.12) and (3.13) may be rewritten as

$$\left(\frac{nh_n\widehat{\sigma}_X.}{2\log_{\theta,K}\left(\frac{|I|}{h_n\widehat{\sigma}_X}\right)}\right)^{1/2} \sup_{x\in I} \pm\, \Theta_{n,1}(x)\left\{Ef_{X,n}\left(x;\widehat{\sigma}_X\widehat{h}_{n,1}(x)\right)\right.$$
$$\left. -f_{X,n}\left(x;\widehat{\sigma}_X\widehat{h}_{n,1}(x)\right)\right\} \xrightarrow{P} 1$$

and

$$\left(\frac{nh_n\widehat{\sigma}_X}{2\log_{\theta,K}\left(\frac{|I|}{h_n\widehat{\sigma}_X}\right)}\right)^{1/2} \sup_{x\in I} \Theta_{n,1}(x)\left|Ef_{X,n}\left(x;\widehat{\sigma}_X\widehat{h}_{n,1}(x)\right) - f_X(x)\right| \xrightarrow{P} 0$$

respectively.

By combining the two last relations, we obtain

$$\left(\frac{n^{4/5}\,\widehat{\sigma}_X}{2\log_{\theta,K}\left(\frac{|I|}{n^{-1/5}\widehat{\sigma}_X}\right)}\right)^{1/2} \sup_{x\in I} \pm\, \Theta_{n,1}(x)\left\{f_{X,n}\left(x;\widehat{H}_{n,1}(x)\right) - f_X(x)\right\} \xrightarrow{P} 1\,,$$

as $n\longrightarrow\infty$, where $\widehat{H}_{n,1}(x) = \widehat{\sigma}_X\widehat{h}_{n,1}(x)$.

Proof of Theorem 3.2. This proof is based on Corollary 2.1, say for every constants c, d and δ such that $0 < c < d < +\infty$ and $1/5 \le \delta < 1$, if

$$P(cn^{-\delta} \le \mathcal{H}_n \le dn^{-\delta}) \longrightarrow 1\,, \quad \text{as } n\rightarrow\infty\,, \tag{3.15}$$

then

$$\left(\frac{n\mathcal{H}_n}{2\log_{\theta,K}\left(\frac{|I|}{\mathcal{H}_n}\right)}\right)^{1/2} \sup_{x\in I} \pm\,\Theta_{n,2}(x)\left\{f_{X,n}\left(x;\mathcal{H}_n\right) - f_X(x)\right\} \tag{3.16}$$
$$\xrightarrow{P} \left\{\sup_{x\in I} \pm\,\Theta_2(x) f_X(x)\int_{\mathbb{R}} K^2(t)\,dt\right\}^{1/2},$$

as $n\longrightarrow\infty$.

Now replacing $\mathcal{H}_n$ by $\widehat{h}_{n,2}$ given by (3.4) and according to (3.6), the condition (3.15) is satisfied. Thus (3.16) becomes

$$\left(\frac{n\widehat{h}_{n,2}}{2\log_{\theta,K}\left(\frac{|I|}{\widehat{h}_{n,2}}\right)}\right)^{1/2} \sup_{x\in I} \pm\,\Theta_{n,2}(x)\Big\{f_{X,n}(x;\widehat{h}_{n,2}) - f_X(x)\Big\}$$

$$\xrightarrow{P} \left\{\sup_{x\in I} \pm\,\Theta_2(x) f_X(x)\int_{\mathbb{R}} K^2(t)\,dt\right\}^{1/2},$$

i.e.

$$\left(\frac{n\widehat{h}_{n,2}}{2\log_{\theta,K}\left(\frac{|I|}{\widehat{h}_{n,2}}\right)}\right)^{1/2} \sup_{x\in I} \pm\,\Theta_{n,2}(x)\left\{f_{X,n}(x;\widehat{h}_{n,2}) - f_X(x)\right\} \xrightarrow{P} 1\,,$$

as $n \longrightarrow \infty$.

Proof of Theorem 3.3. Let

$$C_{n,3}(x) = \left\{\frac{\sqrt{2\pi}\,[K^2](1-\widehat{\rho}^2)\widehat{\sigma}_X^2 \exp\frac{1}{2}\left(\frac{x}{\widehat{\sigma}_X}\right)^2}{4[t^2K]^2\,\widehat{\rho}^2\,x^2}\right\}^{1/5} \tag{3.17}$$

be an estimator of

$$C_3(x) = \left\{\frac{\sqrt{2\pi}\,[K^2](1-\rho^2)\sigma_X^2 \exp\frac{1}{2}\left(\frac{x}{\sigma_X}\right)^2}{4[t^2K]^2\,\rho^2\,x^2}\right\}^{1/5}. \tag{3.18}$$

According to Theorem.2.4, we have

$$\left(\frac{nh_n}{2\log_{\theta,K}\left(\frac{|I|}{h_n}\right)}\right)^{1/2} \sup_{x\in I} \pm\,\Theta_{n,3}(x)\left\{\widehat{E}\,r_n(x;\widehat{h}_{n,3}) - r_n(x;\widehat{h}_{n,3})\right\}$$

$$\xrightarrow{P} \left\{\sup_{x\in I} \frac{\Theta_3^2(x)\,\sigma_I^2(x)}{C_3(x)\,f_X(x)}[K^2]\right\}^{1/2},$$

as $n \longrightarrow \infty$.

Replacing $\Theta_3(x)$ and $C_3(x)$ by the expressions given in (3.9) and (3.18) respectively, we obtain

$$\left(\frac{nh_n}{2\log_{\theta,K}\left(\frac{|I|}{h_n}\right)}\right)^{1/2} \sup_{x\in I} \pm\,\Theta_{n,3}(x)\left\{\widehat{E}\,r_n(x;\widehat{h}_{n,3}) - r_n(x;\widehat{h}_{n,3})\right\} \xrightarrow{P} 1\,, \tag{3.19}$$

as $n \longrightarrow \infty$, when $h_n = n^{-1/5}$.

In addition, for this same choice of h_n, we have

$$\left(\frac{nh_n}{2\log_{\theta,K}\left(\frac{|I|}{h_n}\right)}\right)^{1/2} \sup_{x\in I} \Theta_{n,3}(x)\left|\widehat{E}\, r_n(x;h_n) - r(x)\right| \xrightarrow{P} 0\,, \tag{3.20}$$

(see Nadaraya (1989)).

According to Corollary 2.4, for every fixed λ_1, λ_2 such that $0 < \lambda_1 \le 1 \le \lambda_2 < \infty$, we have

$$\sup_{\lambda_1\le\lambda\le\lambda_2}\left| \left(\frac{n\lambda h_n}{2\log_{\theta,K}\left(\frac{1}{h_n}\right)}\right)^{1/2} \sup_{x\in I}\left\{r_n(x;\lambda h_n) - \widehat{E}\, r_n(x;\lambda h_n)\right\} \right. \tag{3.21}$$

$$\left. - \sup_{x\in I}\left\{\frac{\sigma_I^2(x)}{f_X(x)}[K^2]\right\}^{1/2}\right| = o_P(1)\,,$$

as $n \to \infty$.

As $\widehat{\sigma}_X \xrightarrow{P} \sigma_X \in\,]0,\infty[$, it follows from (3.21), that the expressions (3.20) and (3.19) may be rewritten as

$$\left(\frac{n\widehat{\sigma}_X h_n}{2\log_{\theta,K}\left(\frac{|I|}{\widehat{\sigma}_X h_n}\right)}\right)^{1/2} \sup_{x\in I} \pm\, \Theta_{n,3}(x)\left\{\widehat{E}\, r_n(x;\widehat{H}_{n,3}(x)) - r_n(x;\widehat{H}_{n,3}(x))\right\} \xrightarrow{P} 1$$

and

$$\left(\frac{n\widehat{\sigma}_X h_n}{2\log_{\theta,K}\left(\frac{|I|}{\widehat{\sigma}_X h_n}\right)}\right)^{1/2} \sup_{x\in I} \Theta_{n,3}(x)\left|\widehat{E}\, r_n(x;\widehat{H}_{n,3}(x)) - r(x)\right| \xrightarrow{P} 0\,,$$

as $n \to \infty$, where

$$\begin{aligned}\widehat{H}_{n,3}(x) &= \widehat{\sigma}_X \widehat{h}_{n,3}(x)\\ &= \widehat{\sigma}_X n^{-1/5} C_{n,3}(x)\,.\end{aligned}$$

Using the two last relations, we obtain

$$\left(\frac{n^{4/5}\,\widehat{\sigma}_X}{2\log_{\theta,K}\left(\frac{|I|}{n^{-1/5}\widehat{\sigma}_X}\right)}\right)^{1/2} \sup_{x\in I} \pm\, \Theta_{n,3}(x)\left[r_n\left(x;\widehat{H}_{n,3}(x)\right) - r(x)\right] \xrightarrow{P} 1\,,$$

as $n \to \infty$.

REFERENCES

Akaike, H. (1954). An approximation of the density function. *Ann. Inst. Statist. Math.*, **6**, 127-132.

Bosq, D. et Lecoutre, J. P. (1987). *Théorie de l'Estimation Fonctionnelle.* Economica, Paris.

Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Infer. Soc. Processes*, **7**, 225-277.

Devroye, L. (1978b). The uniform convergence of the Nadaraya-Watson regression function Estimate. *Can. J. Statist.*, **6**, 179-191.

Devroye, L. and Györfi, L. (1985). Nonparametric Density Estimation : The view. Wiley, New York.

Devroye, L. and Lugosi, G. (2001). Combinatorial methods in density estimation. Springer Series in Statistics. Springer-Verlag, New York.

Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoretical Prob.*, **13**, 1-37.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

Härdle, W., Jansen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, 1428-1449.

Izenman, A. J. (1991). Recents developments in nonparametric density estimation. *J. Amer. Statist. Assoc.*, **86**, no. 413, 205-224.

Nadaraya, E. A. (1964). On estimating regression. *Theor. Prob. Appl.*, **9**, 141-142.

Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves.* Kluwer, Dordrecht.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065-1076.

Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832-837.

Roussas, G. (1990). *Nonparametric Functional Estimation and Related Topics.* NATO ASI series 355. Kluwer, Dordrecht.

Scott, D. W. (1992). *Multivariate Density Estimation-Theory, Practice and Visualization.* Wiley, New York.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhya A*, **26,** 359-372.

Chapitre IV

CARACTÉRISATION DES STATISTIQUES D'ORDRE

Soit $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ la statistique d'ordre associée à $X_1, X_2, \ldots, X_n$, une suite de variables aléatoires indépendantes, identiquements distribuées, de fonction de répartition $F(x) = P(X_1 \leq x)$ continue et $X_{n,n} = \max(X_1, \ldots, X_n)$ avec $F_n(x) = P(X_{n,n} \leq x) = F^n(x)$.

4.1 *Etude asymptotique de $X_{n,n}$, $X_{k,n}$ et $X_{n-k+1,n} - X_{n-k,n}$*

4.1.1 Etude asymptotique de $X_{n,n}$

Pour une étude approfondie du comportement limite $F_n(x)$, on se referera à Galombos (1978).

Fréchet (1927) fût le premier à mettre en évidence les lois limites possibles de $F_n(a_n x)$ pour un choix convenables des constantes $a_n > 0$.

Cette classe de lois limites est formée de lois de type suivant

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{pour } x \leq 0, \\ \exp(-x^{-\alpha}) & \text{pour } x > 0, \ \alpha > 0. \end{cases}$$

De même, les lois limites possibles de $F_n(a_n x + x_0)$ où $x_0 = \sup\{x : F(x) < 1\} < \infty$ sont carctérisées par

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{pour } x \leq 0, \alpha > 0, \\ 1 & \text{pour } x > 0. \end{cases}$$

Fisher et Tippett (1928) ont établi que les lois limites de $F_n(a_n x + b_n)$ où $a_n > 0$ et b_n sont des constantes réelles convenablement choisies, se réduisent aux types $\Phi_\alpha(x)$, $\Psi_\alpha(x)$ et $\Lambda(x) = \exp^{-\exp(-x)}$ pour tout $x \in \mathbb{R}$.

Procédant à une étude systématique des lois limites pour la distribution du maximum, vers 1936, Von Mises mit en évidence plusieurs conditions suffisantes pour la convergence de $F_n(a_n x + b_n)$ vers chacun des types citées ci-dessus.

On a le théorème suivant :

Théorème 4.1. *La classe des lois limites de* $F_n(a_n x + b_n)$, *où* $a_n > 0$ *et* b_n *sont des constantes réelles convenablement choisies, ne contiennent que les lois du type* $\Phi_\alpha(x)$, $\Psi_\alpha(x)$ *et* $\Lambda(x)$.

Gnédenko (1943) résolut en grande partie le problème de la caractérisation des domaines d'attraction des lois limites. Il énonça les résultats suivants.

Théorème 4.2. *Soient* $x_0 = \sup\{x : F(x) < 1\} = \infty$, $\alpha > 0$. $X_{n,n}$ *appartient au domaine d'attraction de* $\Phi_\alpha(x)$ *si et seulement si pour tout* $x > 0$,

$$\lim_{t\to+\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}.$$

Théorème 4.3. *Soient* $x_0 = \sup\{x : F(x) < 1\} < \infty$, $\alpha > 0$. $X_{n,n}$ *appartient au domaine d'attraction de* $\Psi_\alpha(x)$ *si et seulement si pour tout* $x > 0$,

$$\lim_{t\to-0} \frac{1 - F(tx + x_0)}{1 - F(t + x_0)} = x^{\alpha}.$$

Théorème 4.4. $X_{n,n}$ *appartient au domaine d'attraction de* $\Lambda(x)$ *si et seulement si pour tout* x,

$$\lim_{n\to\infty} n\left(1 - F_n(a_n x + b_n)\right) = \exp(-x),$$

où

$$a_n = F^{-1}\left(1 - \frac{1}{ne}\right) - F^{-1}\left(1 - \frac{1}{n}\right)$$

et

$$b_n = F^{-1}\left(1 - \frac{1}{n}\right),$$

avec

$$F^{-1}(u) = \inf\{x : F(x) > u\}.$$

De Hann (1970) utilisa la théorie des fonctions à variation régulière développée par Karamata (1930), pour étudier les valeurs extrêmes.

Définition 4.1. *La fonction* $H : \mathbb{R}^+ \to \mathbb{R}^+$ *est à variation régulière d'ordre* ρ, *si et seulement si pour tout réel* x *positif*

$$\lim_{x\to+\infty} \frac{H(tx)}{H(x)} = t^{\rho}.$$

La fonction H est dite à variation lente si $\rho = 0$, si ρ est infini on dit que la fonction à variation rapide.

Théorème 4.5. (De Hann (1970)). *$X_{n,n}$ appartient au domaine d'attraction de Φ_α si et seulement si $(1-F)$ est à variation régulière d'ordre $-\alpha$.*

Remarque 4.1. *$X_{n,n}$ appartient au domaine d'attraction de Φ_α alors*

$$F^n(a_n x + b_n) \xrightarrow{\mathcal{L}} \Phi_\alpha(x)$$

avec

$$a_n = \inf\left\{x : 1 - F(x) \leq \frac{1}{n}\right\} = F^{-1}\left(1 - \frac{1}{n}\right),$$
$$b_n = 0.$$

Théorème 4.6.(De Hann (1970)). *$X_{n,n}$ appartient au domaine d'attraction de Ψ_α si et seulement si pour tout réel positif x, la fonction $G(x) = 1 - F\left(x_0 - \frac{1}{x}\right)$ est à variation régulière d'ordre $-\alpha$, quand $x \longrightarrow +\infty$, où $x_0 = \sup\{x : F(x) < 1\} < \infty$.*

Remarque 4.2. *$X_{n,n}$ appartient au domaine d'attraction de Ψ_α alors*

$$F^n((x_0 - a_n)x + x_0) \xrightarrow{\mathcal{L}} \Psi_\alpha(x)$$

où

$$a_n = \inf\left\{x : 1 - F(x) \leq \frac{1}{n}\right\} = F^{-1}\left(1 - \frac{1}{n}\right).$$

Théorème 4.7. *$X_{n,n}$ appartient au domaine d'attraction de Λ si et seulement si*

$$\lim_{t \to +\infty} \frac{\chi(tx) - \chi(x)}{\chi(ty) - \chi(y)} = \frac{\log x}{\log y}.$$

x, y sont des réels positifs $(y \neq 1)$ et la fonction χ est définie par

$$\chi(x) = \inf\{z : 1 - F(z) \leq x\} = F^{-1}(1 - x).$$

Remarque 4.3. *$X_{n,n}$ appartient au domaine d'attraction de Λ alors*

$$F^n(a_n x + b_n) \xrightarrow{\mathcal{L}} \Lambda(x),$$

avec

$$a_n = F^{-1}_{X_{n,n}}(\Lambda(1)) - F^{-1}_{X_{n,n}}(\Lambda(0)), \; b_n = F^{-1}_{X_{n,n}}(\Lambda(0))$$

et

$$F_{X_{n,n}}^{-1}(x) = \inf\{z : F^n(z) \geq x\}.$$

Exemple 4.1. *Supposons que F est la distribution d'une loi uniforme sur* $[0,\ 1]$, *trouvons la loi limite de $X_{n,n}$. Comme $x_0 = \sup\{x : F(x) < 1\} = 1 < \infty$, alors $X_{n,n}$* n'appartient pas *au domaine d'attraction de $\Phi_\alpha(x)$. Introduisons, pour $x > 0$, $\chi(x) = 1 - F\left(x_0 - \frac{1}{x}\right)$, nous constatons que*

$$\lim_{x\to+\infty} \frac{\chi(tx)}{\chi(x)} = \lim_{x\to+\infty} \frac{1 - F\left(x_0 - \frac{1}{tx}\right)}{1 - F\left(x_0 - \frac{1}{x}\right)} = \lim_{x\to+\infty} \frac{\frac{1}{tx}}{\frac{1}{x}} = t^{-1}.$$

Donc $X_{n,n}$ appartient au domaine d'attraction de $\Psi_1(x)$, c'est à dire

$$F^n((x_0 - a_n)x + x_0) \xrightarrow{\mathcal{L}} \Psi_1(x)$$

où

$$a_n = \inf\left\{x : 1 - F(x) \leq \frac{1}{n}\right\} = F^{-1}\left(1 - \frac{1}{n}\right) = 1 - \frac{1}{n}$$

$$\Psi_1(x) = \begin{cases} \exp(x) & \text{pour } x \leq 0, \\ 1 & \text{pour } x > 0. \end{cases}$$

Le comportement limite de $X_{n,n}$ a beaucoup influencé sur le comportement asymptotique de $X_{k,n}$ et de $X_{n-k+1,n} - X_{n-k,n}$. Cette théorie a fait l'objet de nombreux travaux (voir David (1970), De Hann et Resnick (1984), Deheuvels (1986), Galombos (1978), Gnedenko (1943), Hall (1978), Weissman (1978)). Nous n'allons pas citer tous ces travaux mais on se contente de ce qui nous intéresse.

4.1.2 Etude asymptotique de $X_{k,n}$

D'après les formules binomiales, on a

$$P(X_{k,n} \geq z) = \sum_{t=0}^{k-1} C_n^t (F(z))^t (1 - F(z))^{n-t}$$

et

$$P(X_{n-k+1,n} \leq z) = P\left((1-F)^{-1}(U_{k,n}) \leq z\right) = \sum_{t=0}^{k-1} C_n^t (F(z))^{n-t} (1 - F(z))^t,$$

où $U_{k,n}$ est la k^{ieme} statistique d'ordre d'une suite de variables aléatoires indépendantes et de loi uniforme sur $[0,\ 1]$.

Soient $\alpha(H) = \inf\{x : H(x) > 0\}$ et $\omega(H) = \sup\{x : H(x) < 1\}$, alors

Théorème 4.8. *Pour toutes suites de nombres réels $a_n > 0$, b_n, pour tout entier naturel non nul k, fixé et quand $n \to \infty$,*

$$F_{n-k+1,n}(a_n x + b_n) = P\left(\frac{X_{n-k+1,n} - b_n}{a_n} \leq x\right)$$

converge vers une distribution non dégénérée $H_k(x)$ si et seulement si

$$F^n(a_n x + b_n) = P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right)$$

converge vers une distribution non dégénérée.

Si $H(x)$ existe, alors pour tout $\alpha(H) < x < \omega(H)$,

$$H_k(x) = H(x) \sum_{t=0}^{k-1} \frac{1}{t!} \left[\log \frac{1}{H(x)}\right]^t,$$

où $H(x)$ est l'une des trois fonctions de répartition $\Phi_\alpha(x)$, $\Psi_\alpha(x)$ et $\Lambda(x)$.

Exemple 4.2. *Calculons $H_k(x)$ pour la loi uniforme. D'après l'exemple précédent on a $\alpha(H) = \inf\{x : H(x) > 0\} = -\infty$, $\omega(H) = \sup\{x : H(x) < 1\} = +\infty$ et $\forall x \in \mathbb{R}$,*

$$H_k(x) = H(x) \sum_{t=0}^{k-1} \frac{1}{t!} \left[\log \frac{1}{H(x)}\right]^t,$$

où

$$H(x) = \Psi_1(x) = \begin{cases} \exp(x) & \text{pour } x \leq 0, \\ 1 & \text{pour } x > 0. \end{cases}$$

Par conséquent

$$H_k(x) = \begin{cases} 0 & \text{si } x > 0 \\ e^x \sum_{t=0}^{k-1} \frac{1}{t!}(-x)^t & \text{si } x \leq 0. \end{cases}$$

L'étude asymptotique de $X_{k,n}$ établie ci-dessus est quand k est fixé et ne dépend pas de n, Chibisov (1964) a prouvé que si $k(n) \to \infty$ et $n - k(n) \to \infty$, les lois limites de $X_{k(n),n}$ sont les trois types limites suivantes :

$$G_1^\beta(x) = \begin{cases} 0 & \text{pour } x \leq 0, \\ \Phi(\beta \log(x)) & \text{pour } x > 0,\ \beta > 0, \end{cases}$$

$$\begin{aligned} G_2^\beta(x) &= \begin{cases} 1 & \text{pour } x \geq 0, \\ \Phi(-\beta \log(-x)) & \text{pour } x < 0,\ \beta > 0, \end{cases} \\ G_3^\beta(x) &= \Phi(x), \quad x \in \mathbb{R}, \end{aligned}$$

où $\Phi(x)$ est la distribution d'une loi normale centrée et réduite. Les meilleurs résultats sur l'étude asymptotique de $X_{k(n),n}$ sont donnés par Balkema et De Hann (1978). Ils ont prouvé que si la condition suivante $k(n+1) - k(n) = o\left(\min\left\{\sqrt{k(n)}, \sqrt{n-k(n)}\right\}\right)$ est verifiée alors les lois limites de $X_{k(n),n}$ sont les trois types limites de Chibisov (1964).

4.1.3 Etude asymptotique des espacements $X_{n-k+1,n} - X_{n-k,n}$

Théorème 4.9. (Hall (1978), **Weissman** (1978)). *Supposons que,* $\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = e^{-e^{-x}},$$

alors pour tout $k \geq 1$, *fixé*

$$\left\{\frac{X_{n,n} - X_{n-1,n}}{a_n}, \ldots, \frac{X_{n-k+1,n} - X_{n-k,n}}{a_n}\right\} \xrightarrow{\mathcal{L}} \left\{\omega_1, \frac{\omega_2}{2}, \ldots, \frac{\omega_k}{k}\right\},$$

où $(\omega_i)_{1\leq i\leq k}$ *est une suite de variables aléatoires indépendantes, de loi exponentielle de paramètre égal à* 1.

La démonstration de ce théorème est basée sur les théorèmes de Galombos et Kortz (1978), Pyke (1965) et les lemmes suivants.

Théorème 4.10. (Galombos, Kortz (1978)). *Si* $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ *est la statistique d'ordre associée à* $X_1, X_2, \ldots, X_n$, *une suite de variables aléatoires indépendantes et de loi exponentielle de paramètre égal à* 1, *alors* $X_{n,n} - X_{n-1,n} = \omega_1$, $k(X_{n-k+1,n} - X_{n-k,n}) = \omega_k, \ldots, n(X_{1,n} - X_{2,n}) = \omega_n$, où $(\omega_i)_{1\leq i\leq n}$ *est une suite de variables aléatoires indépendantes et de loi exponentielle de paramètre égal à* 1.

Lemme 4.1. *Soit* U *une variable aléatoire uniforme sur* $[0, 1]$, *alors* -$\log U$ suit une loi *exponentielle de paramètre égal à* 1.

Preuve $P(-\log U > t) = P(U < e^{-t}) = e^{-t}$.

Lemme 4.2. *Soit* ω *une variable aléatoire de loi exponentielle de paramètre égal à* 1, *alors* $\exp(-\omega)$ *est uniformement distribué.*

$$P(\exp(-\omega) < x) = P(-\omega < \log x) = P(\omega > -\log x) = x.$$

A partir de ces lemmes et de ce théorème, nous déduisons que si $U_{0,n} = 0 < U_{1,n} < \ldots < U_{n,n} < U_{n+1,n} = 1$ est la statistique d'ordre associée à $U_1, U_2, \ldots, U_n$, une suite de variables indépendantes de loi uniforme sur $]0,1[$, alors les variables $\left((1-U_{i,n}/1-U_{i-1,n})^{n-i+1}\right)_{1\leq i\leq n}$ sont indépendantes et uniformément distribuées. Nous avons la même constatation pour les variables $\left((U_{i,n}/U_{i+1,n})^{i}\right)_{1\leq i\leq n}$.

Théorème 4.11. *Soient $\omega_1,\omega_2,\ldots$ une suite de variables aléatoires indépendantes et de loi exponentielle de paramètre égal* à 1. *Posons* $S_m=\omega_1+\ldots+\omega_m$, $m\geq 1$, $S_0=0$, alors $(U_{i,n}, 0\leq i\leq n+1)$ et $(S_i/S_{n+1}, 0\leq i\leq n+1)$ *sont identiquement distribuées.*

Preuve (Voir Pyke (1965, p.403))

Posons

$$\chi(x)=\inf\{z:1-F(z)\leq x\}=F^{-1}(1-x),$$

d'après ce qui précède on a $X_{n-k,n}=\chi(U_{n-k+1,n})$, $1\leq j\leq n$

$$\frac{X_{n-k+1,n}-X_{n-k,n}}{a_n}=\frac{\chi(U_{k,n})-\chi(U_{k+1,n})}{\chi\left(\frac{1}{ne}\right)-\chi\left(\frac{1}{n}\right)}$$

$$=\frac{\chi(S_k(n/S_{n+1})/n)-\chi(S_{k+1}(n/S_{n+1})/n)}{\chi\left(\frac{1}{ne}\right)-\chi\left(\frac{1}{n}\right)}$$

$$=\frac{\chi(S_k(n/S_{n+1})/n)-\chi\left(\frac{1}{n}\right)}{\chi\left(\frac{1}{ne}\right)-\chi\left(\frac{1}{n}\right)}-\frac{\chi(S_{k+1}(n/S_{n+1})/n)-\chi\left(\frac{1}{n}\right)}{\chi\left(\frac{1}{ne}\right)-\chi\left(\frac{1}{n}\right)}.$$

D'après le Théorème 4.7 et comme n/S_{n+1} converge en probabilité vers 1, alors l'expression précédente tend en loi vers -log $S_{k+1}+\log S_k=\frac{1}{k}\log(U_{k,n}/U_{k+1,n})^k$, en appliquant le Lemme 4.1 on aboutit au résultat voulu.

Nous terminons ce paragraphe par citer les deux lemmes de Deheuvels (1982) sur la caractérisation des espacements extrêmes quand $X_{n,n}$ appartient au domaine d'attraction de $\Phi_\alpha(x)$, $\Psi_\alpha(x)$.

Lemme 4.3. *Si pour tout* $\alpha>0$ et $x>0$

$$\lim_{n\to\infty}P\left(a_n^{-1}X_{n,n}\leq x\right)=e^{-x^{-\alpha}},$$

alors pour tout $k\geq 1$, fixé

$$\left\{a_n^{-1}(X_{n-k+1,n}-X_{n-k,n}),\ 1\leq k\leq n\right\}\xrightarrow{\mathcal{L}}\left\{\left(\sum_{i=1}^{k}\omega_i\right)^{-\frac{1}{\alpha}}-\left(\sum_{i=1}^{k+1}\omega_i\right)^{-\frac{1}{\alpha}},\ 1\leq k\leq n\right\},$$

où $(\omega_i)_{1\leq i\leq n}$ *est une suite de variables aléatoires indépendantes de loi exponentielle.*

Lemme 4.4. *Si pour tout* $\alpha > 1$ *et* $x < 0$

$$\lim_{n\to\infty} P\left((x_0 - a_n)^{-1}(X_{n,n} - x_0) \leq x\right) = e^{-(-x)^{\alpha}},$$

alors pour tout $k \geq 1$, fixé

$$\left\{(x_0 - a_n)^{-1}(X_{n-k+1,n} - X_{n-k,n}),\quad 1 \leq k \leq n\right\} \xrightarrow{\mathcal{L}} \left\{ \begin{array}{c} \left(\sum_{i=1}^{k+1}\omega_i\right)^{\frac{1}{\alpha}} - \left(\sum_{i=1}^{k}\omega_i\right)^{\frac{1}{\alpha}}, \\ 1 \leq k \leq n \end{array} \right\},$$

où $(\omega_i)_{1\leq i\leq n}$ *est une suite de variable aléatoires indépendantes de loi exponentielle.*

Notre travail consiste à caractériser les distributions telles que les espacements $X_{n-k+1,n} - X_{n-k,n}$ soient asymptotiquement indépendants, il est inspiré au départ du théorème de Hall (1978) et Weissman (1978) et basé sur les lemmes de Deheuvels (1986) cités antérieurement.

4.2 Résultats

Théorème 4.12. Soit $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ *la statistique d'ordre associée à* $X_1, X_2, \ldots, X_n$, *une suite de variables aléatoires indépendantes, identiquement distribuées, de densité* f *et de fonction de répartition* $F(x)$ *continue.*

Si les conditions suivantes sont vérifiées :

(i) Le type limite de $X_{n,n}$ *existe ;*

(ii) $\{X_{n-k+1,n} - X_{n-k,n}\}$ *et* $\{X_{n-k,n} - X_{n-k-1,n}\}$ *sont asymptotiquement indépendants, où* k *fixé ;*

alors, $X_{n,n}$ *appartient au domaine d'attraction de* $\Lambda(x)$.

Nous venons d'énoncer que l'existence du type limite associé à l'indépendance asymptotique des espacements caractérisent les distributions dont les extrêmes sont attirés par la loi de Gumble, nous constatons que ce résultat est toujours vrai en remplaçant la condition

(ii) par la compacité stochastique de $X_{n,n}$.

Nous rappelons que le concept de la compacité stochastique fût étudié initialement par Doeblin (1940), De Hann et Ridder (1979) étendirent cette idée aux extrêmes simples et De Hann et Resnick (1984) donnèrent les conditions nécessaires et suffisantes entre cette compacité et l'existence de toutes les limites non dégénérées de $X_{n,n}$.

Définition 4.1. La *suite $X_{n,n}$ est stochastiquement compacte s'il existe*

$\{a_n > 0,\ b_n \in \mathbb{R},\ n \geq 1\}$ *tels que toute suite*

$$\left\{ \frac{X_{n(k),n(k)} - b_{n(k)}}{a_{n(k)}}, \quad k \geq 1 \right\}$$

contient une sous-suite qui converge en loi vers un distribution non dégénérée.

Théorème 4.13. *Soit* $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ *la statistique d'ordre associée à* $X_1, X_2, \ldots, X_n$, *une suite de variables aléatoires indépendantes, identiquement distribuées, de densité f et de fonction de répartition $F(x)$ continue.*

Si les conditions suivantes sont vérifiées :

(i) $X_{n,n}$ est stochastiquement compacte;

(ii) $\{X_{n-k+1,n} - X_{n-k,n}\}$ et $\{X_{n-k,n} - X_{n-k-1,n}\}$ sont asymptotiquement indépendants, où k fixé;

alors $X_{n,n}$ appartient au domaine d'attraction de $\Lambda(x)$.

Etudions maitenant le cas où k dépend de n.

Théorème 4.14 *Soit* $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ *la statistique d'ordre associée à* $X_1, X_2, \ldots, X_n$, *une suite de variables aléatoires indépendantes, identiquement distribuées.* Si $\lim\limits_{n\to\infty} k_n = +\infty$ et $\lim\limits_{n\to\infty} k_n/n = \alpha$ *avec* $0 < \alpha < 1$, *alors l'indépendance asymptotique de* $\{X_{n-k_n+1,n} - X_{n-k_n,n}\}$ *et* $\{X_{n-k_n,n} - X_{n-k_n-1,n}\}$ *n'implique pas nécessairement l'appartenance de $X_{n,n}$ au domaine d'attraction de $\Lambda(x)$.*

Définissons $G(u)$ et $g(x)$ par

$$G(u) = (1-F)^{-1}(u) = \inf\{x \mid 1 - F(x) \leq u\},$$

$$g(x) = \exp\left(C(x) + \int_B^x \frac{\varepsilon(t)}{t} dt \right),$$

où

B une constante positive, $C(x)$, $\varepsilon(x)$ sont des fonctions réelles telles que $\varepsilon(t)$ *est bornée sur* $\mathbb{R}$ et

$$\lim_{x\to\infty} C(x) = C < \infty.$$

Théorème 4.15. *Soient $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes, identiquement distribuées, de densité f et de fonction de répartition $F(x)$ continue, A* une constante positive.

Si les conditions suivantes sont vérifiées :

1°) $\lim\limits_{n\to\infty} k_n = +\infty$, $\lim\limits_{n\to\infty} k_n/n = 0$

2°) $G(u) = a + g\left(\frac{1}{u}\right) + \int\limits_A^{\frac{1}{u}} \frac{g(v)}{v} dv$,

alors

$$\{X_{n-k_n+1,n} - X_{n-k_n,n}\} \text{ et } \{X_{n-k_n,n} - X_{n-k_n-1,n}\}$$

sont asymptotiquement indépendants.

Lemme 4.1 (Seneta (1975)). *Soit L une fonction définie de $\mathbb{R}^+$ dans $\mathbb{R}^+$.*

Les conditions suivantes sont vérifiées :

(i) $\lim\limits_{u\to\infty} L(ue(u))/L(u) = 1$, *pour toute fonction positive $e(u)$, telle que* $\lim\limits_{u\to\infty} e(u) = 1$.

(ii) $L(x) = \exp\left(\eta(x) + \int\limits_A^x \frac{\varepsilon(t)}{t} dt\right)$, $A > 0$

où $\varepsilon(t)$ et $\eta(x)$ sont des fonctions vérifiant :

$\lim\limits_{x\to\infty} \eta(x) = \eta$, $(\eta < \infty)$ et $\varepsilon(t)$ *est bornée.*

Corollaire 4.1. *Soit* $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ *la statistique d'ordre associée à* $X_1, X_2, \ldots, X_n$, *une suite de variables aléatoires indépendantes, identiquement distribuées, de densité f et de fonction de répartition $F(x)$ continue. Si* (k_n) est une suite croissante de nombres réels positifs telle que :

$$\lim_{n\to\infty} k_n = +\infty, \quad \lim_{n\to\infty} k_n/n = 0,$$

alors l'indépendance asymptotique de $\{X_{n-k_n+1,n} - X_{n-k_n,n}\}$ et $\{X_{n-k_n,n} - X_{n-k_n-1,n}\}$ *n'implique pas nécessairement l'appartenance de* $X_{n,n}$ *au domaine d'attraction de* $\Lambda(x)$.

4.3 Démonstrations

Preuve du Théorème 4.13

La démonstration se fait par l'absurde. Comme la loi limite de $X_{n,n}$ existe, elle ne peut être que les lois du type $\Phi_\alpha(x)$, $\Psi_\alpha(x)$ et $\Lambda(x)$. Supposons que $X_{n,n}$ appartient au domaine d'attaction de $\Psi_\alpha(x)$ alors d'après le Lemme 4.4, nous avons pour tout $k \geq 1$, fixé

$$(x_0 - a_n)^{-1}(X_{n-k+1,n} - X_{n-k,n}) \xrightarrow{\mathcal{L}} \left(\sum_{i=1}^{k+1}\omega_i\right)^{\frac{1}{\alpha}} - \left(\sum_{i=1}^{k}\omega_i\right)^{\frac{1}{\alpha}},$$

où $(\omega_i)_{1\leq i\leq n}$ est une suite de variables aléatoires indépendantes de loi exponentielle.

Posons :

$$\begin{cases} \zeta_k = \left(\sum_{i=1}^{k+1}\omega_i\right)^{\frac{1}{\alpha}} - \left(\sum_{i=1}^{k}\omega_i\right)^{\frac{1}{\alpha}}, \\ \zeta_{k+1} = \left(\sum_{i=1}^{k+2}\omega_i\right)^{\frac{1}{\alpha}} - \left(\sum_{i=1}^{k+1}\omega_i\right)^{\frac{1}{\alpha}}. \end{cases}$$

La condition (ii) est vérifiée si et seulement si ζ_k et ζ_{k+1} sont indépendantes, autrement dit $P\left(\zeta_{k+1} < y \mid \zeta_k = x\right)$ ne dépend pas de x.

Rappelons que

$$P\left(\zeta_{k+1} < y \mid \zeta_k = x\right) = \int_0^y \frac{f_{\zeta_k,\zeta_{k+1}}(x,z)}{f_{\zeta_k}(x)}dz,$$

où

$f_{\zeta_k,\zeta_{k+1}}(x,z)$ est la densité de$\left(\zeta_k,\zeta_{k+1}\right)$

et

$f_{\zeta_k}(x)$ est la densité de ζ_k.

Calculons $f_{\zeta_{k+1},\zeta_k}(x,z)$.

$$P(\zeta_k < x, \zeta_{k+1} < z) =$$
$$= \lambda \int_0^{+\infty} P\left(G_k(u_1;\omega_2,\ldots,\omega_{k+1}) < x,\ G_{k+1}(u_1;\omega_2,\ldots,\omega_{k+2}) < z \mid \omega_1 = u_1\right) e^{-\lambda u_1} du_1,$$

où

$$G_l(u_1,\ldots,u_j;\omega_{j+1},\ldots,\omega_{l+1}) = \left(\sum_{i=1}^{j} u_i + \sum_{i=j+1}^{l+1} \omega_i\right)^{\frac{1}{\alpha}} - \left(\sum_{i=1}^{j} u_i + \sum_{i=j+1}^{l} \omega_i\right)^{\frac{1}{\alpha}}.$$

Comme les $(\omega_i)_{1\le i\le k+2}$ sont indépendants, itérativement nous obtenons,

$$P(\zeta_k < x, \zeta_{k+1} < z) =$$
$$= 1 - \lambda^k \int_0^{+\infty} \ldots \int_0^{+\infty} e^{-\lambda L_{k+1}(0,0,u_1,\ldots,u_{k+1})} du_k \ldots du_1$$
$$-\lambda^{k+1} \int_0^{+\infty} \ldots \int_0^{U_k(x,0,u_1,\ldots,u_k)} e^{-\lambda L_{k+1}(0,z,u_1,\ldots,u_{k+1})} du_{k+1} \ldots du_1,$$

où

$$L_l(x,z,u_1,\ldots,u_l) = \left(x + z + \left(\sum_{i=1}^{l} u_i\right)^{\frac{1}{a}}\right)^{a}.$$

$$U_l(x,0,u_1,\ldots,u_l) = L_l(x,0,u_1,\ldots,u_l) - L_l(0,0,u_1,\ldots,u_l).$$

La dérivée par rapport à $(x,\ z)$ de cette expression nous donne

$$f_{\zeta_{k+1},\zeta_k}(x,z) =$$
$$= \lambda^{k+2} a \int_0^{+\infty} \ldots \int_0^{+\infty} T_{k+1}(x,z,u_1,\ldots,u_{k+1})\, e^{-\lambda L_{k+1}(x,z,u_1,\ldots,u_{k+1})} du_{k+1} \ldots du_1,$$

où

$$T_{k+1}(x,z,u_1,\ldots,u_{k+1}) = L_k^{\frac{a-1}{a}}(x,0,u_1,\ldots,u_k)\, L_{k+1}^{\frac{a-1}{a}}(x,z,u_1,\ldots,u_{k+1}).$$

Calculons maintenant $P(\zeta_k < x)$,

$$P(\zeta_k < x) = P\left(\left(\sum_{i=1}^{k+1} \omega_i\right)^{\frac{1}{\alpha}} - \left(\sum_{i=1}^{k} \omega_i\right)^{\frac{1}{\alpha}} < x\right) =$$
$$= P\left(\left(u_1 + \sum_{i=2}^{k+1} \omega_i\right)^{\frac{1}{\alpha}} - \left(u_1 + \sum_{i=2}^{k} \omega_i\right)^{\frac{1}{\alpha}} < x \mid \omega_1 = u_1\right).$$

En procédant par itération, nous obtenons,

$$P(\zeta_k < x) = \lambda^k \int_0^{+\infty} \dots \int_0^{+\infty} e^{-\lambda L_{k+1}(0,0,u_1,\dots,u_{k+1})} - e^{-\lambda L_{k+1}(x,0,u_1,\dots,u_{k+1})} du_{k+1} du_k \dots du_1,$$

donc $f_{\zeta_k}(x)$ s'écrit

$$f_{\zeta_k}(x) = \lambda^{k+1} a \int_0^{+\infty} \dots \int_0^{+\infty} L_{k+1}^{\frac{a-1}{a}}(x, 0, u_1, \dots, u_{k+1}) \, e^{-\lambda L_{k+1}(x,0,u_1,\dots,u_{k+1})} du_{k+1} du_k \dots du_1.$$

Par conséquent :

$$P(\zeta_{k+1} < y \mid \zeta_k = x) = \int_0^{+\infty} \dots \int_0^{+\infty} L_{k+1}^{\frac{a-1}{a}}(x, 0, u_1, \dots u_{k+1}) \, e^{-\lambda L_{k+1}(x,y,u_1,\dots,u_{k+1})} du_{k+1} \dots du_1.$$

$$\left(\int_0^{+\infty} \dots \int_0^{+\infty} L_{k+1}^{\frac{a-1}{a}}(x, 0, u_1, \dots, u_{k+1}) \, e^{-\lambda L_{k+1}(x,0,u_1,\dots,u_{k+1})} du_{k+1} \dots du_1 \right)^{-1}.$$

$P(\zeta_{k+1} < y \mid \zeta_k = x)$ ne dépend pas de x, si sa dérivée par rapport à x est nulle autrement dit

$$-(a+1) \int_0^{+\infty} \dots \int_0^{+\infty} L_{k+1}^{\frac{a-2}{a}}(x, 0, u_1, \dots, u_{k+1}) \, L_{k+1}^{\frac{a-2}{a}}(x, 0, v_1, \dots, v_{k+1}) \, F_1\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) du_{k+1} dv_{k+1} \dots du_1 dv_1$$

$$+\lambda a \int_0^{+\infty} \dots \int_0^{+\infty} L_{k+1}^{\frac{a-1}{a}}(x, 0, u_1, \dots, u_{k+1}) \, L_{k+1}^{\frac{a-1}{a}}(x, 0, v_1, \dots, v_{k+1}) \, H_1\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) du_{k+1} dv_{k+1} \dots du_1 dv_1 = 0,$$

où

$$F_1\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) = \left(L_{k+1}^{\frac{1}{a}}(0, 0, u_1, \dots, u_{k+1}) - L_{k+1}^{\frac{1}{a}}(0, 0, v_1, \dots, v_{k+1}) \right)$$
$$\times e^{-\lambda L_{k+1}(x,y,u_1,\dots,u_{k+1})} e^{-\lambda L_{k+1}(x,0,u_1,\dots,u_{k+1})}$$

$$H_1\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) = \left(L_{k+1}^{\frac{a-1}{a}}(x, y, u_1, \dots, u_{k+1}) - L_{k+1}^{\frac{a-1}{a}}(x, 0, v_1, \dots, v_{k+1}) \right)$$
$$\times e^{-\lambda L_{k+1}(x,y,u_1,\dots,u_{k+1})} e^{-\lambda L_{k+1}(x,0,u_1,\dots,u_{k+1})}.$$

Ce qui donne finalement :

$$\left(\sum_{i=1}^{k+1} u_i\right)^{\frac{1}{a}} = \left(\sum_{i=1}^{k+1} v_i\right)^{\frac{1}{a}}$$

et

$$\left(\sum_{i=1}^{k+1} u_i\right)^{\frac{1}{a}} = \left(\sum_{i=1}^{k+1} v_i\right)^{\frac{1}{a}} + y, \quad \forall y \in \mathbb{R}.$$

Comme ces égalités sont impossibles, ζ_{k+1} et ζ_k sont dépendants.

Supposons que $X_{n,n}$ appartient au domaine d'attraction de $\Phi_\alpha(x)$, alors d'après le lemme 4.3, on a pour tout $k \geq 1$, fixé

$$a_n^{-1}\left(X_{n-k+1,n} - X_{n-k,n}\right) \xrightarrow{\mathcal{L}} \left(\sum_{i=1}^{k} \omega_i\right)^{-\frac{1}{\alpha}} - \left(\sum_{i=1}^{k+1} \omega_i\right)^{-\frac{1}{\alpha}},$$

où $(\omega_i)_{1 \leq i \leq n}$ est une suite de variables aléatoires indépendantes de loi exponentielle.

Posons :

$$\begin{cases} \zeta_k' = \left(\sum_{i=1}^{k} \omega_i\right)^{-\frac{1}{\alpha}} - \left(\sum_{i=1}^{k+1} \omega_i\right)^{-\frac{1}{\alpha}}, \\ \zeta_{k+1}' = \left(\sum_{i=1}^{k+1} \omega_i\right)^{-\frac{1}{\alpha}} - \left(\sum_{i=1}^{k+2} \omega_i\right)^{-\frac{1}{\alpha}}. \end{cases}$$

La condition (ii) est vérifiée si $P\left(\zeta_{k+1}' < y \mid \zeta_k' = x\right)$ ne dépend de x,

Rappelons que

$$P\left(\zeta_{k+1}' < y \mid \zeta_k' = x\right) = \int_0^y \frac{f_{\zeta_{k+1}',\zeta_k'}(x,z)}{f_{\zeta_k'}(x)} dz,$$

où

$f_{\zeta_{k+1}',\zeta_k'}(x,z)$ est la densité de $\left(\zeta_k', \zeta_{k+1}'\right)$

et

$f_{\zeta_k'}(x)$ est la densité de ζ_k'.

D'une manière analogue que précédemment, en posant

$$S_k(x,z,u_1,\ldots,u_k) = \left(x + z + \left(\sum_{i=1}^{k} u_i\right)^{\frac{-1}{a}}\right)^{-a-1},$$

$$R_{k+1}(x,z,u_1,\ldots,u_{k+1}) = S_{k+1}(x,0,u_1,\ldots,u_{k+1})\, S_{k+1}(x,z,u_1,\ldots,u_{k+1}),$$

$$T_{1k} = \sum_{i=1}^{k} u_i,\ T_{2k} = \sum_{i=1}^{k} v_i,$$

nous obtenons,

$$P\left(\zeta_{k+1}' < y \mid \zeta_k' = x\right) = \tfrac{N}{D},$$

avec

$$N = \lambda^{k+2} a \times$$
$$\int_0^y \int_0^{(z+x)^{-a}} \cdots \int_0^{(z+x)^{-a}-T_{1k}} R_{k+1}(x,z,u_1,\ldots,u_{k+1})\, e^{-\lambda S_{k+1}^{\frac{-a}{-a-1}}(x,z,u_1,\ldots,u_{k+1})} du_{k+1}\ldots dz,$$

$$D = \lambda^{k+1} a \int_0^{(x)^{-a}} \dots \int_0^{(x)^{-a} - T_{2k}} S_{k+1}(x, 0, v_1, \dots, v_{k-1}) \, e^{-\lambda S_{k+1}^{\frac{-a}{-a-1}}(x,0,v_1,\dots,v_{k-1})} dv_{k+1} \dots dv_1.$$

La dérivée de $P\left(\zeta'_{k+1} < y \mid \zeta'_k = x\right)$ par rapport à x est nulle si

$$(a+1) \int_0^{(x)^{-a}(y+x)^{-a}} \int_0^{} \dots \int^{(x)^{-a} - T_{2k-1}(y+x)^{-a} - T_{1k-1}} \int_0^{} M_{k,k+1}(x,0) F_2\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) du_{k+1} dv_{k+1} \dots du_1 dv_1$$

$$+\lambda a \int_0^{(x)^{-a}(y+x)^{-a}} \int_0^{} \dots \int_0^{(x)^{-a} - T_{1k-1}(x)^{-a} - T_{k-1}} \int_0^{} M_{k+1,k+1}^{\frac{-a-1}{-a-2}}(x,0) H_2\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) du_{k+1} dv_{k+1} \dots du_1 dv_1$$

$$= 0$$

où

$$M_{k,l}(x,0) = S_k^{\frac{-a-2}{-a-1}}(x, 0, u_1, \dots, u_k) \, S_l^{\frac{-a-2}{-a-1}}(x, 0, v_1, \dots, v_l),$$

$$F_2\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) = \left(T_{1k+1}^{\frac{-1}{a}} - T_{2k+1}^{\frac{-1}{a}}\right) e^{-\lambda S_{k+1}^{\frac{-a}{-a-1}}(x,y,u_1,\dots,u_{k+1})} e^{-\lambda S_{k+1}^{\frac{-a}{-a-1}}(x,0,u_1,\dots,u_{k+1})},$$

$$H_2\left(x, y, \sum_{i=1}^{k+1} u_i, \sum_{i=1}^{k+1} v_i\right) = \left(S_{k+1}(x, y, u_1, \dots, u_{k+1}) - S_{k+1}(x, 0, v_1, \dots, v_{k+1})\right)$$
$$\times e^{-\lambda S_{k+1}^{\frac{-a}{-a-1}}(x,y,u_1,\dots,u_{k+1})} e^{-\lambda S_{k+1}^{\frac{-a}{-a-1}}(x,0,v_1,\dots,v_{k+1})}.$$

Ce qui donne finalement :

$$\left(\sum_{i=1}^{k+1} u_i\right)^{-\frac{1}{a}} = \left(\sum_{i=1}^{k+1} v_i\right)^{-\frac{1}{a}}$$

et

$$\left(\sum_{i=1}^{k+1} u_i\right)^{-\frac{1}{a}} = \left(\sum_{i=1}^{k+1} v_i\right)^{-\frac{1}{a}} + y, \ \forall y \in \mathbb{R}.$$

Nous avons prouvé que $X_{n,n}$ n'appartient ni au domaine d'attraction de $\Phi_\alpha(x)$ ni au domaine de $\Psi_\alpha(x)$, elle est nécessairement dans le domaine de $\Lambda(x)$.

Preuve du Théorème 1.13

$X_{n,n}$ étant stochastiquement compacte, il existe $\{a_n > 0, \ b_n \in \mathbb{R}, \ n \geq 1\}$ tels que toute suite $\{X_{n(l),n(l)} - b_{n(l)}/a_{n(l)}, \ l \geq 1\}$ contient une sous-suite qui converge en loi vers une distribution non dégénérée.

Comme $\{X_{n-k+1,n} - X_{n-k,n}\}$ et $\{X_{n-k,n} - X_{n-k-1,n}\}$ sont asymptotiquement indépendants , alors selon le Théorème (4.12), toutes les sous-suites de $X_{n(l),n(l)}$ qui convergent en loi, appartiennent au domaine d'attraction de la loi Gumble.

Supposons que $X_{n(1),n(1)}$ est l'une de ces sous suites qui converge en loi, alors

$$F^{n(1)}\left(a_{n(1)}x + b_{n(1)}\right) \xrightarrow{L} \Lambda(x).$$

D'après la remarque 4.3, $a_{n(1)}$ et $b_{n(1)}$ peuvent être remplacés respectivement par

$$\left(F^{n(1)}\right)^{-1}(\Lambda(1)) - \left(F^{n(1)}\right)^{-1}(\Lambda(0)) \quad \text{et} \quad \left(F^{n(1)}\right)^{-1}(\Lambda(0)).$$

En procédant de la même façon pour les autres sous-suites qui convergent, nous constatons que les sous-suites

$$\left\{\left(X_{n(l),n(l)} - \left(F^{n(l)}\right)^{-1}(\Lambda(0))\right)\left(\left(F^{n(l)}\right)^{-1}(\Lambda(1)) - \left(F^{n(l)}\right)^{-1}(\Lambda(0))\right)^{-1},\ l \geq 1\right\}$$

pouvant converger, sont attirées par la loi de Gumble. Par conséquent, la suite

$$\left\{\left(X_{n,n} - (F^n)^{-1}(\Lambda(0))\right)\left((F^n)^{-1}(\Lambda(1)) - (F^n)^{-1}(\Lambda(0))\right)^{-1}\right\}$$

converge en loi vers la loi de Gumble.

Preuve du Théorème 4.14

$U_2, \ldots, U_n$ une suite de variables aléatoires indépendantes , de loi uniforme sur $[0, 1]$ et $U_{1,n} < U_{2,n} < \ldots < U_{n,n}$ sa statistique d'ordre associée.

Selon Siddiqui (1960), nous avons $\{U_{n-[n\alpha]+1,n} - U_{n-[n\alpha],n}\}$ et $\{U_{n-[n\alpha],n} - U_{n-[n\alpha]-1,n}\}$ sont asymptotiquement indépendants, mais $U_{n,n}$ appartient au domaine d'attraction de $\Phi_\alpha(x)$.

Preuve du Corollaire 4.1

Soient $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes, identiquement distribuées, de fonction de répartition $F(t)$ telle que $1 - F(t) = Ct^{-\alpha}, C > 0$, $\alpha > 0$ et $t \geq C^{\frac{1}{\alpha}}$.

En posant

$g(u) = C^{\frac{1}{\alpha}} u^{\frac{1}{\alpha}} (1+\alpha)^{-1}$,

alors

$\lim_{u \to \infty} g(ue(u)) / g(u) = 1$, pour toute fonction $e(u)$ telle que $\lim_{u \to \infty} e(u) = 1$.

De plus nous constatons que

$$G(u) = (1-F)^{-1}(u) = C^{\frac{1}{\alpha}} u^{\frac{-1}{\alpha}}$$
$$= a + g\left(\tfrac{1}{u}\right) + \int_A^{\frac{1}{u}} \tfrac{g(v)}{v} dv,$$

avec

$a = \alpha C^{\frac{1}{\alpha}} A^{\frac{1}{\alpha}} (1+\alpha)^{-1}$.

Selon le Lemme (4.1) et le Théorème (4.15) nous avons $\{X_{n-k_n+1,n} - X_{n-k_n,n}\}$ et $\{X_{n-k_n,n} - X_{n-k_n-1,n}\}$ sont asymptotiquement indépendants et pourtant $X_{n,n}$ appartient pas au domaine d'attraction de $\Phi_\alpha(x)$

Preuve du Lemme 4.1

Montrons que si

$$L(x) = \exp\left(\eta(x) + \int_A^x \frac{\varepsilon(t)}{t} dt\right), \ A > 0,$$

alors

$$\lim_{u \to \infty} \frac{L(ue(u))}{L(u)} = 1.$$

Pour toute fonction positive $e(u)$, telle que $\lim_{u \to \infty} e(u) = 1$,

$$L(ue(u)) / L(u) = \exp(\eta(ue(u)) - \eta(u)) \exp\left(\int_u^{ue(u)} \varepsilon(t)/t \, dt\right).$$

Pour $T = \log t$

$$L(ue(u)) / L(u) = \exp(\eta(ue(u)) - \eta(u)) \exp\left(\int_{\log u}^{\log u + \log e(u)} \varepsilon\left(e^T\right) dT\right).$$

En appliquant le théorème de la moyenne

$$\lim_{u\to\infty} L(ue(u))/L(u) = \lim_{u\to\infty} \exp(\eta(ue(u)) - \eta(u)) \exp\left(\varepsilon\left(e^{\zeta}\right)\log e(u)\right),$$

où

$$\log u \le \zeta \le \log u + \log e(u).$$

Puisque

$$\lim_{u\to\infty} \eta(u) = \eta \ \ (\eta < \infty),$$

alors

$$\lim_{u\to\infty} L(ue(u))/L(u) = e^0 \lim_{u\to\infty} \exp\left(\varepsilon\left(e^{\zeta}\right)\log e(u)\right).$$

Comme $\varepsilon\left(e^{\zeta}\right)$ est bornée et $\lim_{u\to\infty} \log e(u) = 0$,

$$\lim_{u\to\infty} L(ue(u))/L(u) = e^0 e^0 = 1.$$

Montrons que si

$$\lim_{u\to\infty} L(ue(u))/L(u) = 1,$$

alors

$$L(u) = \exp\left(\eta(u) + \int_A^u \frac{\varepsilon(t)}{t} dt\right).$$

Soient

$$f(t) = \log L\left(e^t\right),\ t \ge \gamma = \log A,$$

$$n_0 = \inf\{n \in \mathbb{N} :\ n \ge \gamma\},$$

$$f_1(t) = f(n) + 3\left(f(n+k_n) - f(n) + C_2\right) \int_0^{t-n} u(1-u)\,du,$$

où (k_n) une suite réelle vérifiant, $0 \le k_n \le 1$, $\lim_{n\to\infty} k_n = 0$ et C_2 est une constante positive ou nulle.

La dérivée de $f_1(t)$ s'écrit :

$$f_1'(t) = 3\left(f(n+k_n) - f(n) + C_2\right)(t-n)\left(1-(t-n)\right).$$

Pour $n \ge n_0$ $f_1'(n) = 0$, $f_1'(t)$ est continue et

$$\left|f_1'(t)\right| \le 3\left|f(n+k_n) - f(n) + C_2\right|(t-n)\left(1-(t-n)\right).$$

Pour n tel que $k_n \geq \frac{1}{2}$, le maximum de $(t-n)(1-(t-n))$ est atteint pour $t = n + \frac{1}{2}$, dans ce cas $(t-n)(1-(t-n)) \leq \frac{1}{4}$.

Pour n tel que $k_n \leq \frac{1}{2}$, le maximum de $(t-n)(1-(t-n))$ est atteint pour $t = n + k_n$, d'où

$$(t-n)(1-(t-n)) \leq \frac{1}{4}.$$

Par conséquent $\forall t,\ n \leq t \leq n + k_n$, nous avons

$$(t-n)(1-(t-n)) \leq \frac{1}{4},\quad \left|f_1'(t)\right| \leq \frac{3}{4}\left|f(n+k_n) - f(n) + C_2\right|.$$

En appliquant le théorème de la moyenne nous avons

$$|f_1(t) - f(t)| \leq |f(n) - f(t)| + 3\left|f(n+k_n) - f(n) + C_2\right|(t-n)\xi(1-\xi),$$

où $\xi \in [0,\ (t-n)]$.

Le calcul du maximum de $(t-n)(1-(t-n))$ implique que

$$|f_1(t) - f(t)| \leq |f(n) - f(t)| + \frac{3}{4}\left|f(n+k_n) - f(n) + C_2\right|.$$

Comme $n \leq t \leq n + k_n$ et $0 \leq k_n \leq 1$, nous constatons

$\lim\limits_{t \to \infty} (f(n) - f(t)) = 0$;

$\lim\limits_{t \to \infty} (f(n+k_n) - f(n)) = 0$;

$\lim\limits_{t \to \infty} (f_1(t) - f(t)) \leq \frac{3}{4} C_2$.

Par conséquent, pour t très grand $(f_1(t) - f(t))$ et $f_1'(t)$ sont bornées. Ainsi pour tout t, $t \geq n_0$, $f_1(t)$ peut s'écrire

$$f_1(t) = f_1(n_0) + \int_{n_0}^{t} f_1'(u)\,du.$$

Posons $L_1(x) = \exp(f_1(\log x))$ donc ona

$$L_1(x) = C_1 \exp\left(\int_{n_0}^{\log x} f_1'(u)\,du\right)$$

où, C_1est une constante positive et $x \geq \exp n_0 = k$.

En posant

$$L(x)/L_1(x) = C(x),$$

alors $\forall x \geq k$,

$$\begin{aligned} L(x) &= C_1 C(x) \exp\left(\int_k^x \frac{f_1'(\log t)}{t} dt\right) \\ &= \exp\left(\eta(x) + \int_A^x \frac{\varepsilon(t)}{t} dt\right), \end{aligned}$$

où

$$\varepsilon(t) = f_1'(\log t) \text{ est bornée,}$$

$$\lim_{x\to\infty} C(x) = \lim_{x\to\infty} L(x)/L_1(x) = \exp(f_1(\log x) - f(\log x)) = C < \infty,$$

et

$$\eta(x) = \log(C_1 C(x)), \quad \lim_{x\to\infty} \eta(x) = \log(C_1 C) = \eta < \infty.$$

Preuve du Théorème 4.17

Posons

$$\begin{cases} S_{k_n+1,n} = X_{n-k_n+1,n} - X_{n-k_n,n} \\ S_{k_n,n} = X_{n-k_n,n} - X_{n-k_{\bar{n}}-1,n}, \end{cases}$$

donc

$$\begin{cases} S_{k_n+1,n} = G(U_{k_n,n}) - G(U_{k_n+1,n}) \\ S_{k_n,n} = G(U_{k_n,+1n}) - G(U_{k_n+2,n}). \end{cases}$$

où

$U_{1,n} < U_{2,n} < \ldots < U_{n,n}$ est la statistique d'ordre associée à $U_1, \ldots, U_n$ une suite de variables aléatoires indépendantes de loi uniforme.

D'après la condition 2 du Théorème.(4.17), on a :

$$\begin{cases} S_{k_n+1,n} = g(1/U_{k_n,n}) - g(1/U_{k_n+1,n}) + \int\limits_{1/U_{k_n+1,n}}^{1/U_{k_n,n}} \frac{g(v)}{v} dv; \\ S_{k_n,n} = g(1/U_{k_n+1,n}) - g(1/U_{k_n+2,n}) + \int\limits_{1/U_{k_n+2,n}}^{1/U_{k_n,+1n}} \frac{g(v)}{v} dv. \end{cases}$$

$$\begin{cases} S_{k_n+1,n} = g\left(1/U_{k_n,n}\right) - g\left(1/U_{k_n,n}\left(1 - \left(U_{k_n+1,n} - U_{k_n,n}\right)/U_{k_n+1,n}\right)\right) + \int\limits_{1/U_{k_n+1,n}}^{1/U_{k_n,n}} \frac{g(v)}{v} dv; \\ S_{k_n,n} = g\left(1/U_{k_n+1,n}\right) - g\left(1/U_{k_n+1,n}\left(1 - \left(U_{k_n+2,n} - U_{k_n,+1n}\right)/U_{k_n+2,n}\right)\right) + \int\limits_{1/U_{k_n+2,n}}^{1/U_{k_n,+1n}} \frac{g(v)}{v} dv.; \end{cases}$$

Comme

$\left(U_{k_n+1,n} - U_{k_n,n}\right)/U_{k_n+1,n} = O_p(1)$,

$P\left(\left|U_{k_n+1,n}/\frac{k_n}{n} - 1\right| \geq \varepsilon\right) \to 0$, quand $k_n \to \infty$ et $k_n = o(n)$; et pour tout $\varepsilon > 0$ il existe

$0 < \lambda \equiv \lambda_\varepsilon < 1$, $A_{n\varepsilon}$ tel que $P\left(A_{n\varepsilon}\right) > 1 - \varepsilon$,

alors sur $A_{n\varepsilon}$ nous avons

$$\begin{cases} S_{k_n+1,n} = g\left(1 - \left(k_n + 1/n\right)\lambda\varepsilon_n\right) - g\left(\left(1 - \left(k_n + 1/n\right)\lambda\varepsilon_n\right)\left(1 - \lambda_n\right)\right) \\ \qquad\qquad +g\left(\zeta_n\right)\log U_{k_n+1,n}/U_{k_n,n}; \\ S_{k_n,n} = g\left(\left(1 - \lambda\left(k_n/n\right)\varepsilon_n'\right)\right) - g\left(\left(1 - \lambda\left(k_n/n\right)\varepsilon_n'\right)\left(1 - \lambda_n'\right)\right) \\ \qquad\qquad g\left(\zeta_n'\right)\log U_{k_n+2,n}/U_{k_n+1,n} \end{cases}$$

(voir Shorack et Wellner (1986, p. 419))

où

$\left(1 - \left(k_n + 1/n\right)\lambda\varepsilon_n\right)\left(1 - \lambda_n\right) \leq \zeta_n \leq 1 - \left(k_n + 1/n\right)\lambda\varepsilon_n$ et $\left(1 - \lambda\left(k_n/n\right)\varepsilon_n'\right)\left(1 - \lambda_n'\right) \leq$

$\zeta_n' \leq 1 - \lambda\left(k_n/n\right)\varepsilon_n'$ et $\varepsilon_n \to 0$, $\lambda_n \to 0$, $\lambda_n' \to 0$, $\varepsilon_n' \to 0$ quand $n \to \infty$.

A l'infini, on a

$$\begin{cases} S_{k_n+1,n} = \frac{-1}{k_n} g\left(\zeta_n\right)\log\left(U_{k_n,n}/U_{k_n+1,n}\right)^{k_n} \\ S_{k_n,n} = \frac{-1}{k_n+1} g\left(\zeta_n'\right)\log\left(U_{k_n+1,n}/U_{k_n+2,n}\right)^{k_n+1} \end{cases}$$

Comme

$\left(U_{k_n,n}/U_{k_n+1,n}\right)^{k_n}$ et $\left(U_{k_n+1,n}/U_{k_n+2,n}\right)^{k_n+1}$ sont asymptotiquement indépendants ,

donc

$S_{k_n+1,n}$ et $S_{k_n,n}$ le sont aussi.

4.4 Application des statistiques d'ordre à l'estimation

Discutons maintenant le rôle des statistiques d'ordre dans l'estimation statistique. Illustrons d'abord l'influence des statistiques d'ordre dans l'estimation de la densité par la méthode de l'histogramme avec des partitions aléatoires.

Soient $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ la statistique d'ordre associée à $X_1, \ldots, X_n$, une suite de variables aléatoires de densité $f(x)$ et de fonction de répartition $F(x)$ et ß_n la σ-algèbre engendrée par $X_1, \ldots, X_n$. Pour tout réel x, pour tout entier naturel n, nous considérons des suites aléatoires $\{A_n(x)\}$ et $\{B_n(x)\}$ telles que, $A_n(x)$ et $B_n(x)$ soient ß_n mesurables et

$$P\left[\begin{array}{c} 1 \leq B_n(x) - A_n(x) \leq k_n,\ 1 \leq A_n(x) \\ < B_n(x) \leq n,\ \ X_{A_n(x),n} \leq x \leq X_{B_n(x),n} \end{array}\right] = 1$$

où $\{k_n\}$ est une suite d'entiers positifs. Par analogie avec la définition de l'histogramme qui est un estimateur de la densité basé sur des intervalles non aléatoires, l'estimateur suivant

$$f_n(x) = \begin{cases} \frac{B_n(x) - A_n(x)}{n\left(X_{B_n(x),n} - X_{A_n(x),n}\right)} & \text{si } X_{A_n(x),n} \leq x \leq X_{B_n(x),n}, \\ 0 & \text{si } x < X_{1,n} \text{ ou } x > X_{n,n}. \end{cases}$$

est un estimateur de $f(x)$, basé sur des intervalles aléatoires. Si pour tout n, les suites $A_n(x)$ et $B_n(x)$ sont choisies telles que $X_{A_n(x),n} \leq x \leq X_{B_n(x),n}$, alors $f_n(x)$ peut s'écrire

$$f_n(x) = \frac{F_n\left(X_{B_n(x),n}\right) - F_n\left(X_{A_n(x),n}\right)}{X_{B_n(x),n} - X_{A_n(x),n}},$$

où $F_n(.)$ est la fonction de distribution empirique, qu'on peut considerer aussi comme un estimateur de $f(x)$. D'autres choix des suites $\{A_n(x)\}$ et $\{B_n(x)\}$ peuvent donner une alternative d'estimateurs de $f_n(x)$. A savoir, si $\{k_n\}$ est une suite d'entiers positifs, $A_n(x) = \max\left\{R_n(x) - \frac{k_n}{2},\ 1\right\}$ et $B_n(x) = \min\left\{R_n(x) + \frac{k_n}{2},\ n\right\}$, où $R_n(x) = nF_n(x)$. $f_n(x)$ peut s'écrire

$$f_{n_1}(x)=\begin{cases}\frac{k_n}{n\left(X_{j+k_n/2,n}-X_{j-k_n/2,n}\right)} & \text{si } X_{j,n}\le x<X_{j+1,n},\\ \qquad j=k_n/2+1,\ldots,n-k_n/2\,, & \\ \frac{j-1+k_n/2}{n\left(X_{j+k_n/2,n}-X_{1,n}\right)} & \text{si } X_{j,n}\le x<X_{j+1,n},\\ \qquad j=1,\ldots,k_n/2\,, & \\ \frac{n-j+k_n/2}{n\left(X_{n,n}-X_{j-k_n/2,n}\right)} & \text{si } X_{j,n}\le x<X_{j+1,n},\\ \qquad j=n+1-k_n/2\,,\ldots,n-1, & \\ 0 & \text{si } x<X_{1,n} \text{ ou } x>X_{n,n}.\end{cases}$$

cet estimateur ainsi définie est appelé estimateur symétrique de $f(x)$.

D'autre part, si on considère $\{k_n\}$ une suite d'entiers positifs, $I=\{1,\ k_n+1,\ 2k_n+1,\ldots\}$, $A_n(x)=\max\{\alpha\mid\alpha\in I,\ X_{\alpha,n}\le x\}$, $B_n(x)=\min\{A_n(x)+k_n,\ n\}$ et r_n est le plus grand entier appartenant à I, tel que $r_n\le n-k_n$, l'estimateur $f_n(x)$ devient

$$f_{n_2}(x)=\begin{cases}\frac{k_n}{n\left(X_{\alpha+k_n,n}-X_{\alpha,n}\right)} & \text{si } X_{\alpha,n}\le x<X_{\alpha+k_n,n},\\ \qquad \alpha=1,\ k_n+1,\ldots,r_n, & \\ \frac{n-r_n-k_n}{n\left(X_{n,n}-X_{r_n+k_n,n}\right)} & \text{si } X_{r_n,n}\le x<X_{n,n}\\ 0 & \text{si } x<X_{1,n} \text{ ou } x>X_{n,n}\end{cases}$$

Nous constatons que $f_{n_2}(x)$ est constant sur les intervalles aléatoires disjoints $[X_{1,n},\ X_{k_n+1,n})$, $[X_{k_n+1,n},\ X_{2k_n+1,n}),\ldots,[X_{r_n-k_n,n},\ X_{r_n,n})$, $[X_{r_n,n},\ X_{n,n})$, de plus il a la forme de l'histogramme, cet estimateur ainsi défini est appelé estimateur non symétrique de $f(x)$.

$f_{n_2}(x)$ et $f_{n_1}(x)$ sont des estimateurs consistants de $f(x)$ et si de plus $f(x)$ est continue, $k_n\to\infty$, $\frac{k_n}{n}\to 0$, quand $n\to\infty$,

alors

$$\frac{F_n\left(X_{B_n(x),n}\right)-F_n\left(X_{A_n(x),n}\right)}{X_{B_n(x),n}-X_{A_n(x),n}}\longrightarrow f(x)\ \ p.s,\ \text{ quand } n\to\infty.$$

Nous achevons ce paragraphe par l'estimateur de Hill construit à partir des statistiques d'ordre. Cet estimateur caractérise les distributions dont la taille varie régulièrement à l'infini. Il appartient à la classe des estimateurs à noyaux introduite par Csörgo, Deheuvels, et

Mason (1983). Soit k_n une suite d'entiers telle que

$$0 < k_n < n, \quad k_n \to \infty, \quad k_n/n \to 0, \text{ quand } n \to \infty. \tag{4.1}$$

La statistique

$$T_n = k_n^{-1} \sum_{i=1}^{k_n} (\log X_{n-i+1,n} - \log X_{n-k_n,n})$$

est appelée l'estimateur de Hill.

Si k_n vérifie la condition (4.1) et si pour tout $\lambda > 0$,

$$\lim_{x \to \infty} \frac{1 - F(\lambda x)}{1 - F(x)} = \lambda^{-\alpha},$$

alors

$$T_n \xrightarrow{P} \frac{1}{\alpha}, \quad \text{quand } n \to \infty.$$

Mason (1982) montre que l'estimateur de Hill caractérise les distributions appartenant au domaine d'attraction de Fréchet, Lo Gane (1985) a donné le comportement asymptotique de cet estimateur quand la distribution appartient au domaine d'attraction de Gumble.

RÉFÉRENCES

Akaike, H. (1954). An approximation of the density function. *Ann. Inst. Statist. Math.*, **6**, 127-132.

Balkema, A.A. and De Hann, L. (1978). *Limit distributions for order statistics.* I. Teor. Veroyat. Primen. 80-96; II. 358-75. [258].

Barlett. M. S. (1963). Statistical estimation of density function.*Sankhyã Ser.* **25**, 245-245.

Berlinet, A. (1993). Hierarchies of higher order kernels. *Prob. Theor. Rel. Fields* **94,** 489-504.

Bochner, S. (1955). *Harmonic Analysis and the Theory of probability.* University of Chicago Press, Chicogo, Illinois.

Bosq, D. et Lecoutre, J. P. (1987). *Théorie de l'Estimation Fonctionnelle.* Economica, Paris.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71** 353-360.01

Chibisov. D.M. (1964). Some theorems on the limiting behavior of empirical distribution functions. *Selected Transl. Math. Statist. Prob.*,**6**, 147-156.

Chow, Y. S, Geman, S. and Wu, L. D. (1983). Consistent cross validation density estimation. *The annals of statistics.* **11**, 125-35

Csörgo, S., Deheuvels, P., Mason, D. M. (1984). Kernel estimates of a tail of of a distribution. *Preprint n°* **15**, L.S.T.A.

David. H. A. (1970). *Order Statistic.* John Wiley and Sons, New York.

Deheuvels. P. (1977). La densité par histogrammes généralisés. *Revue de Stat. Appl.* **25,** 5-42.

Deheuvels. P. (1982). Strong limiting bounds for maximal uniform spacing. *Ann. Probab.* **10,** 1058-1065.

Deheuvels, P. (1983). *Upper bounds for k-th maximal spacings. Z. Wahrsch. verw. Gebiete.* **62**, 465-474.

Deheuvels, P. (1986). On the influence on the extremes of an i.i.d sequences onthe maximal spacings. *Ann.Probab.* **1**, 194-208.

Deheuvels, P. and Mason, D. M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.* **20**, 1248-1287.

Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Infer. Soc. Processes*, **7**, 225-277.

De Hann, L. (1975). *On regular variation and its application the Weak convergence of sample extremes.* Mathematical Centre tracts 32, Amsterdam.

De Hann, L. and Risnick, SI. (1984). Asymptically balanced functions and stochastic compactenes of sample extremes. *Ann. Probab.* **2**, 588-608.

Derzko, G. and Deheuvels, P. (2002). Estimation non-paramétrique de la régression dichotomique-application biomédicale. *C.R. Acad. Sci. Paris Ser.* **I 334** 59-63.

Devroye, L. (1978b). The uniform convergence of the Nadaraya-Watson regression function Estimate. *Can. J. Statist.*, **6**, 179-191.

Devroye, L. (1981). Laws of the iterated logarithm for ordr statistics of uniforms spacings. *Ann. Probab.* **9** 806-867.

Devroye, L. (1982). A log log law for maximal uniform spacings. *Ann. Probab.* **10** 863-868.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation :* The view. Wiley, New York.

Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation.* Springer Series in Statistics. Springer-Verlag, New York.

Doeblin, W.(1940). Sur l'ensemble de puissances d'une loi de probabilité. *Studia. Math.* **9** 71-96.

Duin, R. P. W. (1976). On the choise of smoothing j parameters for Parzen estimators of probability density functions. *IEEE Trans.Comput.* C-**25**, 1175-1179.

Dvoretzky, A.(1956). Asymptotic minimax character of the sample distribution function and the classical multinomal estimator. *Ann Math. Statist.* **27**, 642-669.

Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoretical Prob.*, **13,** 1-37.

Epanechnikov,V. A. (1969). Nonparametric estimation of multi-dimensional probability density. *Theory Probab. Appl.* **14**, 153-158.

Fix, E. and Hodges, J. L. Jr. (1951). Discriminatory analysis, nonparametric discrimination : Small sample properties. *Report No. 11, USAF School of Aviation Medicine, Texas.*

Galombos, J. (1978). *The asymptotic theory of extreme order statistics.* Wiley, New York.

Galombos, J. and Kortz, S. (1978). *Characterizations of Probability Distributions.* Lecture of Notes in Mathematics 675. Springer, Berlin and New York.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44,** 423-453.

Habbema, J. D. F., Hermans, J. and Van Der Broek, K.(1974). A stepwise discriminant analysis program using density estimation. In *COMPSTAT 1974, Proceedings in Computational Statistics, Vienna (G. Bruckman ed.)* 101-110. Physica, Heidelberg.

Hall, P. (1978). Representations and limit theorems for extremes value distributions. *J. Appl. Prob.* **15,** 639-644.

Hall, P. (1982). Cross validation in density estimation. *Biometrika.* **69,** 383- 390.

Hall, P. (1983). Large Sample optimality of least Squares Cross-validation in density estimation. *Ann. Stat.* **11**, 1156-1174.

Hall, P. and Marron, J. S. (1991). Lower bounds for bandwidth selection in density estimation. *Probab. Theor. Rel. Fields.* **90** 149-173.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

Härdle, W., Jansen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, 1428-1449.

Izenman, A. J. (1991). Recents developments in nonparametric density estimation. *J. Amer. Statist. Assoc.*, **86**, no. 413, 205-224.

Kappenman, E. L. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Leadbetter, M. R. (1963). *On the nonparametric estimation of probability densities.* Ph. D. Thesis, University of North Carolina

Loftsgaarden, D. O. and Quesenberry, C. P. (1965). A nonparamitric estimate of multivariate density function. *Ann. Math. Statist.* **36**, 1049-1051.

Lo Gane Samb (1985). Asymptotic behavior of the C-D-M's kernel estimate of the tail of distribution. *Preprint* , L.S.T.A

Mack, Y. P., Rosenblatt, M. (1979). Multivariate K-nearest neighbor density estimates. *J. Multivariate. Anal.* **9**, 1-15.

Marron, J. S. (1989). Comments, on data driven bandwith selector. *Comp. Stat. Data Anal.* **8**, 155-170.

Marron, J. S. (1992). *Bootstrap Bandwidth Selection , in Exploring the Limits of Bootstrap*, eds. R. LePage and L. Billard. New York : John Willey.

Mielniczuk, J., Sarda, P., and Vieu, P. (1989). Local Data-Driven Bandwidth Choice for Density Estimation. *Journal of Statistical Planning and Inference.* **23**, 53-69.

Moore, D. S. and Yackel, J. W. (1977a). Consitency properties of nearest neighbor density function estimators. *Ann. Statist.* **5**, 143-154.

Nadaraya, E. A. (1964). On estimating regression. *Theor. Prob. Appl.*, **9**, 141-142.

Nadaraya, E. A. (1976). On the nonparametric estimator of Bayesian risk in the classification problem, *Proc. AN. Georg SSR*, **82**, 2, pp. 277-280 (in Rusisan).

Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves.* Kluwer, Dordrecht

Nemouchi, N. et Mohdeb, Z.(2005). Regression non paramétrique dans modèle gaussien. *Sciences & Technologie* C. N^0 **24.**

Nemouchi, N. and Mohdeb, Z. (2007). Asymptotic confidence bands for nonparametric function estimates in the Gaussian cases. *Far East. J. Theo. Stat.*, **21,** 1, 65-81.

Park, B. U. and Marron, J. S. (1990). Comparison of data- driven bandwidth Selectors. *J. Amer. Statist Assoc.* 8566-8572.

Parthazarthy, K.R. (1969). *Probability Measures on Metric Spaces.* Academic Press, New York.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065-1076.

Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

Pyke, R. (1965). Spacings. *J. Roy. Statist. Soc.* **27,** 395-436.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832-837.

Roussas, G. (1990). *Nonparametric Functional Estimation and Related Topics.* NATO ASI series 355. Kluwer, Dordrecht.

Rumedo, M. (1982). Empirical choise of Histograms and kernel density estimators. *Scand. J. Statist.* **9**, 65-78.

Schucany, W. R. (1989). Locally optimal window widths for kernel density estimation with large samples. *Statist. Probab. Letters* **7** 401-405.

Schuster, E.F. (1970). Note on uniform convergence of density estimates. *Ann. Math. Statist.* **41,** 1347-1348.

Schuster, E.F. and Gregory, G. G. (1981). On the nonconsistency of maximum likelihood nonparametric density estimators. In computer Science and statistic; *Proceedings of the 13th Symposium on the Interface (W. F. Eddy, ed).* 295-298. Springer, Berlin.

Scott, D. W. (1992). *Multivariate Density Estimation-Theory, Practice and Visualization.* Wiley, New York.

Scott, D. W. and Terrel, G. R. (1987). Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* **82,** 1131-1146.

Seneta, E. (1975). *Regulary Varying Functions,* Lecture Notes in Mathematics 508. Springer, Berlin.

Serfling, R. J. (1979). A variation on Scheffe's theorem with application to nonparametric density estimation. *Technical report M 502. Florida State University.*

Sheather, S. J. (1992). The performance of six popular bandwidth selection methods on some real datasets. *Comput. Statist.* **7,** 225-250

Shorack, G. and Wellner, J.(1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons, New York.

Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *J. Res. Nat. Bur. Stand.* **64B**, 145-50.[258]

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Stone, C. J. An asymptotically optimal window selection rule for kernel estimates. *Ann. Stat.* **12**, 1285-1297.

Stute, W. (1986a). Conditional empirical processes. *Ann. Statist.* **14**, 638-647.

Stute, W. (1986b). On almost sure convergence of conditional empirical distribution functions. *Ann. Prob.* **14,** 891-901.

Terrel, G. R. (1990). The maximal Smoothing Principle in Density Estimation. *Journal of American Statistical Association.* **85,** 470-477.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhya A*, **26,** 359-372.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Stastist. Ass.* **364,** 812-815.

Woodroofe, M. (1970). On choosing a Delta-Sequence. *Annals of Mathematical statistics.* **41,** 1665-1671.

Résumé : L'objet de ce travail est de construire des estimateurs de régression non paramétrique asymptotiquement optimaux, sous l'hypothèse que les lois sous-jacentes sont gaussiennes. Les résultats que nous obtenons présentent l'intérêt d'être directement applicables en analyse exploratoire des données. Nous établissons aussi la caractérisation des lois dont les espacements d'une statistique d'ordre d'une suite de variables aléatoires indépendantes et identiquement distribuées sont asymptotiquement indépendants. En se basant sur les résultats de Deheuvels (1982-1983) et en s'inspirant d'un Théorème de Hall (1978) et Weissman (1978) , nous montrons que cette indépendance asymptotique associée à certaines conditions peut caractériser les lois dont les extrêmes sont attirées par la loi de Gumble.

Mots clés. Estimation de la densité, estimation de la régression, estimation à noyau, bandes de confiance, Statistique d'ordre, indépendance asymptotique, compacité stochastique, domaine d'attraction.

Abstract : In this work, a method of constructing nonparametric asymptotically optimal estimators is presented. Assuming Gaussian underlying distributions, straightforward applications in exploratory data analysis may be deduced from our results. We etablish the characterisation of laws of which spacings of a order statistics of an independent random variables identically distriduted are asymptotically independent. While taking results of Deheuvels as basis (1982-1983) and while being inspired by a Hall theorem (1978) and Weissman (1978), we show that this asymptotic independence associaded to others conditions can characterize laws whose extremes are attracted by the Gumble law.

Keys words : Order statistics, Asymptotic independence, Stochastic compacteness, Attraction domain, Density estimation, Regression estimation, Kernel estimation, Confidence bands.