

50376
1980
208

N° d'ordre 844

50376
1980
208

THÈSE

présentée à

l'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE I

pour obtenir

LE GRADE DE DOCTEUR DE 3^{ème} CYCLE



Bernard DEREMETZ

**EFFICACITÉ DE
CERTAINS ESTIMATEURS FONCTIONNELS**

Membres du Jury : D. BOSQ, Président
D. BOSQ, Rapporteur
P. JACOB
F. MIGNOT Examineurs
J. BLEUEZ

Soutenue le 10 septembre 1980

Que Monsieur le Professeur Denis BOSQ trouve ici l'expression de ma profonde reconnaissance, tout d'abord pour la qualité de ses enseignements de second et troisième cycles qui m'ont donné le goût de faire cette thèse, ensuite pour l'aide efficace et les conseils éclairés qu'il m'a prodigués pendant toute la durée de mes travaux, et enfin pour avoir accepté de présider un jury à une époque de l'année peu propice à une soutenance.

Messieurs les Professeurs Fulbert MIGNOT et Pierre JACOB et Monsieur Jacques BLEUEZ m'ont fait l'honneur de juger ce travail et je les en remercie vivement.

J'exprime aussi toute ma gratitude aux membres du Séminaire de Statistique de l'U.E.R. de Mathématiques Pures et Appliquées de Lille I et en particulier à Luc NOBÉCOURT pour sa précieuse aide sur l'estimation bayésienne.

Je tiens aussi à rendre un hommage ému à notre regretté camarade Christian GUILBART, pour le soutien qu'il m'avait apporté au début de mes recherches.

Tous mes remerciements aussi aux personnes qui ont contribué à la réalisation de cette thèse et principalement à Raymonde BÉRAT et Arlette LENGAIGNE pour la compétence et le zèle dont elles ont fait preuve lors de la dactylographie.

INTRODUCTION

L'estimation de la densité a déjà fait l'objet de nombreux travaux, et on pourra consulter [32], qui comporte une revue bibliographique très complète sur les différentes méthodes d'estimation et sur les principaux résultats obtenus. Les deux grandes méthodes spécialement étudiées sont :

- * la méthode du noyau d'après Rosenblatt en 1956.
- * la méthode par projection d'après Cencov en 1962.

Certains articles contiennent des résultats de convergence de ces estimateurs dans le cas de l'estimation ponctuelle et globale, et il nous a semblé intéressant de faire le point sur ce sujet, en apportant des améliorations dans quelques cas précis. La publication [64] nous a permis d'unifier les résultats précédents, en donnant la vitesse de convergence de la famille générale d'estimateurs de Bosq et Bleuez [10], [11] et ceci sous certaines conditions vérifiées par les noyaux.

Nous avons ensuite donné les minoration des risques d'estimateurs de la fonction de répartition, de la loi et de la densité à l'aide de l'inégalité de Cramer-Rao pour un paramètre hilbertien d'après [8].

Les travaux de Farrell et Wahba, [22] et [59], Bretagnolle et Hubert [13] établissent les bornes inférieures de ces inégalités de Cramer-Rao dans le cas de l'estimation ponctuelle pour les premiers, et de l'estimation globale pour les seconds, ce qui nous a permis de montrer que certains estimateurs possèdent la propriété d'être asymptotiquement minimax sur des ensembles de densités bien précis.

Dans le premier chapitre de cette thèse, nous avons donc rappelé ce qu'est un estimateur minimax, ce mode d'estimation étant peu connu et peu utilisé depuis les travaux anciens de Wald en 1945, Girshick et Savage en 1951, Hodges et Lehman en 1950, etc. Récemment, Aggarwal (1967), Phadia (1973), Fergusson (1974) ont déterminé les estimateurs minimax de la fonction de répartition pour différentes fonctions de perte. L'approche de Phadia et Fergusson est la plus intéressante, car la détermination de l'estimateur minimax se fait à partir d'une suite de lois a priori sur l'ensemble des fonctions de répartition.

Nous avons utilisé leur méthode afin de déterminer l'estimateur minimax de la densité dans le cas de la famille paramétrée normale. Cet exemple explicite l'estimateur, dont l'existence est assurée par le travail de Wertz (en 1974) et dont nous avons étendu les résultats à différentes fonctions de la densité.

Enfin dans le dernier chapitre nous avons étudié plus spécialement l'estimateur par projection et les estimateurs de Whittle [69], Wahba [60] et Brunk [14], qui apparaissent comme des améliorations de l'estimateur précédent. Les auteurs utilisent l'information sur l'ensemble des densités afin de pondérer les éléments de la base hilbertienne. Les coefficients alors introduits ont pour effet d'atténuer les oscillations des courbes, et d'améliorer dans certains cas la vitesse de convergence de l'estimateur, ainsi que d'obtenir un estimateur positif.

Efficacité de certains estimateurs fonctionnels.

CHAPITRE I : Estimation minimax.

(A) Estimation ponctuelle.

- I - Position du problème. Notations.
- II - Estimation minimax.
- III - Estimation I-minimax.

(B) Estimation globale.

- I - Estimation de la fonction de répartition.
 - 1. Première technique d'obtention de l'estimateur minimax.
 - a. Dans le cas de la fonction de perte L_1 .
 - b. Dans le cas de la fonction de perte L_3 .
 - 2. Seconde technique d'obtention de l'estimateur minimax.
 - a. Premier exemple d'une suite de lois a priori.
 - b. Deuxième exemple d'une suite de lois a priori.
- II - Estimation de fonctions de la densité.
 - 1. Existence d'estimateurs Minimax de $\Psi(f)$.
 - a. $F \xrightarrow{\Psi} A$ $A = \{\text{fonctions mesurables de } \mathbb{R} \rightarrow \mathbb{R}\}$
 - b. $F \xrightarrow{\Psi} \mathbb{R}$.
 - 2. Situation dans laquelle l'estimateur minimax de $\Psi(f) = f$ est unique, $f \in F_{\theta}$.
 - 3. Compléments sur la continuité des applications R_2 et δ_2 pour la distance de l'espace Ω_2 .
 - 4. Détermination de l'unique estimateur minimax de la densité dans le cas de la famille normale.
 $F_{\theta} = \{N(\theta, 1)(x)\}$.

CHAPITRE II : Efficacité des estimateurs de la loi.

(A) Efficacité des estimateurs de la densité dans le cas de perte quadratique.

- I - Etude de la convergence en moyenne quadratique sur $W_P^m(M), Z_P^m$.

1. Minoration du risque R sur $W_P^m(M)$.
2. Vitesse atteinte par certains estimateurs sur $W_P^m(M)$ et Z_2^m .
3. Conditions sur le noyau $K_{r(n)}$ pour que la famille générale d'estimateurs \hat{h}_n converge à la même vitesse qu'en 2. sur Z_P^m et W_P^m .

II - Etude de la convergence en moyenne quadratique sur C_λ .

1. Minoration du risque R sur C_λ^u .
2. Conditions sur le noyau $K_{r(n)}$ pour que la famille générale d'estimateurs \hat{h}_n converge à la vitesse maximale sur C_λ .

(B)

Efficacité des estimateurs dans le cas de la perte quadratique intégrée.

I - Inégalité de Cramer-Rao pour un paramètre hilbertien.

1. Notations et principaux résultats de [8].
2. Application à l'estimation de la fonction de répartition.
3. Application à l'estimation de la loi.
4. Complément sur l'inégalité de Cramer-Rao dans le cas de l'estimateur \hat{h}_n de la densité.

II - Etude des vitesses de convergence d'estimateurs de la densité dans le cas de la perte quadratique intégrée.

1. Minoration du risque sur W^m .
2. Vitesse atteinte par l'estimateur \hat{h}_n^1 sur W^m .
3. Vitesse atteinte par l'estimateur $\hat{h}_n^{2,H}$ sur $W_{L_0}^m$.
4. Vitesse atteinte par l'estimateur $\hat{h}_n^{2,T}$ sur W_π^m .

III - Détermination d'une famille d'estimateurs asymptotiquement minimax pour le mise.

1. Première détermination d'une famille asymptotiquement minimax sur W^m pour le mise.
2. Conditions générales sur le noyau $K_{r(n)}$.

CHAPITRE III : Etude de la famille d'estimateurs de noyaux

$$K_n(x,t) = \sum_{i=1}^{\infty} \lambda_{i,n} e_i(x) e_i(t).$$

I - Comparaison des estimateurs \hat{h}_n^1 et \hat{h}_n^2 .

II - Effet d'un changement de base sur l'estimateur \hat{h}_n^2 .

III - Etude de la famille d'estimateurs \hat{h}_n de noyaux

$$K_n(x,t) = \sum_{i=0}^{\infty} \lambda_i(n) e_i(x) e_i(t).$$

1. Rappels bibliographiques.

2. Estimateurs de Whittle-Brunk.

a. Comparaison du mise entre \hat{h}_n^2 et $\hat{h}_n^{(\lambda_i)}$.

b. Utilisation des données afin de déterminer les variances a priori $\sigma_i^2 = 1/\pi_i$.

c. Estimation de la hauteur des vagues.

3. Estimateurs de Whittle-Wahba.

a. Vitesse de convergence au sens du mise de $\hat{h}_n^{\lambda,m}$.

b. Détermination de λ et m .

4. Méthodes d'estimation préconisées.

IV - Détermination d'une nouvelle fonction de perte, qui tient compte de l'information sur F .

1. Etude dans le cas de l'estimateur $\hat{h}_n^{2,T}$.

2. Etude dans le cas de l'estimateur \hat{h}_n^1 .

V - Courbes.

CHAPITRE 1

ESTIMATION MINIMAX

(A) Estimation ponctuelle.

Avant de traiter le problème de l'estimation globale de la densité, nous allons énoncer les théorèmes fondamentaux sur l'estimation minimax en traitant l'estimation d'un paramètre réel.

I - Position du problème - Notations.

. Soient (Ω, \mathcal{A}, P) un espace probabilisé et (E, \mathcal{B}, μ) un espace mesuré : μ étant une mesure σ finie sur \mathcal{B} .

. Soit $\mathcal{P}_\theta = \{P_\theta, \theta \in \Theta\}$ $\Theta \subset \mathbb{R}$ une famille paramétrée de probabilités sur (E, \mathcal{B}) dominée par μ et X une variable aléatoire $(\Omega, \mathcal{A}, P) \xrightarrow{X} (E, \mathcal{B})$ dont la loi P_X inconnue appartient à \mathcal{P}_θ .

On prendra $E = \mathbb{R}$, $\mathcal{B} = \mathcal{B}_{\mathbb{R}}$ tribu des boréliens de \mathbb{R} .

. Soit X_1, \dots, X_n un échantillon de taille n de la variable aléatoire X , on se propose de déterminer l'estimateur minimax et I minimax de $\theta \in \Theta$ sachant que $P_X \in \mathcal{P}_\theta$.

. On désigne par \mathcal{H} l'ensemble des mesures de probabilités définies sur l'espace mesuré $(\Theta, \mathcal{L}, \zeta)$ où

L	est la tribu des Boréliens de Θ .
ζ	est une mesure σ finie sur (Θ, \mathcal{L}) .

. On note \mathcal{D} l'ensemble des estimateurs de θ : applications mesurables définies sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ et à valeurs dans (Θ, \mathcal{L}) et on note :

- . $f(x, \theta)$ la densité $\frac{dP_\theta}{d\mu}$ $P_\theta \in \mathcal{P}_\theta$
- . $h(\theta)$ la densité $\frac{dH}{d\zeta}$ $H \in \mathcal{H}$.

. La fonction de perte étant quadratique, la fonction de risque sera égale à $d \in \mathcal{D}$

$$R(\theta, d) = \int_{\mathbb{R}^n} [d(x_1, \dots, x_n) - \theta]^2 f(x_1, \theta) \dots f(x_n, \theta) d\mu(x_1) \dots d\mu(x_n)$$

On se placera dans l'hypothèse suivante :

$$\forall d \in \mathcal{D}$$

$$\forall \theta \in \Theta \quad R(\theta, d) \leq M < \infty$$

. Pour une loi a priori $H \in \mathcal{H}$ le risque bayésien sera égal à

$$r(H, d) = \int_{\Theta} R(\theta, d) dH(\theta) = \int_{\Theta} R(\theta, d) h(\theta) d\zeta(\theta)$$

Dans certaines situations, l'existence d'une loi a priori sur l'ensemble des paramètres Θ ne peut être justifiée, ou alors cette loi existe, mais l'information dont on dispose est insuffisante pour la déterminer. Dans ce cas l'estimation bayésienne (voir par exemple [27], [71], ...) est difficilement justifiable et une procédure plus adaptée et surtout plus prudente est la procédure minimax (voir [62], [26], [27], [31] ... [43] etc).

II - Estimation minimax.

Les définitions suivantes sont équivalentes (voir [71] ou [27])

Définition 1. - Un estimateur d^* minimax réalise

$$\sup_{\theta \in \Theta} R(\theta, d^*) = \inf_{d \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, d)$$

Définition 2. - Un estimateur d^* minimax réalise

$$\sup_{H \in \mathcal{H}} r(H, d^*) = \inf_{d \in \mathcal{D}} \sup_{H \in \mathcal{H}} r(H, d)$$

Nous pouvons alors citer deux théorèmes qui donnent des conditions suffisantes pour qu'un estimateur soit minimax.

Théorème 1. - voir [30], [31] ...

Soient $P_\theta = \{P_\theta, \theta \in \Theta\}$ et \mathcal{D} une famille d'estimateurs de θ . Supposons que d^* élément de \mathcal{D} soit bayésien pour une distribution a priori ξ^* élément de H , et tel que $R(\theta, d^*) = \text{cte}$ $\forall \theta \in \Theta$: Alors d^* est un estimateur minimax de θ .

Exemple (1) [30]. Problème de l'estimation de θ paramètre d'une loi binomiale, en une observation.

Soit X la variable aléatoire suivant une loi $B(m', \theta)$ avec θ inconnu, $0 < \theta < 1$. Dans le cas de la fonction de perte quadratique, on montre que l'estimateur

$$d^*(X) = \frac{X}{m'} \frac{\sqrt{m'}}{\sqrt{m'+1}} + \frac{1}{2(\sqrt{m'+1})}$$
 est bayésien pour la loi beta

a priori ξ^* de paramètres $\sqrt{\frac{m'}{2}}$. Notation $B_e(\sqrt{\frac{m'}{2}}, \sqrt{\frac{m'}{2}})$.

On vérifie aisément que :

$$R(\theta, d^*) = \frac{1}{4(\sqrt{m'+1})^2} = \text{cte} \text{ et } d^* \text{ est un } \underline{\text{estimateur minimax}} \text{ de } \theta.$$

Théorème 2. - Soit $\{\xi_k : k = 1, 2, \dots\} \subset H$ une suite de lois a priori sur Θ et $\{d_k : k = 1, 2, \dots\}$; $\{r(\xi_k, d_k) : k = 1, 2, \dots\}$

les suites correspondantes des estimateurs bayésiens et des risques bayésiens. Si d^* est un estimateur de θ dont la fonction de risque satisfait à

$$\sup_{\theta \in \Theta} R(\theta, d^*) \leq \limsup_{k \rightarrow \infty} r(\xi_k, d_k)$$

Alors d^* est un estimateur minimax de θ .

Corollaire. - Soit d^* un estimateur ayant un risque constant : $R(\theta, d^*) = \rho^*$, $\forall \theta \in \Theta$. S'il existe $\{\xi_k : k = 1, 2, \dots\} \subset H$ une suite de lois a priori telle que $\lim_{k \rightarrow \infty} r(\xi_k, d_k) = \rho^*$ alors d^* est un estimateur minimax de θ .

Dans certains cas, on peut obtenir un estimateur admissible.

Théorème 3. - [30].

1) Si d_0 est l'unique estimateur minimax alors d_0 est admissible.

2) Si d_0 est admissible et a un risque constant, alors d_0 est minimax.

Exemple (2) illustrant le théorème (2).

Problème de l'estimation du paramètre θ d'une loi gaussienne réduite en n observations.

Soit X la variable aléatoire suivant une loi $N(\theta, 1)$.
 $\theta \in \Theta = \mathbb{R}$, θ considéré comme une variable aléatoire suivant une loi a priori $\xi_k : N(0, k^2)$.

L'estimateur bayésien correspondant (pour n observations indépendantes) est d'après [71], (en notant : $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$)

$$d_k(X_1, \dots, X_n) = \frac{\bar{X}_n}{1 + \frac{1}{nk^2}}$$

. Soit alors l'estimateur $\bar{X}_n = d^*(X_1, \dots, X_n)$, dont le risque est :

$$R(d^*, \theta) = \frac{1}{n} \quad \forall \theta \in \Theta.$$

On montre que : $r(\xi_k, d_k) \xrightarrow{k \rightarrow \infty} R(d^*, \theta)$.

L'estimateur d^* est donc un estimateur minimax de θ .

On peut d'ailleurs faire plusieurs remarques à partir de cet exemple.

Remarque 1.- Dans ce cas l'estimateur minimax coïncide avec l'estimateur sans biais.

Remarque 2.- L'estimateur $d^*(X_1, \dots, X_n) = \bar{X}_n$ apparaît comme l'estimateur bayésien généralisé et est formellement un estimateur bayésien avec comme loi a priori la mesure de Lebesgue sur $\Theta = \mathbb{R}$.

Remarque 3.- Soit $d_1(X_1, \dots, X_n) = \frac{\bar{X}_n}{1 + \frac{1}{n}}$ l'estimateur bayésien correspondant à ξ_1 .

On obtient par un calcul immédiat que :

$$R(d_1, \theta) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left(\frac{x_1 + \dots + x_n}{n+1} - \theta \right)^2 dP_\theta(x_1) \dots dP_\theta(x_n) = \frac{n+\theta^2}{(n+1)^2}$$

$$R(d^*, \theta) = \frac{1}{n} \quad d^*(X_1, \dots, X_n) = \bar{X}_n.$$

Alors $R(d_1, \theta) < R(d^*, \theta)$ si $\theta^2 < 2$.

Résultat tout à fait logique, d'après l'hypothèse a priori qui privilégie la région $\theta^2 < 2$.

III - Estimation I minimax.

L'estimation bayésienne correspond à une très bonne information sur Θ ensemble des paramètres. A l'opposé l'estimation minimax ne suppose aucune connaissance sur Θ , ou plutôt ne tient compte d'aucune

information sur θ : en effet, dans la pratique, l'on dispose souvent d'une famille $I \subset H$ de lois a priori possibles. La question est donc de déterminer sous de telles informations l'estimateur optimal.

On pourra consulter dans cette direction : [3], [31], [50], [52], [66].

Définition.- Soit $I \subset H$ une famille de lois a priori sur θ .

Un estimateur d_I est "I minimax" s'il réalise

$$\sup_{\xi \in I} r(\xi, d_I) = \inf_{d \in \mathcal{D}} \sup_{\xi \in I} r(\xi, d).$$

Remarques :

1) Si $I = \{\xi\}$ l'estimateur I minimax est alors l'estimateur bayésien correspondant à ξ .

2) Si $I = H$. On a cette fois l'estimateur minimax.

(Il suffit d'ailleurs que I contiennent toutes les distributions dégénérées pour que l'on obtienne l'estimateur minimax.

Voir [3], [62]).

Exemple.- Reprenons l'étude faite dans l'exemple précédent.

(Estimation du paramètre θ : X suivant une loi $N(\theta, 1)$). On considère comme famille de lois a priori sur $\theta = \mathbb{R}$

$$I = I_\rho = \{\xi_H / \xi_H = \rho \xi_0 + (1-\rho)H\}$$

$$0 < \rho < 1$$

$$\xi_0 = \delta_0 \text{ dirac en } 0$$

H une loi a priori arbitraire $H \in H$.

Il a été vu précédemment que dans le cas d'une observation l'estimateur minimax est :

$$d^*(X) = X \text{ et que } R(d^*, \theta) = 1.$$

Par contre, l'estimateur I minimax (voir [3]) est égal à :

$$d_I(X) = \begin{cases} 0 & \text{si } |X| \leq 1 \\ X & \text{si } |X| > 1 \end{cases}$$

et pour un choix convenable de ρ on obtient que : $R(d_I, \theta) < 1$;
 d_I est donc préférable à d^* tout en conservant son caractère d'estimateur "prudent".

Théorème [52] analogue au théorème 1, I.2.

Soit d_0 un estimateur bayésien pour une loi a priori $\xi^* \in I \subset H$ sur θ et tel que : $R(\theta, d_0) = \text{cte.}$
 $\forall \theta \in \Theta$

Alors d_0 est un estimateur I minimax.

Remarque (1).- Si $P_\theta = \{P_\theta, \theta \in \Theta\}$ est la famille exponentielle à un paramètre et

$$I = \left\{ \xi / \int_\theta \theta d\xi(\theta) = v_1 \text{ et } \int_\theta \theta^2 d\xi(\theta) = v_2 \quad \left. \begin{array}{l} v_1 \\ v_2 \end{array} \right\} \text{ fixés} \right\}$$

[33] Jackson & others traitent la I-admissibilité et l'estimation asymptotiquement I-minimax : Notion proche de l'estimation empirique bayésienne asymptotiquement optimal. (cf. Singh [54]).

Remarque (2).- Dans [34], Katz considère des estimateurs minimax et admissibles dans des espaces tronqués :

$$\theta' = \{\theta \in \Theta / \theta \geq \theta_0\}.$$

On peut interpréter son travail en langage I-minimax, I étant l'ensemble des lois a priori sur θ qui ne chargent que $\theta / \theta \geq \theta_0$.

(B) Estimation globale.

1 - Estimation de la fonction de répartition F.

. On désignera par G_1 l'ensemble des fonctions de répartition continues G_2 l'ensemble des fonctions de répartition

. \hat{F} étant un estimateur de F, on utilisera les fonctions de perte L_i $i \in \{1,2,3,4\}$ définies par

$$L_1(\hat{F}, F) = \int [\hat{F}(x) - F(x)]^2 dF(x)$$

$$L_2(\hat{F}, F) = \int [\hat{F}(x) - F(x)]^2 d\lambda(x), \lambda \text{ étant une mesure finie sur } \mathbb{R}.$$

$$L_3(\hat{F}, F) = \int \frac{[\hat{F}(x) - F(x)]^2}{F(x)(1-F(x))} dF(x)$$

$$L_4(\hat{F}, F) = \int \frac{[F(x) - \hat{F}(x)]^2}{F(x)(1-F(x))} d\lambda(x).$$

. On notera : $R_i = R_i(\hat{F}, F) = E_F[L_i(\hat{F}, F)]$

le risque correspondant

$$r_i = r_i(\xi) = E_\xi[R_i(\hat{F}, F)]$$

le risque bayésien correspondant pour la loi a priori ξ sur G_1 (resp. G_2).

. \hat{F}_i^* l'estimateur minimax de F correspondant à la fonction de perte L_i

. $\|\cdot\|_p$ désignera la norme de l'espace de Banach $L^p(\lambda)$
avec $\|f\|_p = \left[\int f^p(x) d\lambda(x) \right]^{1/p}$

1. Première technique d'obtention de l'estimateur minimax de F.

On considère les fonctions de perte L_1 et L_3 , indépendantes de F à estimer, $F \in G_1$. On obtient alors des risques constants qu'il suffit de minimiser en accord avec la définition de l'estimateur minimax. Citons deux résultats significatifs ([2], [49]).

a. Dans le cas de la fonction de perte L_1 .

\hat{F} désignant ici l'estimateur de F défini par : X_1, \dots, X_n étant un échantillon ordonné tiré suivant X (X_0 et X_{n+1} étant respectivement égaux à $-\infty$ et $+\infty$)

$$\hat{F}(x) = \text{cte} = c_j \quad \text{pour } X_j \leq x < X_{j+1}.$$

On montre ([2]) que le risque R_1 est constant, pour

$$c_j = \frac{j+1}{n+2}, \quad j = 0, \dots, n.$$

d'où l'estimateur minimax \hat{F}_1^* .

Remarque. Si on désigne par F_n la fonction de répartition empirique (estimateur sans biais de F) un calcul simple montre que :

$$R_1(\hat{F}_1^*, F) = \frac{1}{6(n+2)} \quad \text{et} \quad R_1(F_n, F) = \frac{1}{6n}.$$

b. Dans le cas de la fonction de perte L_3 .

Le risque R_3 correspondant est minimisé pour :

$$c_j = \frac{j}{n}, \quad j = 0, \dots, n; \quad \text{d'où l'estimateur minimax } \hat{F}_3^*.$$

Remarque. \hat{F}_3^* coïncide avec F_n et $R_3(F_n, F) = R_3(\hat{F}_3^*, F) = \frac{1}{n}$.

2. Seconde technique d'obtention de l'estimateur minimax de F.

On utilise cette fois la version du théorème (2) II, correspondante à l'estimation fonctionnelle non paramétrique (d'après [47]).

Soient alors les estimateurs \hat{F} de $F \in G_2$ définis par :

$$F(x) = a + \sum_{i=1}^n b_i \delta_{X_i}((-\infty, t]) \quad a \geq 0 \quad b_i \geq 0$$

$$a + \sum_{i=1}^n b_i \leq 1.$$

Théorème d'après Phadia [47].

Soit \hat{F}_k une suite d'estimateurs bayésiens de F pour les lois a priori ξ_k sur G_2 et soit $r(\xi_k, \hat{F}_k)$ la suite correspondante des risques bayésiens.

Si $r(\xi_k, \hat{F}_k) \xrightarrow{k \rightarrow \infty} C$ et si \hat{F}^* est un estimateur de F tel que $\forall F \in G_2, R(\hat{F}^*, F) \leq C$ alors

\hat{F}^* est un estimateur minimax de F .

La difficulté consiste alors à déterminer des suites de lois a priori sur G_2 .

a) Premier exemple de détermination d'une suite de lois a priori sur G_2 d'après [47].

le $k^{\text{ième}}$ membre de cette suite (ξ_k) est défini de la façon suivante

$$\xi_k\{F/F(t) = 0 \text{ si } t < -k\} = 1$$

$$\xi_k\{F/F(t) = 1 \text{ si } t \geq k\} = 1$$

$$\xi_k\{F/F(t) \leq C \text{ si } -k \leq t < k\} = \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \int_0^C x^{a-1} (1-x)^{b-1} 1_{[0,1]}(x) dx$$

C'est-à-dire pour $-k \leq t < k$. $F(t)$ est distribuée suivant une loi bêta de paramètres a et b .

Ces paramètres a et b dépendent de n , taille de l'échantillon). On peut donc dire que ξ_k est essentiellement concentrée sur le sous-espace des distributions qui ont seulement des sauts en $+k$ et $-k$.

● Dans le cas de la fonction de perte L_2 .

Pour $a = b = \frac{\sqrt{n}}{2}$ (on obtient ainsi une suite croissante de risques bayésiens). On a alors l'estimateur bayésien pour ξ_k défini par

$$\left| \begin{array}{ll} \hat{F}_k(t) = \left[2(n^{\frac{1}{2}} + 1) \right]^{-1} + \frac{n F_n(t)}{n+n^{1/2}} & \text{si } -k \leq t < k \\ \hat{F}_k(t) = 0 & \text{si } t < -k \\ \hat{F}_k(t) = 1 & \text{si } t \geq k. \end{array} \right.$$

Cet estimateur bayésien est tel que

$$r_2^k = r_2(\xi_k, \hat{F}_k) = \int_{[-k, +k]} 1(t) \left\{ \frac{1}{4(n^{1/2}+1)^2} \right\} d\lambda(t)$$

$$r_2^k \rightarrow R_2(\hat{F}_2^*, F) = \int \frac{1}{4(n^{1/2}+1)^2} d\lambda(t) \text{ avec}$$

$$\hat{F}_2^*(t) = \left[2(n^{\frac{1}{2}} + 1) \right]^{-1} + \frac{n}{n+n^{1/2}} F_n(t), \text{ qui est alors l'estimateur$$

minimax de F pour L_2 .

● Dans le cas de la fonction de perte L_4 .

(Pour $a = b = 1$). Un raisonnement analogue nous donne l'estimateur minimax de F pour L_4 égal à :

$$\hat{F}_4^*(t) = F_n(t).$$

b) Deuxième exemple de détermination d'une suite de lois a priori sur G_2 d'après [24], [25], [44].

Au lieu de mettre directement une loi a priori sur G_2 , Fergusson [24] considère une loi a priori P sur $(P_{(R)}, B_{P(R)})$ où $P_{(R)}$ est l'ensemble des probabilités sur (R, B_R) , muni de la

topologie de la convergence faible, $\mathcal{B}_{\mathcal{P}(\mathbb{R})}$ sa tribu borélienne.

et où P est la mesure de probabilité aléatoire correspondante

à \mathcal{P} telle que $\int x dP(x)$ et $P(B) \forall B \in \mathcal{B}_{\mathbb{R}}$ sont des variables aléatoires.

Remarque : G_2 et $\mathcal{P}_{\mathbb{R}}$ sont isomorphes.

Le problème consiste à déterminer P telle que

1 - Le support à \mathcal{P} (pour la topologie de la convergence faible) est $\mathcal{P}(\mathbb{R})$.

2 - La distribution a posteriori de P (étant donné X_1, \dots, X_n échantillon tiré suivant P) doit être facilement calculable.

Pour cela, il est nécessaire de rappeler les définitions de la distribution de Dirichlet et d'un processus de Dirichlet.

Définition 1.- La distribution de Dirichlet de dimension m et de paramètres $(\alpha_1, \dots, \alpha_m)$ $\alpha_i \geq 0$ et $\sum_{i=1}^m \alpha_i > 0$, est la distribution de $(Z_1/S, Z_2/S, \dots, Z_m/S)$, où

. Z_1, \dots, Z_m sont indépendantes, Z_i suivant une loi gamma : $g(\alpha_i, 1)$. $i = 1, \dots, m$, et où

$$. S = \sum_{i=1}^m Z_i.$$

Définition 2.- Soient $\alpha(\cdot)$ une mesure finie σ additive et non nulle sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et $P(\cdot)$ un processus stochastique indexé par les éléments de $\mathcal{B}_{\mathbb{R}}$. On dira que P est un processus de Dirichlet de paramètre α (notation : $P \in \mathcal{D}(\alpha)$), si pour toute partition mesurable et finie de \mathbb{R} :

$\{B_1, \dots, B_m\}$ $B_i \in \mathcal{B}_{\mathbb{R}}$, $i=1, \dots, m$, le vecteur aléatoire $(P(B_1), \dots, P(B_m))$ suit une distribution de Dirichlet de paramètres $(\alpha(B_1), \dots, \alpha(B_m))$.

Fergusson [24] montre que l'on peut appliquer le théorème de Kolmogorov (voir [7] par exemple). On en déduit alors l'existence d'une probabilité P sur $(\mathcal{P}(\mathbb{R}), \mathcal{B}_{\mathcal{P}(\mathbb{R})})$ loi d'un processus de Dirichlet $\mathcal{D}(\alpha)$: c'est notre loi a priori qui répond bien aux deux exigences précédentes car :

1. Le support de P est l'ensemble des probabilités sur \mathbb{R} dont le support est dans celui de α . (On prendra : support de α égal à \mathbb{R}).

2. La loi a posteriori de $P \in \mathcal{D}(\alpha)$ est la loi d'un processus $P' \in \mathcal{D}(\alpha + \sum_{i=1}^n \delta_{X_i})$.

Dans le cas de la fonction de perte L_2 , on peut obtenir un estimateur bayésien de la fonction de répartition.

En effet, soit $F(t) = P((-\infty, t])$. Alors si $P \in \mathcal{D}(\alpha)$
 $\implies F(t)$ suit une loi $\mathcal{B}_e[\alpha((-\infty, t]), \alpha(]t, \infty))$.

L'estimateur bayésien de $F \in G_2$ si $P \in \mathcal{D}(\alpha)$ sera égal à

$$\left\{ \begin{array}{l} \hat{F}_\alpha(t) = E_P[\bar{F}(t)/X_1 \dots X_n] = p_n F_0(t) + (1-p_n)F_n(t) \\ F_0(t) = \alpha((-\infty, t]) \\ p_n = \frac{\alpha(\mathbb{R})}{\alpha(\mathbb{R})+n} \end{array} \right.$$

. On peut maintenant définir l'estimateur minimax de F :

En effet dans le cas où $\alpha(\cdot) = \alpha_k(\cdot) = \frac{\sqrt{n}}{2} \delta_{-k}(\cdot) + \frac{\sqrt{n}}{2} \delta_k(\cdot)$,

l'estimateur \hat{F}_α coïncidera avec \hat{F}_k obtenu en a . pour $-k \leq t < k$.

D'autre part, si l'on considère une suite (α_k) de mesures, α_k définie comme ci-dessus, on obtient une suite P_k de lois a priori sur $(\mathcal{P}(\mathbb{R}), \mathcal{B}_{\mathcal{P}(\mathbb{R})})$ et comme il y a équivalence entre :

$F(t)$ suit une loi $B_e(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$ - $k \leq t < k$ hypothèse utilisée en a.

et

$F(t)$ suit une loi $B_e(\alpha((-\infty, t]), \alpha([t, \infty)))$ - $k \leq t < k$ hypothèse utilisée en b.

avec $\alpha(\cdot) = \alpha_k(\cdot) = \frac{\sqrt{n}}{2} \delta_{-k}(\cdot) + \frac{\sqrt{n}}{2} \delta_k(\cdot)$.

Le calcul du risque bayésien r_2^k est donc identique à celui effectué en a.

$$r_2^k = E_P \left\{ \int [\hat{F}_{\alpha_k}(t) - F(t)]^2 dP_F(x_1) \dots dP_F(x_n) d\lambda(t) \right\}$$

$$r_2^k \xrightarrow[k \rightarrow \infty]{} R_2(\hat{F}_2^*, F) \quad \text{avec}$$

$$\hat{F}_2^*(t) = [2(n^{1/2} + 1)]^{-1} + \frac{n}{n+n^{1/2}} F_n(t) \quad \text{Estimateur minimax de } F \text{ pour } L_2.$$

Remarque (complétant les résultats du A).

Si $P \in \mathcal{D}(\alpha)$, on peut utiliser les processus de Dirichlet pour définir les estimateurs bayésiens de $m = \int x dP(x)$ dans le cas de la fonction de perte quadratique. On obtient dans [24] que l'estimateur bayésien \hat{m} de m est alors défini par :

$$\left| \begin{aligned} \hat{m}(X_1 \dots X_n) &= P_n m_0 + (1 - P_n) \bar{X}_n \\ P_n &= \frac{\alpha(R)}{\alpha(R) + n} \\ m_0 &= \int x d\alpha(x) / \alpha(R). \end{aligned} \right.$$

Cet estimateur coïncide avec celui obtenu

1. Par [31] exemple (1), chap. I, A, II

dans le cas où $\alpha(R) = \frac{1}{\sqrt{m'}}$ et $E(\alpha) = \frac{\sqrt{m'}}{2}$, $n = 1$.

(Il suffit de multiplier l'estimateur de θ par m' pour obtenir un estimateur de la moyenne).

2. Par [71] exemple (2), chap. I, A, II

dans le cas où $\alpha_k(\mathbb{R}) = \frac{1}{k^2}$ et $E(\alpha_k) = 0$.

Alors si on considère la suite (α_k) qui est telle que :
 $\alpha_k(\mathbb{R}) \rightarrow 0$, on retrouve bien le résultat que \bar{X}_n est un estimateur
 $k \rightarrow \infty$
minimax pour la fonction de perte quadratique.

II - Estimation de fonctions de la densité f .

. f désigne une version spécifiée de la densité.

. On désigne maintenant par $P_F = \{P_f, f \in F\}$ la famille de probabilités sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ dominée par μ , avec $f = \frac{dP_f}{d\mu}$, et μ mesure σ finie.

1. Existence de l'estimateur minimax de $\Psi(f)$,

a) Ψ définie sur F à valeurs dans A .

On désigne par A l'ensemble des fonctions mesurables définies et à valeurs dans \mathbb{R} .

. La fonction de perte utilisée sera L_2 (λ étant σ finie et non plus finie comme dans le I).

. h étant un estimateur mesurable

$$\text{Soit } N_2(h, F, \lambda) = \left[\sup_{f \in F} \int h^2 dP_f^n \otimes \lambda(x^n, x) \right]^{1/2}$$

où $x^n = (x_1, \dots, x_n)$.

. On identifiera les fonctions h à leurs classes d'équivalence pour la relation :

$$h_1 \sim h_2 \text{ si } h_1 = h_2 P_f^n \otimes \lambda_P, \forall f \in F.$$

. On notera alors : $\Omega_2(F, \lambda) = \{h/N_2(h, F, \lambda) < \infty\}$,
ou Ω_2 s'il n'y a pas de confusion possible.

La structure de $\Omega_2(F, \lambda)$ est précisée dans le lemme suivant :

Lemme 1. -

$\Omega_2(F, \lambda)$ est un espace de Banach pour la norme $N_2(h, F, \lambda)$.

Démonstration : Voir [48].

. On suppose que $B_{\mathbb{R}^n} \otimes B_{\mathbb{R}}$ est fermée par rapport à $P_F \otimes \lambda = \{P_{f_i}^n \otimes \lambda, f_i \in F\}$, c'est-à-dire qu'elle contient les ensembles $P_{f_i} \otimes \lambda$ nuls.

. Si $M_2 = M_2(F, \lambda)$ désigne l'ensemble des combinaisons linéaires finies $\sum_{i=1}^k \alpha_i g_i$,

$$[k \in \mathbb{N}, \alpha_i \in \mathbb{R}, g_i \in L^2(\mathbb{R}^n \times \mathbb{R}, B_{\mathbb{R}^n} \otimes B_{\mathbb{R}}, P_{f_i}^n \otimes \lambda)].$$

$i=1 \dots k \quad f_i \in F$

Alors, on obtient une forme bilinéaire sur $\Omega_2 \times M_2$ en posant : $\langle h, \sum_{i=1}^k \alpha_i g_i \rangle = \sum_{i=1}^k \alpha_i \int h \cdot g_i \, d P_{f_i}^n \otimes \lambda$.

On appelle $\sigma(F, \lambda)$ la topologie la moins fine qui rende continues les applications linéaires $\langle \cdot, g \rangle$, $g \in M_2$.

Lemme 2. - [68]

La $\sigma(F, \lambda)$ topologie est moins fine que la topologie faible de $\Omega_2(F, \lambda)$.

Démonstration : $M_2(F, \lambda)$ est inclus dans $\Omega_2^*(F, \lambda)$ d'où le résultat. (cf. Bourbaki livre 5, Corollaire 3, page 92).

Afin de caractériser les compacts de cette $\sigma(F, \lambda)$ topologie, il est nécessaire de définir la notion d'espace probabilisé compact (notion due à Pitcher [48]).

Définition. - Si B_2 désigne la boule unité fermée de $\Omega_2(F, \lambda)$, l'espace $(\mathbb{R}^n \times \mathbb{R}, \mathcal{B}_{\mathbb{R}^n} \otimes \mathcal{B}_{\mathbb{R}}, P_F \otimes \lambda = \{P_F^n \otimes \lambda, f \in F\})$ sera compact si et seulement si B_2 est compacte pour la $\sigma(F, \lambda)$ topologie.

Théorème (1). - [48]

Soit ν une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Si l'ensemble des mesures de probabilités $P_F \otimes \nu$ est dominé, alors l'espace $(\mathbb{R}^n \times \mathbb{R}, \mathcal{B}_{\mathbb{R}^n} \otimes \mathcal{B}_{\mathbb{R}}, P_F \otimes \nu)$ sera compact.

. La notion de compacité étant plus générale que la notion de domination, il existe des situations où la compacité n'entraîne pas la domination (Voir [48], page 600).

. Avant de citer des résultats d'existences d'estimateurs minimax de $\Psi(f)$, Ψ fonction définie sur F , nous aurons besoin de quelques lemmes préparatoires.

Lemme 3. d'après [68].

Il existe ν une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,

ψ une fonction telle que $\psi > 0$ (λ .ps).

ψ et ν étant telles que l'on ait l'équivalence

$$h \in \Omega_2(F, \lambda) \iff h \psi^{-1/2} \in \Omega_2(F, \nu).$$

Démonstration : Comme λ est σ finie, on peut trouver une mesure de probabilité ν équivalente à λ et telle que :

$$\psi = \frac{d\nu}{d\lambda} > 0 \quad (\lambda\text{-ps}). \quad \text{On peut considérer que :}$$

$$\psi(x) > 0 \quad \text{et} \quad \psi = \frac{d(P_F^n \otimes \nu)}{d(P_F^n \otimes \lambda)} \quad \text{alors on a bien}$$

$$\forall x \in \mathbb{R}$$

$$h \in \Omega_2(F, \lambda) \iff h \psi^{-1/2} \in \Omega_2(F, \nu).$$

. On utilisera enfin l'expression :

$$\begin{aligned} \delta(h, \lambda) &= \sup_{f \in F} N_2(h - \Psi(f), \{f\}, \lambda) \\ &= \sup_{f \in F} \left[\int \left[h(x^n, x) - \Psi(f)(x) \right]^2 dP_f^n \otimes \lambda(x^n, x) \right]^{1/2} \end{aligned}$$

expression qu'il va falloir minimiser afin de déterminer l'estimateur minimax de $\Psi(f)$ pour L_2 .

On peut alors citer les deux théorèmes suivants :

Théorème (2) (le cas de $\Psi = \text{Id}$ a été traité par Wertz [68]).

Soit Ψ la fonction définie sur F telle que :

$$\sup_{f \in F} \int \Psi^2(f)(x) d\lambda(x) < \infty.$$

Alors il existe dans $\Omega_2(F, \lambda)$ un estimateur minimax de $\Psi(f)$ pour L_2 .

C'est-à-dire que pour toute fonction $\Psi : F \rightarrow L^2(\lambda)$, il existe un estimateur minimax de $\Psi(f)$.

Démonstration :

Soit ν la mesure de probabilité définie au lemme 3. Soit alors

$$\delta'(h, \nu) = \sup_{f \in F} N_2(h - \Psi(f) \cdot \varphi^{-1/2}, \{f\}, \nu).$$

L'ensemble des mesures de probabilité $P_F \otimes \nu$ est alors dominé et d'après le théorème (1) et le lemme 2, les boules fermées de $\Omega_2(F, \nu)$ sont compactes pour la $\sigma(F, \nu)$ topologie.

Soit $\alpha = \inf_{h \in \Omega_2(F, \nu)} \delta'(h, \nu)$ si ω est la fonction nulle

$$0 \leq \alpha \leq \delta'(\omega, \nu) = \left[\sup_{f \in F} \int (\Psi(f) \cdot \varphi^{-1/2})^2 dP_f^n \otimes \nu \right]^{1/2} = \left[\sup_{f \in F} \int \Psi^2(f) d\lambda < \infty \right]^{1/2}.$$

Alors il existe (h_k) incluse dans une boule fermée telle que :

$$\alpha = \lim_{k \rightarrow \infty} \delta'(h_k, \nu)$$

(h_k) possède donc un point d'accumulation h' pour la $\sigma(F, \nu)$ topologie. Tout voisinage de la forme : $\{h \in \Omega.(F, \nu) \mid |h - h', g| \leq \varepsilon\}$ avec $\varepsilon > 0$ et $g \in L_2(P_f^n \otimes \nu)$ contient une sous-suite (h_{k_j}) de (h_k) . Alors $\forall g \exists (h_{k_j})$ avec $\lim_{j \rightarrow \infty} |h_{k_j} - h', g| = 0$ d'où

$$\begin{aligned} |h' - \Psi(f) \cdot \psi^{-1/2}, g| &= \lim_{j \rightarrow \infty} |h_{k_j} - \Psi(f) \cdot \psi^{-1/2}, g| \\ &\leq \limsup_{j \rightarrow \infty} N(h_{k_j} - \Psi(f) \cdot \psi^{-1/2}, \{f\}, \nu) \cdot N(g, \{f\}, \nu) \\ &\leq \alpha \cdot N(g, \{f\}, \nu). \end{aligned}$$

D'autre part, considérant $h' - \Psi(f) \cdot \psi^{-1/2}$ comme un élément de $L_2(P_f^n \otimes \nu)$, on peut l'interpréter comme un élément du bidual $L_2^{**}(P_f^n \otimes \nu)$ et on a : $\forall f \in F$

$$N(h' - \Psi(f) \cdot \psi^{-1/2}, \{f\}, \nu) = \sup_{g \in L_2 : N(g) < 1} |h' - \Psi(f) \cdot \psi^{-1/2}, g| \leq \alpha.$$

Alors $\delta'(h', \nu) = \alpha$. Comme

$$\begin{aligned} \delta(h' \cdot \psi^{1/2}, \lambda) &= \left[\sup_{f \in F} \int [h' \cdot \psi^{1/2} - \Psi(f)]^2 dP_f^n \otimes \lambda \right]^{1/2} \\ &= \left[\sup_{f \in F} \int (h' - \Psi(f) \cdot \psi^{-1/2}) dP_f^n \otimes \nu \right]^{1/2} = \alpha. \end{aligned}$$

On pose : $h^* = h' \cdot \psi^{1/2}$ estimateur minimax de $\Psi(f)$ pour L_2

car :

si $\exists h_1$ optimal dans $\Omega_2(F, \lambda)$ alors

$$\delta(h_1, \lambda) < \delta(h^*, \lambda) = \alpha \text{ et on aurait}$$

$$\delta'(h_1 \psi^{-1/2}, \nu) < \alpha, \text{ ce qui est impossible.}$$

Corollaire 1. - Si λ est finie et si $\Psi(f)$ est une fonction bornée pour tout f élément de F , alors il existe un estimateur minimax de $\Psi(f)$ pour L_2 .

Démonstration : Il est évident que la condition :

$$\sup_{f \in F} \int \Psi^2(f)(x) d\lambda(x) < \infty \text{ est alors vérifiée.}$$

Exemple 1.- Soit Ψ telle que $\Psi(f) = F$, $F(x) = \int_{-\infty}^x f(t) dt$.

Il existe bien un estimateur minimax de F pour L_2 .

L'estimateur a d'ailleurs été calculé dans [24], [25], [47], dans le cas où Ψ est surjective, les auteurs n'ayant cependant pas justifié le choix de l'hypothèse sur λ .

Exemple 2.- $\Psi(f) = af + b$ possèdera un estimateur minimax si a et b sont finis.

Corollaire 2.- Soit (X, Y) un couple de variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) de densité $f(x, y)$ par rapport à $\mu \otimes \mu$ et $\Psi(f)(x) = \int y f(y/x) d\mu(y)$ la fonction de régression avec $f(y/x) = \frac{f(x, y)}{\int f(x, y) d\mu(y)}$ la densité conditionnelle.

Alors si

- 1) μ admet un moment d'ordre 2 fini
- 2) $\sup_{f \in F} \int f^2(y/x) d\mu(y) < \infty \quad \forall x \quad \lambda(p.s)$
- 3) $\sup_{f \in F} \iint f^2(y/x) d\mu(y) d\lambda(x) < \infty$

il existe un estimateur minimax de la fonction de régression pour L_2 dans $\Omega_2(F, \lambda)$.

Démonstration :

Avec les hypothèses ci-dessus la condition

$$\sup_{f \in F} \int \Psi^2(f)(x) d\lambda(x) < \infty \text{ est bien vérifiée.}$$

En effet, on a alors :

$$\begin{aligned} \iint \left[\int y f(x/y) d\mu(y) \right]^2 d\lambda(x) &\leq \iint \left[\int y^2 d\mu(y) \cdot \int f^2(y/x) d\mu(y) \right] d\lambda(x) \\ &\leq \int y^2 d\mu(y) \cdot \iint f^2(y/x) d\mu(y) d\lambda(x) < \infty. \end{aligned}$$

b) Ψ définie sur F et à valeurs dans \mathbb{R} .

On peut citer le théorème suivant analogue au théorème (2) précédent.

Théorème. -

Soit Ψ la fonction définie sur F et à valeurs réelles $\Psi : F \rightarrow \mathbb{R}$. Alors si Ψ est bornée sur F , il existe un estimateur minimax de $\Psi(f)$ pour la perte quadratique.

Démonstration : On considère cette fois la norme :

$$N_2(h, F) = \left[\sup_{f \in F} \int h^2(x^n) dP_f^n(x^n) \right]^{1/2} \quad \text{qui fait de}$$

$\Omega_2(F) = \{h/N_2(h, F) < \infty\}$ un espace de Banach (analogue au lemme 1 précédent).

Soit la $\sigma(F)$ topologie définie par la forme bilinéaire

$$\langle h, \sum_{i=1}^k \alpha_i g_i \rangle = \sum_{i=1}^k \alpha_i \int h g_i dP_{f_i}^n$$

$$\begin{aligned} k \in \mathbb{N} \quad \alpha_i \in \mathbb{R} \quad g_i \in L_2(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, P_{f_i}^n) \\ h \in N_2(h, F) \end{aligned}$$

alors la boule unité de $\Omega_2(F)$ est compacte pour la $\sigma(F)$ topologie.

La démonstration est alors identique à celle du théorème (2).

Corollaire 1. - Si λ est une mesure σ finie telle que :

$$\sup_{f \in F} \int f^2 d\lambda < \infty \quad \text{alors}$$

$\Psi(f) = \int f^2 d\lambda$ admet un estimateur minimax pour la perte quadratique.

Corollaire 2. - Si :

$$\left| \begin{array}{l} \sup_{f \in F} \int f^2 d\lambda < \infty \\ \sup_{f \in F} \int (\text{Log } f)^2 d\lambda < \infty. \end{array} \right.$$

Alors, il existe un estimateur minimax de l'entropie

$$\Psi(f) = \int f(y) \text{Log } f(y) dy.$$

2. Situation dans laquelle l'estimateur minimax de

$$\Psi(f) = f \text{ est unique } f \in F_{\Theta}.$$

. Dans le cas d'une famille paramétrée et pour la fonction de perte $L_2(\lambda)$, on peut, avec des hypothèses supplémentaires, obtenir l'unicité et donc l'admissibilité de l'estimateur minimax.

. L'espace probabilisé est ici $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathcal{P}_{\Theta}, \Theta \subset \mathbb{R})$. Les mesures λ et μ sont égales à la mesure de Lebesgue sur \mathbb{R} .

$$\text{On notera } f(\theta, x) = \frac{dP_{\theta}}{dx} \quad P_{\theta} \in \mathcal{P}_{\Theta}$$

- . $F_{\Theta} = \{f(\theta, x), \theta \in \Theta\}$
- . Ω_2 au lieu de $\Omega_2(F_{\Theta}, \lambda)$
- . $N_2(h)$ au lieu de $N_2(h, F_{\Theta})$
- . $\Omega_{2, K} = \{h \in \Omega_2 / N_2(h) \leq K\}$.

On pose les hypothèses supplémentaires suivantes : d'après [26]

- H.1. $F_{\Theta} \subset L^2(\mathbb{R})$ et $\int f^2(x, \theta) dx$ uniformément bornée sur Θ
- H.2. $\forall \theta \in \Theta \quad f(\theta, x) > 0$ (sauf sur un ensemble de mesure nulle indépendant de θ).
- H.3. $\exists \theta_0 \in \Theta / \forall M \subset \mathbb{R} \quad M$ ensemble borné $\exists A_M$ (cte) vérifiant $\frac{1}{A_M} f(x, \theta_0) < f(x, \theta) < A_M f(x, \theta_0)$
 $\forall \theta \in \Theta \quad ; \quad \forall x \in M$ (sauf sur un ensemble de mesure nulle).
- H.4. Θ compact.
- H.5. $f(\theta, x)$ fonction continue de θ pour tout x , (sauf peut-être sur un ensemble de x de mesure nulle).

H.6. Une suite (h_i) d'estimateurs de $f(\theta, \cdot)$ satisfera la condition H.6. Si :

$$\forall \varepsilon > 0 \quad \exists M(\varepsilon) \subset \mathbb{R} \quad (M(\varepsilon) \text{ région bornée}) \text{ telle que}$$

$$\int_{[\mathbb{R}-M(\varepsilon)]^n} \int_{\mathbb{R}} [h_i(x^n, x) - f(\theta, x)]^2 dP_{\theta}^n(x^n) dx < \varepsilon \quad \begin{matrix} \forall \theta \in \Theta \\ \forall \theta \in N \end{matrix}.$$

Nous allons tout d'abord préciser dans quel cas l'hypothèse H.6. est vérifiée pour une suite (h_i) d'estimateurs de $f(\theta, \cdot)$. Ce travail est une extension de Gosh [26], sur l'estimation d'un paramètre réel.

Lemme. -

Soit $(h_i)_{i \in \mathbb{N}}$ une suite incluse dans Ω_2 . Alors si les hypothèses H.1. à H.5. sont vérifiées, la suite $(h_i)_{i \in \mathbb{N}}$ vérifie H.6.

Démonstration :

$\forall \theta \in \Theta$ on choisit $M_{\theta} \subset \mathbb{R}$ telle que $\int_{\mathbb{R}-M_{\theta}} f(\theta, x) dx < \varepsilon'$.
Comme $\int_{M_{\theta}} f(\theta, x) dx$ est une fonction continue par rapport à θ

$$\forall \theta' \in v(\theta) \quad \left| \int_{M_{\theta}} f(x, \theta) dx - \int_{M_{\theta}} f(x, \theta') dx \right| < \varepsilon'$$

[$v(\theta)$ voisinage de θ]

et f étant une densité, on a alors

$$\int_{\mathbb{R}-M_{\theta}} f(x, \theta') dx < 2\varepsilon' \quad \forall \theta' \in v(\theta)$$

d'autre part $L_2(h_i(x^n, x), f(\theta, x))$ est bornée $\left| \begin{matrix} \forall x^n \in \mathbb{R}^n \\ \forall \theta \in \Theta \end{matrix} \right.$ en effet

$$\int [h_i(x^n, x) - f(\theta, x)]^2 dx \leq \int h_i^2(x^n, x) dx - 2 \int h_i(x^n, x) f(\theta, x) dx + \int f^2(\theta, x) dx$$

$$\leq \int \quad \quad \quad + 2 \sqrt{\int h_i^2(x^n, x) dx} \sqrt{\int f^2(\theta, x) dx} + \quad \quad \quad "$$

car $h_i \in \Omega_2$ et l'hypothèse H.1. est vérifiée.

$$\int_{[\mathbb{R}-M_\theta]^n} L_2[h_i(x^n, x), f(\theta, x)] f_\theta(x_1) \dots f_\theta(x_n) dx_1 \dots dx_n dx$$

$$\leq \sup_{x^n, \theta} L_2[h_i(x^n, x), f(\theta, x)] \times [2\varepsilon']^n = \varepsilon.$$

Comme θ est compact, on peut le recouvrir par un nombre fini de voisinages $v(\theta_1) \dots v(\theta_k)$ et $M_{\theta_1} \dots M_{\theta_k}$ étant les sous-ensembles correspondant à $\theta_1 \dots \theta_k$. On a alors :

$$\int_{\mathbb{R}} \int_{[\mathbb{R}-(M_{\theta_1} \cup \dots \cup M_{\theta_k})]^n} [h_i(x^n, x) - f_\theta(x)]^2 dP^n(x^n) dx < \varepsilon \quad \forall \theta \in \theta$$

$$\forall i \in \mathbb{N}$$

Théorème.- Avec les hypothèses H.1. à H.5. il existe un unique estimateur minimax h^* de $f(\theta, \cdot) \in F_\theta$ dans l'ensemble Ω_2 .

Démonstration :

Soient h_1^* et h_2^* deux estimateurs minimax dans Ω_2 .

La fonction de L_2 étant une fonction strictement convexe par rapport à h , on a pour θ fixé

$$E_{P_\theta} \left\{ \int \left[\frac{1}{2} [(h_1^*(x^n, x) + h_2^*(x^n, x))] - f(\theta, x) \right]^2 dx / \theta \right\} <$$

$$\frac{1}{2} \left[E \left\{ \int [h_1^*(x^n, x) - f(\theta, x)]^2 dx / \theta \right\} + E \left\{ \int [h_2^*(x^n, x) - f(\theta, x)]^2 dx / \theta \right\} \right]$$

sauf si $h_1^* = h_2^*$ λ^{n+1} (p.s) λ mesure de Lebesgue.

D'après le lemme précédent :

\exists $M \subset \mathbb{R}$ telle que

$$\int_{[\mathbb{R}-M]^n} \int_{\mathbb{R}} [h(x^n, x) - f(\theta, x)]^2 dP_\theta^n(x^n) dx < \varepsilon \quad \forall \theta \in \theta$$

$$\forall h \in \Omega_2.$$

Comme $f(\theta, x)$ est une fonction continue de θ (sauf sur un ensemble de mesure nulle) et comme $h \in \Omega_2$

$\int_{M^n} \cdot \int_{\mathbb{R}} [h(x^n, x) - f(\theta, x)]^2 dP_\theta^n(x^n) dx$ est une fonction continue de θ et on obtient que : $E\left\{\int [h(x^n, x) - f(\theta, x)]^2 dx / \theta\right\}$ est une fonction continue de θ .

Comme θ est compact, $\exists \theta_0 \in \theta /$

$$\begin{aligned} & \sup_{\theta} E\left\{\int \left[\frac{1}{2} [h_1^*(x^n, x) + h_2^*(x^n, x)] - f(\theta, x)\right]^2 dx / \theta\right\} = \\ & = E\left\{\int \left[\frac{1}{2} [h_1^*(x^n, x) + h_2^*(x^n, x)] - f(\theta_0, x)\right]^2 dx / \theta_0\right\} < \\ & < \frac{1}{2} \{E\left[\int [h_1^*(x^n, x) - f(\theta_0, x)]^2 dx / \theta_0\right] + E\left[\int [h_2^*(x^n, x) - f(\theta_0, x)]^2 dx / \theta_0\right]\} \\ & < \frac{1}{2} \left\{\sup_{\theta} E\left[\int [h_1^*(x^n, x) - f(\theta, x)]^2 dx / \theta\right] + \sup_{\theta} E\left[\int [h_2^*(x^n, x) - f(\theta, x)]^2 dx / \theta\right]\right\}. \end{aligned}$$

L'inégalité précédente est en contradiction avec le fait que h_1^* et h_2^* sont des estimateurs minimax

sauf si : $h_1^* = h_2^* \quad \lambda^{n+1}$ (p.s) et comme Ω_2 est un ensemble quotient (voir II.1.a) alors $h_1^* = h_2^*$.

3. Compléments sur la continuité des applications R_2 et δ_2 pour la distance de l'espace Ω_2 (H.1 à H.5 vérifiées).

On a vu (lemme 1.a.) que Ω_2 est un espace de Banach :

Soit ρ la distance déduite de la norme N_2 .

Soit $\|h\|_{\theta} = \sqrt{\int h^2(x^n, x) dP_{\theta}^n \otimes \lambda(x^n, x)}$.

On note $L_{P_{\theta}^n \otimes \lambda}^2$ l'espace de Hilbert des fonctions h telles que : $\|h\|_{\theta} < \infty$. λ mesure de Lebesgue

Il est immédiat que :

$$\forall \theta \in \theta \quad \Omega_2 \subset L_{P_{\theta}^n \otimes \lambda}^2.$$

Lemme 1.-

Soient $h_1 \in \Omega_{2,K}$ $h_2 \in \Omega_{2,K}$ Alors $\frac{1}{2} (h_1 + h_2) \in \Omega_{2,K}$.

Démonstration :

Rappel : $\Omega_{2,K} = \{h \in \Omega_2 / N_2(h) \leq K\}$.

Soient $h_1 \in \Omega_{2,K}$ et $h_2 \in \Omega_{2,K}$.

Puisque N_2 est une norme, on obtient que :

$$N_2^2 \left[\frac{1}{2} (h_1+h_2) \right] + N_2^2 \left[\frac{1}{2} (h_1-h_2) \right] \leq \frac{1}{2} \left[N_2^2(h_1) + N_2^2(h_2) \right] \quad \text{et}$$

$$N_2^2 \left[\frac{1}{2} (h_1+h_2) \right] \leq \frac{1}{2} \left[N_2^2(h_1) + N_2^2(h_2) \right]$$

$$\text{alors si : } \left. \begin{array}{l} N_2^2(h_1) \leq K^2 \\ N_2^2(h_2) \leq K^2 \end{array} \right| \implies N_2^2\left(\frac{1}{2} (h_1+h_2)\right) \leq K^2$$

d'où le résultat.

Lemme 2.-

$\forall \theta \in \Theta$ si $h(x^n, x) \in \Omega_2$, alors :

$$h(x^n, x) - f(\theta, x) \in L^2_{P_{\theta}^{n, \lambda}}$$

Démonstration : Elle est immédiate d'après l'hypothèse H.1.

et l'application de l'inégalité de Schwartz.

Proposition 1.-

Soit $\theta \in \Theta$, θ fixé, alors $R_2[h(x^n, x), f(\theta, x)]$ est une fonction continue (resp. uniformément continue) pour la distance ρ , $h \in \Omega_2$ (resp. $h \in \Omega_{2,K}$).

Démonstration :

Soient $u_1 \in L^2_{P_{\theta}^{n, \lambda}}$
 $u_2 \in L^2_{P_{\theta}^{n, \lambda}}$ alors : $\left| \|u_1\|_{\theta} - \|u_2\|_{\theta} \right| \leq \|u_1 - u_2\|_{\theta}$.

Soient alors $h_1 \in \Omega_2$ comme : $\left\| h_1(x^n, x) - f(\theta, x) \right\|_{\theta}$
 $h_2 \in \Omega_2$ $= R_2^{1/2}(h_1(x^n, x), f(\theta, x))$

on a d'après le lemme 2 et l'inégalité du dessus :

$$\begin{aligned} & |R_2^{1/2} [h_1(x^n, x), f(\theta, x)] - R_2^{1/2} [h_2(x^n, x), f(\theta, x)]| \\ & \leq \|h_1 - h_2\|_{\theta} \leq \rho(h_1, h_2) \end{aligned}$$

alors $R_2^{1/2} [h(x^n, x), f(\theta, x)]$ est une fonction uniformément continue pour ρ avec $h \in \Omega_2$.

L'application : $y \rightarrow y^2$ est continue (resp. uniformément continue) si $y > 0$ (resp. $0 < y < M < \infty$).

On en déduit alors le résultat.

Proposition 2. - L'application $\delta_2(h)$ est une fonction uniformément continue de h élément de $\Omega_{2,K}$ pour la distance ρ .

Démonstration : Nous allons utiliser la fonction

$$\delta_2^2(h) = \sup_{\theta \in \Theta} R_2(h(x^n, x), f(\theta, x)).$$

Soient : $h_1 \in \Omega_{2,K}$. On peut trouver $\theta_i \in \Theta$ tel que $\forall \theta \in \Theta$:

$$h_2 \in \Omega_{2,K} \quad i=1,2$$

$$R_2(h_i(x^n, x), f(\theta, x)) < \delta_2^2(h_i) < R_2(h_i(x^n, x), f(\theta_i, x)) + \varepsilon .$$

$$i=1,2$$

D'après la proposition 1., si $\rho(h_1, h_2) < \eta$, alors :

$$|R_2(h_1(x^n, x), f(x, \theta_i)) - R_2(h_2(x^n, x), f(x, \theta_i))| < \varepsilon' \quad i = 1, 2$$

alors :

$$\begin{aligned} R_2(h_2(x^n, x), f(x, \theta_2)) & < R_2(h_1(x^n, x), f(x, \theta_2)) + \varepsilon' \\ & < R_2(h_1(x^n, x), f(x, \theta_1)) + \varepsilon + \varepsilon' \\ & < R_2(h_2(x^n, x), f(x, \theta_1)) + \varepsilon + 2\varepsilon'. \end{aligned}$$

On obtient alors :

$$|R_2(h_2(x^n, x), f(x, \theta_2)) - R_2(h_2(x^n, x), f(x, \theta_1))| < \varepsilon + 2\varepsilon'$$

alors :

$$\begin{aligned} |\delta_2^2(h_1) - \delta_2^2(h_2)| & < |R_2(h_1(x^n, x), f(x, \theta_1)) - R_2(h_2(x^n, x), f(x, \theta_2))| + 2\varepsilon \\ & < |R_2(h_2(x^n, x), f(x, \theta_2)) - R_2(h_2(x^n, x), f(x, \theta_1))| \\ & + |R_2(h_1(x^n, x), f(x, \theta_1)) - R_2(h_2(x^n, x), f(x, \theta_1))| + 2\varepsilon. \\ & < \varepsilon + 2\varepsilon' + \varepsilon' + 2\varepsilon \\ & < 3(\varepsilon + \varepsilon') \end{aligned}$$

et pour $h \in \Omega_{2,K}$ on aura aussi

δ_2 uniformément continue.

4. Détermination de l'unique estimateur minimax de la densité

dans le cas de la famille $F_\theta = \{N(\theta, 1)(x)\}$ avec

$$N(\theta, \sigma)(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right).$$

. De façon analogue au paragraphe BI2 où nous avons déterminé l'estimateur minimax de $\Psi(f) = F$, nous allons, dans le cas de la famille paramétrée normale F_θ , déterminer l'estimateur minimax de $\Psi(f) = f$, $f \in F_\theta$, à l'aide d'une suite de lois a priori sur F_θ .

. Une loi a priori sur F_θ est obtenue de la façon suivante :

. Soit g l'application que l'on suppose continue, définie

par :

$$\begin{array}{ccc} (\theta, B_\theta, \xi) & \xrightarrow{g} & (F_\theta, B_{F_\theta}, \xi') \\ \theta & \rightsquigarrow & g(\theta) = f(\theta, \cdot) \end{array}$$

ξ' sera la mesure image de ξ par g .

On prendra ici $\theta = \mathbb{R}$.

. On notera $H = \{\text{lois a priori } \xi_\sigma \text{ définies sur } \theta\}$.

$$H' = \{\text{lois a priori } \xi'_\sigma = \xi_\sigma g^{-1} \text{ définies sur } F_\theta\}.$$

. On rappelle que $h\xi'_\sigma$, noté encore h_σ étant l'estimateur bayésien correspondant à ξ'_σ , le risque bayésien est :

$$\begin{aligned} r_2(\sigma) &= \int_{F_\theta} R_2[h_\sigma, g(\theta)] d\xi'_\sigma(g(\theta)) \\ &= \int_\theta R_2[h_\sigma, g(\theta)] d\xi_\sigma(\theta). \end{aligned}$$

Soit alors la version du théorème de Phadia [47] dans le cas paramétrique de l'estimation de la densité.

Théorème.-

Soit h_σ une suite d'estimateurs bayésiens de $g(\theta)$ pour les lois a priori (ξ'_σ) sur F_θ et soit $r_2(\sigma)$ la suite correspondante des risques bayésiens.

Si $r_2(\sigma) \xrightarrow{\sigma \rightarrow \infty} C$ et si h^* est un estimateur de $g(\theta)$ tel que $\forall g(\theta) \in F_\theta, R_2[h^*, g(\theta)] \leq C$ alors h^* est un estimateur minimax de $g(\theta)$.

Tout comme dans B.I.2, nous allons voir que l'estimateur minimax sera en fait la limite des estimateurs bayésiens correspondant à une suite infinie de lois a priori sur F_θ . Sans énoncer de théorème général, il semble que cela soit toujours vrai lorsque l'estimateur minimax est unique (voir B.II.2), et nous pensons pouvoir le préciser dans un travail ultérieur.

a) Détermination de l'estimateur bayésien correspondant à la loi image par g de ξ de densité $N(0, \sigma^2)(x)$.

L'estimateur bayésien noté h_σ est donné par : (cf. Nobécourt [44]).

$$h_\sigma(x_1 \dots x_n, x) = \frac{\int_{\mathbb{R}} N(\theta, 1)(x) \cdot N(\theta, 1)(x_1) \dots N(\theta, 1)(x_n) \cdot N(0, \sigma^2)(\theta) d\theta}{\int_{\mathbb{R}} N(\theta, 1)(x_1) \dots N(\theta, 1)(x_n) N(0, \sigma^2)(\theta) d\theta}$$

$$h_\sigma(x_1 \dots x_n, x) = \frac{1}{\sqrt{2\pi}} \frac{\int e^{-\frac{1}{2}[(\theta-x)^2 + (\theta-x_1)^2 + \dots + (\theta-x_n)^2 + \frac{\theta^2}{\sigma^2}]} d\theta}{\int e^{-\frac{1}{2}[(\theta-x_1)^2 + \dots + (\theta-x_n)^2 + \frac{\theta^2}{\sigma^2}]} d\theta}$$

$$h_\sigma(x_1 \dots x_n, x) = \frac{1}{\sqrt{2\pi}} \frac{\int e^{-\frac{1}{2}\left[\left(n+1+\frac{1}{\sigma^2}\right)\theta^2 - 2\theta(x + \sum x_i) + \sum x_i^2\right] - \frac{x^2}{2}} d\theta}{\int e^{-\frac{1}{2}\left[\left(n+\frac{1}{\sigma^2}\right)\theta^2 - 2\theta(\sum x_i) + \sum x_i^2\right]} d\theta}$$

$$h_{\sigma}(x_1 \dots x_n, x) = \frac{1}{\sqrt{2\pi}} \frac{\int e^{-\frac{1}{2} (n+1 + \frac{1}{\sigma^2}) \left[\theta - \frac{x + \sum x_i}{n+1 + \frac{1}{\sigma^2}} \right]^2} e^{\frac{1}{2} \frac{(x + \sum x_i)^2}{n+1 + \frac{1}{\sigma^2}} - \frac{x^2}{2}} d\theta}{\int e^{-\frac{1}{2} (n + \frac{1}{\sigma^2}) \left[\theta - \frac{\sum x_i}{n + \frac{1}{\sigma^2}} \right]^2} e^{\frac{1}{2} \frac{(\sum x_i)^2}{n + \frac{1}{\sigma^2}}} d\theta}$$

$$h_{\sigma}(x_1 \dots x_n, x) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n + \frac{1}{\sigma^2}}{n + 1 + \frac{1}{\sigma^2}}} \frac{e^{\frac{1}{2} \frac{(x + \sum x_i)^2}{n + 1 + \frac{1}{\sigma^2}} - \frac{x^2}{2}}}{e^{\frac{1}{2} \frac{(\sum x_i)^2}{n + \frac{1}{\sigma^2}}}}$$

$$h_{\sigma}(x_1 \dots x_n, x) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n\sigma^2 + 1}{(n+1)\sigma^2 + 1}} e^{-\frac{1}{2} \left\{ -\frac{(x + \sum x_i)^2}{n + 1 + \frac{1}{\sigma^2}} + x^2 + \frac{(\sum x_i)^2}{n + \frac{1}{\sigma^2}} \right\}}$$

$$h_{\sigma}(x_1 \dots x_n, x) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n\sigma^2 + 1}{(n+1)\sigma^2 + 1}} e^{-\frac{1}{2} \left\{ x^2 \left(1 - \frac{1}{n + 1 + \frac{1}{\sigma^2}} - \frac{2x \sum x_i}{n + 1 + \frac{1}{\sigma^2}} + (\sum x_i)^2 \alpha \right) \right\}}$$

$$\text{Avec } \alpha = \frac{\sigma^4}{(n\sigma^2 + 1) [(n+1)\sigma^2 + 1]}$$

On obtient alors

$$h_{\sigma}(x_1 \dots x_n, x) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n\sigma^2 + 1}{(n+1)\sigma^2 + 1}} \exp \left\{ -\frac{\left[x - \frac{n\sigma^2 \bar{X}_n}{n\sigma^2 + 1} \right]^2}{2 \cdot \frac{(n+1)\sigma^2 + 1}{n\sigma^2 + 1}} \right\}$$

b) Calcul du risque $R_2(h_{\sigma}, g(\theta))$ de cet estimateur bayésien.

avec $R_2(h_{\sigma}, g(\theta)) = \int \text{var}(h_{\sigma}(x^n, x)) dx.$

Nous allons d'abord calculer la variance de l'estimateur et pour cela nous aurons besoin du lemme suivant de la théorie des distributions normales (voir par exemple [71], p. 134).

Lemme. -

Soient Y et W deux variables aléatoires réelles telles que Y/W suit une loi normale $N(W, \sigma^2)$

W suit une loi normale $N(\xi, \tau^2)$ alors la distribution marginale de Y est une loi normale $N(\xi, \sigma^2 + \tau^2)$.

D'autre part, l'espérance d'une densité conditionnelle étant une densité marginale, on obtient :

$$\text{var} [\bar{h}_\sigma(x^n, x)] = E_{P_\theta} [\bar{h}_\sigma^2(x^n, x)] - 2E_{P_\theta} [\bar{h}_\sigma(x^n, x)] \cdot f(\theta, x) + f^2(\theta, x)$$

avec

$$\bullet \quad E_{P_\theta} [\bar{h}_\sigma^2(x^n, x)] = E \left[\frac{1}{2\pi} \cdot \frac{n\sigma^2 + 1}{(n+1)\sigma^2 + 1} \exp \left\{ - \frac{\left[x - \bar{X}_n \frac{n\sigma^2}{n\sigma^2 + 1} \right]^2}{\frac{(n+1)\sigma^2 + 1}{n\sigma^2 + 1}} \right\} \right]$$

$$= \frac{1}{2\sqrt{2\pi}} \frac{\sqrt{2(n\sigma^2 + 1)}}{\sqrt{(n+1)\sigma^2 + 1}} E \left[\frac{1}{\sqrt{2\pi}} \sqrt{\frac{2(n\sigma^2 + 1)}{(n+1)\sigma^2 + 1}} \exp \left\{ \begin{matrix} \text{''} & \text{''} \\ \text{''} & \text{''} \end{matrix} \right\} \right]$$

l'expression sous le symbole espérance est une densité conditionnelle

avec $\frac{n\sigma^2}{(n+1)\sigma^2 + 1} \cdot \bar{X}_n$ suivant une loi normale :

$$N\left(\frac{n\sigma^2}{(n+1)\sigma^2 + 1} \cdot \theta, \frac{n\sigma^4}{[(n+1)\sigma^2 + 1]^2}\right)$$

d'où : $\int E_{P_\theta} [\bar{h}_\sigma^2(x^n, x)] dx = \frac{1}{2\sqrt{2\pi}} \frac{\sqrt{2(n\sigma^2 + 1)}}{\sqrt{(n+1)\sigma^2 + 1}}$

d'autre part •

$$E_{P_\theta} [\bar{h}_\sigma(x^n, x)] = E \left[\frac{1}{\sqrt{2\pi}} \sqrt{\frac{n\sigma^2 + 1}{(n+1)\sigma^2 + 1}} \exp \left\{ - \frac{\left[x - \frac{n\sigma^2 \bar{X}_n}{n\sigma^2 + 1} \right]^2}{2 \cdot \frac{(n+1)\sigma^2 + 1}{n\sigma^2 + 1}} \right\} \right]$$

Ici encore l'expression sous le symbole espérance est une densité conditionnelle. On obtient alors une densité marginale normale

$$N\left(\frac{n\sigma^2}{n\sigma^2 + 1} \cdot \theta, \frac{(n+1)\sigma^2 + 1}{n\sigma^2 + 1} + \frac{n\sigma^4}{(n\sigma^2 + 1)^2}\right)$$

On a alors

$$- 2 \int E_{P_\theta^n}(h_\sigma(x^n, x)) \cdot f(\theta, x) dx =$$

$$= - 2 \int \frac{n\sigma^2 + 1}{\sqrt{2\pi} \sqrt{\sigma^4(n^2 + 2n) + \sigma^2(2n+1) + 1}} \exp\left\{-\frac{\left(x - \frac{n\sigma^2}{n\sigma^2 + 1} \theta\right)^2 (n\sigma^2 + 1)^2}{2(\sigma^4(n^2 + 2n) + \sigma^2(2n+1) + 1)}\right\} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} dx$$

Un calcul élémentaire nous permet d'écrire l'expression sous la forme :

$$\frac{- 2(n\sigma^2 + 1)}{\sqrt{2\pi} \sqrt{2\pi} \sqrt{\sigma^4(n^2 + 2n) + \sigma^2(2n+1) + 1}} \exp\left\{-\frac{\left[x - \theta \frac{2n\sigma^4(n+1) + \sigma^2(3n+1) + 1}{2n\sigma^4(n+1) + \sigma^2(4n+1) + 2}\right]^2}{2 \cdot \frac{\sigma^4(n^2 + 2n) + \sigma^2(2n+1) + 1}{2n\sigma^4(n+1) + \sigma^2(4n+1) + 2}} - \frac{\theta^2}{2} \beta\right\} dx$$

avec $\beta = \frac{1}{2n\sigma^4(n+1) + \sigma^2(4n+1) + 2}$

finalement :

$$- 2 \int E_{P_\theta^n}(h_\sigma(x^n, x)) \cdot f(\theta, x) dx = - \frac{2(n\sigma^2 + 1)}{\sqrt{2\pi} \sqrt{2n\sigma^4(n+1) + \sigma^2(4n+1) + 2}} \times \dots / \dots$$

$$\dots / \dots \times \exp\left\{-\frac{\theta^2}{2[2n(n+1)\sigma^4 + \sigma^2(4n+1) + 2]}\right\}$$

et :

$$R(h_\sigma, g(\theta)) = \frac{1}{2\sqrt{\pi}} \left[\sqrt{\frac{n\sigma^2 + 1}{(n+1)\sigma^2 + 1}} + 1 \right] - \dots/\dots$$

$$- \frac{\sqrt{2}(n\sigma^2 + 1)}{\sqrt{\pi} \sqrt{2n(n+1)\sigma^4 + \sigma^2(4n+1) + 2}} \exp\left\{-\frac{\theta^2}{2[2n(n+1)\sigma^4 + \sigma^2(4n+1) + 2]}\right\}$$

c) Calcul du risque bayésien $r_{2\sigma}$ correspondant à la loi
a priori sur $\theta = \mathbb{R}$ de densité $N(0, \sigma^2)$

avec $r_{2\sigma} = \int R_2(h_\sigma, g(\theta)) \cdot N(0, \sigma^2) d\theta.$

Dans l'expression de $R(h_\sigma, g(\theta))$, deux termes sont indépendants de θ . Pour déterminer r_σ , il suffit donc de calculer

$$- \int \frac{\sqrt{2}}{\sqrt{\pi}} \frac{(n\sigma^2 + 1)}{\sqrt{2n(n+1)\sigma^4 + \sigma^2(4n+1) + 2}} \exp\left\{-\frac{\theta^2}{2[2n(n+1)\sigma^4 + \sigma^2(4n+1) + 2]}\right\} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\theta^2}{2\sigma^2}} d\theta$$

ou

$$- \int \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2}}{\sqrt{\pi}} \frac{\sigma^{-1}(n\sigma^2 + 1)}{\sqrt{2n(n+1)\sigma^4 + \sigma^2(4n+1) + 2}} \exp\left\{-\frac{\theta^2[2n(n+1)\sigma^4 + \sigma^2(4n+2) + 2]}{2\sigma^2[2n(n+1)\sigma^4 + 4n+1)\sigma^2 + 2]}\right\} d\theta$$

égal à :

$$- \frac{\sqrt{2}(n\sigma^2 + 1)}{\sqrt{\pi} \sqrt{2n(n+1)\sigma^4 + \sigma^2(4n+2) + 2}}$$

alors :

$$r_{2\sigma} = \frac{1}{2\sqrt{\pi}} \left[\sqrt{\frac{n\sigma^2 + 1}{(n+1)\sigma^2 + 1}} + 1 \right] - \frac{\sqrt{2}(n\sigma^2 + 1)}{\sqrt{\pi} \sqrt{2n(n+1)\sigma^4 + \sigma^2(4n+2) + 2}}$$

d) Détermination d'une famille d'estimateurs stricts de la densité, à risque constant sur F_θ .

On peut énoncer :

Proposition 1. - Tout estimateur $h(x^n, x) = \frac{1}{\sqrt{2\pi} A(n)} e^{-\frac{(x-T_n(x^n))^2}{2 \cdot A^2(n)}}$,

où $T_n(x^n)$ suit une loi normale de moyenne $\theta \in \mathbb{R}$, est tel que son risque est indépendant de $g(\theta) \in F_\theta$.

On nomme Ω_2^{cte} cette sous-famille d'estimateurs de Ω_2 .

Démonstration :

$$R_2(h, g(\theta)) = \int \{E[h^2(x^n, x)] - 2E[h(x^n, x)] \cdot f(\theta, x) + f^2(\theta, x)\} dx$$

est indépendant de θ car : $E[h^2(x^n, x)]$ et $E[h(x^n, x)]$ le sont si $E(T_n) = \theta$. Il suffit en effet d'appliquer le lemme cité dans b. et la remarque qui suit ce lemme.

Soit alors l'estimateur $h^*(x^n, x)$ élément de Ω_2^{cte} correspondant à $A^2(n) = \frac{n+1}{n}$ et à $T_n(x^n) = \bar{X}_n$. On a donc

$$h^*(x^n, x) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+1}} \exp\left\{-\frac{(x - \bar{X}_n)^2}{2 \cdot \frac{n+1}{n}}\right\}$$

Le choix d'un tel estimateur est justifié car h^* est tel que

1. $h^*(x^n, x) = \lim_{\sigma \rightarrow \infty} h_\sigma(x^n, x)$.

2. L'estimateur h^* est asymptotiquement sans biais

car $E[h^*(x^n, x)] = E\left[\frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{n}{n+1}} \exp\left\{-\frac{(x-\bar{X}_n)^2}{2 \cdot \frac{n+1}{n}}\right\}\right]$

$$= \frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{n}{n+2}} \exp \left\{ -\frac{(x-\theta)^2}{2 \cdot \frac{n+2}{n}} \right\}$$

en effet \bar{X}_n suit une loi $N(\theta, \frac{1}{n})$.

3. $h^*(x^n, x)$ est strict.
4. $h^*(x^n, x)$ est à risque constant.

Remarque : Le choix de $T_n(x^n) = \bar{X}_n$ est dicté par le fait que \bar{X}_n est l'unique estimateur minimax de θ pour la perte quadratique et que c'est aussi l'estimateur sans biais à variance minimale.

Calcul du risque de l'estimateur h^* .

D'après la méthode étudiée en b., on obtient de façon immédiate que

$$R_2(h^*, g(\theta)) = \frac{1}{2\sqrt{\pi}} \left[1 - \sqrt{\frac{n}{n+1}} \right].$$

On peut alors citer :

Proposition 2. - L'estimateur h^* défini par :

$$h^*(x^n, x) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+1}} \exp \left\{ -\frac{(x-\bar{X}_n)^2}{2 \cdot \frac{n+1}{n}} \right\} \text{ est l'estimateur}$$

minimax de la densité appartenant à $F_\theta = \{N(\theta, 1)(x) \quad \theta \in \mathbb{R}\}$ pour la fonction de perte $L_2(\lambda)$, λ mesure de Lebesgue.

Il est de plus tel que

1. $h^* \in \Omega_2^{cte} \subset \Omega_2$
2. $h^*(x^n, x) = \lim_{\sigma \rightarrow \infty} h_\sigma(x^n, x)$
3. h^* est asymptotiquement sans biais.

4. h^* est strict et à risque constant sur F_θ .

Démonstration :

Soit $r_2(\sigma)$ obtenu en c.

Il est immédiat que :

$$\lim_{\sigma \rightarrow \infty} r_2(\sigma) = R_2(h^*, g(\theta)) = \frac{1}{2\sqrt{\pi}} \left[1 - \sqrt{\frac{n}{n+1}} \right].$$

Comme $R_2(h^*, g(\theta)) = \text{cte} \quad \forall \theta \in \Theta$, on peut conclure à l'aide de la version du théorème de Phadia [47] citée au début du 4.

L'existence de cet estimateur étant assurée par le fait que $F_\theta \subset L^2(\mathbb{R})$ (condition d'existence étudiée en B.II.1). Cependant, on ne peut utiliser B.II.2 pour montrer l'unicité car ici Θ n'est pas compact.

Remarque :

Nous reviendrons au chapitre II, B, I,4, sur l'efficacité de cet estimateur minimax et sur l'efficacité d'un estimateur sans biais analogue.

CHAPITRE II

EFFICACITE DES ESTIMATEURS DE LA LOI

Dans le 1er chapitre, nous avons vu que l'estimation minimax dépend étroitement de :

1. l'ensemble des fonctions à estimer.
2. la fonction de perte choisie.

(Voir B.I. par exemple).

Cette classe d'estimateurs est donc différente par nature des estimateurs classiques. L'exemple B.II.4. prouve la difficulté à déterminer l'estimateur minimax (ou les estimateurs car l'unicité dans le cas non paramétrique semble très délicate à prouver). On comprend donc pourquoi ce mode d'estimation est peu utilisé. (La bibliographie en témoigne). Cependant une étude de l'efficacité des estimateurs pour la méthode du noyau, et pour la méthode par projection va montrer que ceux-ci (dans le cas de certaines fonctions de perte et pour certains ensembles précis de densité) sont des estimateurs minimax de façon asymptotique.

Ils possèdent donc dans certains cas le caractère d'estimateurs prudents au même titre que certains estimateurs sont stricts ou sans biais.

Des conditions seront données sur le noyau $K_{r(n)}(x,t)$ pour que la famille générale d'estimateurs de Bosq-Bleuez [5,10] soit asymptotiquement minimax pour la moyenne quadratique et dans le cas de la fonction de perte $L_2(\lambda)$.

Nous étudierons principalement la méthode du noyau et la méthode par projection.

Rappel.

Soient (Ω, \mathcal{A}, P) un espace probabilisé, (E, \mathcal{B}, μ) un espace mesuré, et X une variable aléatoire définie sur (Ω, \mathcal{A}) , à valeurs dans (E, \mathcal{B}) , dont la loi P_X admet une densité $\frac{dP_X}{d\mu}$ par rapport à μ , que l'on cherche à estimer

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (E, \mathcal{B}) .$$

. Méthode du noyau.

On suppose que E est égal à \mathbb{R}^S , $S \in \mathbb{N}^*$. On se donne une application K de E dans \mathbb{R} , une suite (k_n) , de nombres strictement positifs tendant vers zéro, et on prend comme estimateur de la densité :

$$\hat{h}_n^1(x, \omega) = \frac{1}{n k_n^S} \sum_{j=1}^n K\left(\frac{x - X_j(\omega)}{k_n}\right), \quad x \in E, \quad \omega \in \Omega.$$

En général, on choisit comme fonction K une densité de probabilité. Cette méthode a été introduite par Rosenblatt [51].

. Méthode des fonctions orthogonales ou méthode par projection.

On suppose que la densité $\frac{dP_X}{d\mu}$ appartient à $L^2(\mu)$, et on désigne par (e_i) $i \in I$ une base hilbertienne de $L^2(\mu)$ (supposée séparable). On se donne une suite d'entiers $q(n)$ tendant vers l'infini et on prend comme estimateur de la densité :

$$h_n^2(x, \omega) = \sum_{i=1}^{q(n)} \hat{a}_{i,n}(\omega) e_i(x), \quad x \in E, \quad \omega \in \Omega$$

avec :

$$\hat{a}_{i,n} = \frac{1}{n} \sum_{j=1}^n e_i(X_j).$$

Cette méthode fut introduite par Cencov [17] en 1962.

On notera aussi :

- . $\hat{h}_n^{2,H}$ L'estimateur correspondant à la méthode par projection,
cas des fonctions d'hermite,
- . $\hat{h}_n^{2,T}$ " " à la méthode par projection,
cas des fonctions trigonométriques,
- . \hat{h}_n^3 " " à l'algorithme polynomial.

. Soit $(K_r)_{r \in I}$, une famille de fonctions réelles mesurables, à 2 variables, éléments de E . On associe à cette famille une suite d'estimateurs de la densité définie de la façon suivante :

$$\hat{h}_n(x) = \frac{1}{n} \sum_{j=1}^n K_{r(n)}(X_j, x), \quad x \in E$$

I partie infinie et non bornée de \mathbb{R}^+ et où $r(n) \in I$ et $r(n) \rightarrow +\infty$.

. W_p^m désignera l'espace de Sobolev des densités f telles que les dérivées $f^{(i)}$ ($i = 1, \dots, m-1$) sont absolument continues et telles que $f^{(m)}$ soit élément de $L^p(\lambda)$, λ mesure de Lebesgue.

. $W_p^m(M)$ sera alors constitué par les éléments de W_p^m tels que

$$\|f^{(m)}\|_p \leq M \quad (M > 0).$$

. Z_p^m est formé des éléments de W_p^m tels que $f^{(i)} \in L^p(\lambda)$
 $i = 0, 1, \dots, m-1$.

. C_λ est l'ensemble des densités bornées et satisfaisant une condition de Lipschitz d'ordre λ en $x = 0$, $0 < \lambda < 1$.

. C_{λ}^U est l'ensemble des densités continues f telles que $\exists \alpha \in \mathbb{R}^+$
 $|f(x) - \alpha| \leq 2(\lambda+1) x^{\lambda} \quad 0 < \lambda < 1$ $\forall x \in \mathbb{R}$

. f désignera une version spécifiée de la densité.

. D'autres notations relatives à $p = 2$ seront données au début de B.II.

(A) Effacité des estimateurs de la densité dans
le cas de la fonction de perte quadratique.

La plupart des travaux sur l'estimation des densités établissent les propriétés asymptotiques de certaines classes d'estimateurs, qu'il s'agisse de l'étude des vitesses de convergence comme dans [6, 21, 42, 59, 63] ou d'études de convergence en loi, comme dans [46]. Ces propriétés sont en général uniformes sur certaines classes de densités. Ces travaux concernent soit l'estimation d'une densité en un point (estimation ponctuelle), soit l'estimation de la densité toute entière (estimation globale).

En ce qui concerne l'estimation ponctuelle, le premier qui ait abordé le problème inverse, c'est-à-dire celui de la minoration du risque, est Farrell : ayant remarqué que les estimateurs optimaux dans chaque classe, qu'il s'agisse des estimateurs à noyau de Parzen [46], de ceux du maximum de vraisemblance, ou de ceux construits à partir de séries orthogonales [5, 6, 53, 63], atteignaient tous le même ordre en n , taille de l'échantillon, il a démontré dans [22] qu'il était impossible de faire mieux sur certaines classes de densités proches d'un polynôme, établissant ainsi une borne de Cramer-Rao en dimension infinie. Sa méthode consiste à employer un couple de densités de l'ensemble choisi, globalement voisines mais perturbées au voisinage du point où se fait l'estimation. Le couple est adapté à n .

Wahba dans [59] a étendu ce résultat ponctuel à des espaces de Sobolev de densités, en calculant par ailleurs la vitesse effectivement atteinte sur ces mêmes espaces pour quatre classes distinctes d'estimateurs.

Définition.

Soient F un ensemble de densités et \hat{h}_n un estimateur de f en x ; si il existe des constantes réelles, positives c, d, u ne dépendant que de F , telles que l'on ait simultanément :

$$\cdot \inf_{\hat{h}_n} \sup_{f \in F} R(\hat{h}_n, f) \geq c.n^{-u}$$

et pour un estimateur \hat{h}_n^* au moins :

$$\cdot \sup_{f \in F} R(\hat{h}_n^*, f) \leq d.n^{-u},$$

alors on dira que \hat{h}_n^* est un estimateur asymptotique minimax.

Nous allons énoncer les résultats obtenus pour deux ensembles de densités : $W_p^m(M)$ et C_λ .

• On pose :

$$\phi(m, p) = \frac{2m - \frac{2}{p}}{2m + 1 - \frac{2}{p}} \quad (\text{fonction croissante de } p).$$

I - Etude de la convergence en moyenne quadratique sur $W_p^m(M)$ et Z_p^m .

1. - Minoration du risque R sur $W_p^n(M)$.

• On travaille avec $p \geq 1$. R désigne ici le risque associé à la perte quadratique.

Sans perdre de généralité, on supposera que l'on estime f en $x = 0$.

On a alors le théorème suivant :

Théorème [59].-

Soit $(a_n)_{n \geq 1}$ une suite de nombres positifs telle que

$$\lim_{n \rightarrow \infty} \inf_{f \in W_p^m(M)} \inf P_f (|\hat{h}_n - f(0)| \leq a_n) = 1.$$

Alors $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \inf n^{\Phi(m, p+\varepsilon)} a_n^2 = \infty.$

Démonstration :

La démonstration est adaptée de celle de Farrell [22].

Corollaire.-

Supposons que pour tout $\varepsilon > 0$, on ait :

$$\sup_{f \in W_p^m(M)} R(\hat{h}_n, f) = b_n n^{-\Phi(m, p+\varepsilon)}$$

alors il existe $D_0 > 0$ telle que $b_n > D_0$ pour une infinité de n .

Alors pour la convergence en moyenne quadratique, la meilleure vitesse qui est uniforme sur $W_p^m(M)$ n'est pas supérieure à $\Phi(m, p+\varepsilon)$, ε arbitrairement petit.

Définition.-

Un noyau K (déterminant l'estimateur \hat{h}_n^1) sera de "type m"

si :

(1) $\sup_{-\infty < y < +\infty} K(y) < \infty$	(2) $\int_{-\infty}^{+\infty} K(y) dy < \infty$
(3) $\lim_{y \rightarrow \infty} y K(y) = 0$	(4) $\int_{-\infty}^{+\infty} K(y) dy = 1$
(5) $\int_{-\infty}^{+\infty} y^i K(y) dy = 0 \quad i=1, 2, \dots, m-1$	(6) $\int_{-\infty}^{+\infty} y ^{m-\frac{1}{p}} K(y) dy < \infty$

2. - Vitesse atteinte par certains estimateurs sur $W_p^m(M)$ et Z_2^m .

a. On peut citer :

Proposition (d'après [58], [59]).- Soit $F \subset W_p^m(M)$.

Il existe $D > 0$ telle que :

$$\bullet \sup_{f \in F} R(\hat{h}_n^i, f) \leq D \cdot n^{-\Phi(m,p)} \quad i = 1, 2, 3$$

dans le cas de l'estimateur :

- \hat{h}_n^1 pour K noyau de type m .
- $\hat{h}_n^{2,T}$ f à support compact et $p = 2$
- \hat{h}_n^3 f étant bornée supérieurement pour tout x et inférieurement au voisinage de 0 .
 $|x(1-F(x))|$ (resp. $|x F(x)|$) étant bornée pour $x \geq 0$ (resp. $x \leq 0$). (F fonction de répartition).

b. Dans le cas de $\hat{h}_n^{2,H}$, on peut énoncer une proposition qui complète les résultats de Schwartz [53] et Walter [63].

Proposition.- Soit $F \subset Z_2^m$ l'ensemble des densités f bornées à support compact dont le développement en série d'hermite converge vers f en tout point. Alors pour $q(n) = O(n^{\frac{1}{2m-1}})$ et $f \in F$.

$$\bullet R(\hat{h}_n^{2,H}, f) = O(n^{-2m-2/2m-1})$$

Démonstration :

Sur cet ensemble de densités, la variance de l'estimateur converge vers 0 avec $\frac{q(n)}{n^2}$ (voir [5] page 59).

D'autre part, le biais converge vers 0 avec $q(n)^{-m+1}$, d'où le résultat.

Walter [63] majore la variance à l'aide de l'inégalité de Schwartz, avec : $h_n^{2,H} = E_f[\hat{h}_n^{2,H}]$ et $a_{i,n} = E_f(\hat{a}_{i,n})$ (e_i désigne ici la $i^{\text{ème}}$ fonction d'hermite et obtient :

$$E_f [\hat{h}_n^{2,H}(x) - h_n(x)]^2 = E_f \left[\sum_{i=1}^{q(n)} (\hat{a}_{i,n} - a_{i,n}) \cdot e_i(x) \right]^2$$

$$\leq E_f \left[\sum_{i=1}^{q(n)} (\hat{a}_{i,n} - a_{i,n})^2 \right] \sum_{i=1}^{q(n)} e_i^2(x)$$

d'autre part $E_f (\hat{a}_{i,n} - a_{i,n})^2 < \frac{(i+1)^{-1/2}}{n}$ car d'après [56] ; p. 242,

si $x \in \underset{A < \infty}{[-A, +A]}$ $|e_i(x)| < \frac{C_A}{(i+1)^{1/4}}$

d'où :

$$E_f [\hat{h}_n^{2,H}(x) - h_n^{2,H}(x)]^2 \leq A \cdot \frac{q(n)}{n},$$

qui constitue une majoration de la variance, moins fine que celle obtenue grâce à [5].

3. - Conditions sur le noyau $K_{r(n)}$ pour que la famille générale d'estimateurs (\hat{h}_n) converge à la même vitesse qu'en 2.
sur $Z_p^m \subset W_p^m$.

On rappelle que :

$$\hat{h}_n(x) = \frac{1}{n} \sum_{j=1}^n K_{r(n)}(X_j, x).$$

E désigne ici un intervalle ouvert (a,b) de \mathbb{R} .

Le symbole $(-m)$ désignera l'antidérivée d'ordre m, ce qui permettra de "mesurer" la vitesse de convergence de la suite $K_{r(n)}$.

Rappels : d'après Korevaar tome I [35] et Mikusinsky [40].

. Soit g une fonction intégrale (ainsi que sa valeur absolue) sur $E = (a,b)$.

Alors $g^{(-1)}$ l'antidérivée d'ordre 1 ou l'intégrale indéfinie est définie par $g^{(-1)}(t) = \gamma + \int_c^t g(u)du$, γ constante.

$g^{(-m)}$ l'antidérivée d'ordre m ou l'intégrale indéfinie d'ordre m est définie par :

$$g^{(-m)}(t) = p(t) + \int_c^t dt_1 \int_c^{t_1} dt_2 \dots \int_c^{t_{m-1}} g(t_m) dt_m$$

où $p(t)$ est un polynôme de degré strictement inférieur à m .

. Le symbole δ représente la distribution de Dirac et δ_x est telle que $\delta_x(t) = \delta(x-t)$.

δ définie par exemple par la suite fondamentale

$$f_k(u) = \begin{cases} k & \text{pour } 0 < u < \frac{1}{k} \\ 0 & \text{sinon} \end{cases} \quad k = 1, 2, \dots$$

et $\int \delta(u)\psi(u)du = \lim_{k \rightarrow \infty} \int f_k(u)\psi(u)du = \psi(0)$, pour toute fonction continue.

Soient alors les hypothèses suivantes sur $K_{r(n)}$ d'après Walter-Blum [64]

H.1. $(K_{r(n)}(x,t))$ forme une suite de fonctions bornées et mesurables sur $E \times E$ ($E = (a,b)$).

H.2. $\forall x \in E$
 $\forall \psi \in C^\infty / \text{Supp } \psi \subset E$
 $\lim_{r(n) \rightarrow \infty} \int K_{r(n)}(x,t)\psi(t)dt = \psi(x)$.

H.3. $\forall x \in E$, $K_{r(n)}(x,.) \in L^2(E)$.

H.4. $\|K_{r(n)}(x,.)\|_2^2 = o(r(n))$ uniformément pour $x \in E$

. $K_{r(n)}^{(-m)}(x,t)$ désignant l'antidérivée d'ordre m par rapport à t de $K_{r(n)}(x,t)$ qui est nulle en a , ainsi que ces $(m-1)$ premières dérivées.

$$\delta_x^{(-m)}(t) = \delta_x^{(-m)}(x-t) = \frac{(x-t)_+^{m-1}}{(s-1)!}.$$

On suppose pour cet m il existe $q \geq 1$ et β avec $0 < \beta \leq m + \frac{1}{q} - 1$ tel que :

H.5. $K_{r(n)}^{(-m)}(x, \cdot) - \delta_x^{(-m)}(\cdot) \in L^q(E)$ pour $x \in E$

H.6. $\|K_{r(n)}^{(-m)}(x, \cdot) - \delta_x^{(-m)}(\cdot)\|_q = O(r(n)^{-\beta})$ uniformément pour $x \in E$.

H.7. $|K_{r(n)}^{(-k)}(x, b) - \delta_x^{(-k)}(b)| = O(r(n)^{-\beta})$
uniformément pour $x \in E$ et $k \in \{1, \dots, m\}$.

Avant d'énoncer le théorème général, nous avons besoin de deux lemmes préparatoires.

Lemme 1.- Soit f une densité définie sur E telle que $f \in Z_p^m$ alors $f \in L^\infty(E)$.

Démonstration : On a :

$$|f^p(x) - f^p(a)| \leq p \int_a^x f^{p-1}(t) \cdot |f'(t)| dt$$

et d'après l'inégalité de Holder et Jensen :

$$" \leq p \|f\|_p^{p/q} \cdot \|f'\|_p < \infty, \text{ d'après l'hypothèse}$$

d'où $f^p \in L^\infty(E)$ et $f \in L^\infty(E)$.

Lemme 2.- Soit f une densité définie sur E telle que $f \in Z_p^m$ alors $f \in L^2(E)$.

Démonstration : D'après le lemme 1, $f \in L^\infty(E)$.

Comme $f \in L^1(E)$: $\int f^2(x) dx \leq \|f\|_\infty \cdot \|f\|_1 < \infty$.

On a alors :

Théorème d'après [64].-

Soit $m \geq 1$ et l'ensemble F des densités définies sur E avec $F \subset Z_p^m$. Alors si

* $(K_{r(n)}(x,t))$ satisfait H.1. à H.7. avec $\frac{1}{p} + \frac{1}{q} = 1$

* $f \in F$ et $r(n) = O(n^{\frac{1}{1+2\beta}})$ $\beta = m - \frac{1}{p}$.

On a, uniformément sur E :

$$R(\hat{h}_n, f) = O(n^{-\phi(m,p)})$$

Démonstration :

$$E \left[\hat{h}_n(x) - f(x) \right]^2 = \frac{1}{n} \left[\int_E K_{r(n)}^2(x,t) f(t) dt - \left(\int_E K_{r(n)}(x,t) f(t) dt \right)^2 \right] + \left[\int_E K_{r(n)}(x,t) f(t) dt - f(x) \right]^2$$

et • $\int_E K_{r(n)}^2(x,t) f(t) dt \leq \|f\|_\infty \|K_{r(n)}(x, \cdot)\|_2^2$ d'après lemme 1 et H.3.
 " $= O(r(n))$ d'après H.4.

• $\left(\int_E K_{r(n)}(x,t) f(t) dt \right)^2 \leq \|f\|_2^2 \|K_{r(n)}(x, \cdot)\|_2^2$
 $= O(r(n))$ d'après lemme 2 et H.3. et H.4.

D'autre part, on majore le biais en intégrant m fois par parties

$$\left| \int_E K_{r(n)}(x,t) f(t) dt - f(x) \right| = \left| \int (K_{r(n)}(x,t) - \delta(x-t) \cdot f(t)) dt \right| \text{ d'après H.7.}$$

$$\leq \left| \int [K_{r(n)}^{(-m)}(x,t) - \delta^{(-m)}(x-t)] f^{(m)}(t) dt \right| + O(r^{-\beta}(n))$$

et, avec l'inégalité de Hölder

$$\leq \|K_{r(n)}^{(-m)}(x, \cdot) - \delta_x^{(-m)}(\cdot)\|_q \|f^{(m)}\|_p + O(r^{-\beta}(n)).$$

On peut alors conclure grâce à H.6.

Remarque 1.- Dans le cas où :

$$K_{r(n)}(x,t) = \sum_{j=1}^{r(n)} e_j(x) e_j(t) \quad (e_j) \text{ fonctions trigonométriques}$$

H.1. à H.7. sont bien vérifiées et pour $p = 2$, on retrouve le résultat précédent.

Remarque 2. - La vitesse précédente n'est pas une vitesse maximale "absolue", en effet si l'on considère l'estimateur \hat{h}_n^1 , associé au noyau intégral de Fourier et si la densité f à estimer est telle que la fonction caractéristique décroît exponentiellement (voir Davis [18]), alors

$$\bullet R(\hat{h}_n^1, f) = O\left(\frac{\log n}{n}\right).$$

Plus simplement encore, si on utilise une densité élément de F_θ , $F_\theta = \{N(\theta, 1)(x)\}$, Chap. I. B. II. 4 dans le cas de l'estimateur minimax h^* , on a alors pour $x \in \mathbb{R}$

$$\bullet R(h^*, f) = O\left(\frac{1}{n}\right).$$

II - Etude de la convergence en moyenne quadratique sur C_λ .

1. - Minoration du risque R sur C_λ^U .

On peut énoncer le théorème suivant dû à Farrell [21] et [22]

En fait nous n'utilisons qu'un cas particulier, Farrell [21] et [22] ayant travaillé sur des ensembles de densités proches d'un polynôme.

Théorème. -

Soit $(a_n)_{n \geq 1}$ une suite de nombres positifs telle que :

$$\liminf_{n \rightarrow \infty} \inf_{f \in C_\lambda^U} \inf_{2\lambda} P_f \left[\left| \hat{h}_n - f(0) \right| \leq a_n \right] = 1.$$

$$\text{Alors } \liminf_{n \rightarrow \infty} \inf_{2\lambda+1} \frac{2}{a_n} = \infty.$$

Démonstration : Elle est identique à celle de Farrell [22],

en remarquant que, avec les notations de Farrell, $C_\lambda^U = C_{0,n}$ avec $\eta(x) = x^{1+\lambda}$.

Corollaire. -

$$\text{Si } \sup_{f \in C_\lambda^U} R(\hat{h}_n, f(0)) = b_n \cdot n^{-\frac{2\lambda}{2\lambda+1}} .$$

Alors il existe $D_0 > 0$ tel que $b_n \geq D_0$ pour une infinité de n .

Démonstration :

D'après l'inégalité de Bienaymé Tchebycheff et les hypothèses

$$P_f \left[\left| \hat{h}_n - f(0) \right| \leq a_n \right] \geq 1 - \frac{E_f \left[\hat{h}_n - f(0) \right]^2}{a_n^2}$$

$$\geq 1 - \frac{b_n \cdot n^{-\frac{2\lambda}{2\lambda+1}}}{a_n^2}$$

Si $a_n = n^{-\frac{\lambda}{2\lambda+1}}$ et si $\limsup b_n = 0$, la conclusion du théorème précédent est mise en défaut, alors que son hypothèse est vérifiée.

2. - Conditions sur le noyau $K_{r(n)}$ pour que la famille générale d'estimateurs \hat{h}_n converge à la vitesse maximale sur C_λ .

Nous aurons besoin de la notion de suite $(K_{r(n)})$ dominée par un noyau K . (On considère ici $E = \mathbb{R}$).

Soit un noyau K satisfaisant aux conditions suivantes :

C.1. $K \in L^\infty(\mathbb{R})$

C.2. $K \geq 0$

C.3. $\int K(x) dx = 1$

C.4. $\int |x|^q K(x) dx < \infty$ pour $q = 1$ ou :

C.4. $\int [\log_+ |x|] \cdot |x|^q K^q(x) dx$ pour $q > 1$.

Définition 1. (d'après Walter-Blum [64]).-

Une suite de noyaux $(K_{r(n)})$ satisfaisant les hypothèses H.1. et H.2. précédentes, sera dite dominée par le noyau K satisfaisant C.1., C.2., C.3., C.4. ou C'.4. si

$$\forall x \in \mathbb{R} \quad a. \quad K_{r(n)}(x, t) \geq 0$$

$$\forall t \in \mathbb{R} \quad b. \quad \int_{-\infty}^{+\infty} K_{r(n)}(x, t) dt = 1$$

$$c. \quad K_{r(n)}(x, t) \leq c.r(n).K[r(n)(t-r)]$$

c constante strictement positive.

On obtient alors la proposition suivante :

Proposition.-

Soit $(K_{r(n)})$ une suite de noyaux vérifiant H.1. et H.2. et dominée par K (satisfaisant C.1., C.2., C.3., C.4., ou C'.4.).

Alors $(K_{r(n)})$ satisfait aux hypothèses H.3. à H.7., avec $m = 1$ et $\beta = \frac{1}{q}$.

Démonstration : contenue implicitement dans [64].

Définition 2. d'après [64].-

Une suite de noyaux $(K_{r(n)})$ vérifiant H.1. et H.2. et dominée par K défini par $K(t) = \frac{1}{(1+t^2)}$, sera dite de type Fejer.

Remarque. Exemples.

. Il est immédiat que la suite de type Fejer vérifie la proposition précédente avec $q > 1$.

. La condition de la définition est vérifiée pour le noyau de Fejer, pour tout noyau de Parzen à support compact et positif.

Soit alors le théorème (comme précédemment, on étudie la convergence en $x = 0$).

Théorème d'après [64].

Soient f une densité bornée, élément de C_λ , et $K_{r(n)}$ une suite de type Fejer. Alors l'estimateur \hat{h}_n associé sera pour

$$r(n) = n^{\frac{1}{1+2\lambda}} \text{ tel que :}$$

$$\bullet R(\hat{h}_n, f) = O(n^{-\frac{2\lambda}{2\lambda+1}}) .$$

Démonstration :

Comme on estime en $x = 0$, on pourra considérer que les noyaux $K_{r(n)}$ sont fonctions d'une seule variable t . Alors :

$$R(\hat{h}_n, f) = \frac{1}{n} \left\{ \int K_{r(n)}^2(t) f(t) dt - \left[\int K_{r(n)}(t) f(t) dt \right]^2 \right\} + \left[\int K_{r(n)}(t) f(t) dt - f(0) \right]^2 .$$

$$\text{Soit alors } r(n) = n^{\frac{1}{1+2\lambda}} .$$

Comme $(K_{r(n)})$ est de type Fejer : $\|K_{r(n)}\|_\infty = O(r(n))$ et que $\int K_{r(n)}(t) f(t) dt \leq \|f\|_\infty$, le premier terme est majoré par $C \cdot \frac{r(n)}{n}$.

Nous aurons le résultat annoncé si l'on a de plus :

$$\left| \int K_{r(n)}(t) f(t) dt - f(0) \right| = O(r^{-\lambda}(n)) .$$

Soit alors :

$$\begin{aligned} \int K_{r(n)}(t) f(t) dt - f(0) &= \int K_{r(n)}(t) [f(t) - f(0)] dt \\ &= \left(\int_{-\infty}^{-\delta} + \int_{-\delta}^{-\frac{1}{r(n)}} + \int_{-\frac{1}{r(n)}}^{+\frac{1}{r(n)}} + \int_{\frac{1}{r(n)}}^{\delta} + \int_{\delta}^{\infty} \right) K_{r(n)}(t) [f(t) - f(0)] dt \end{aligned}$$

où δ est choisi tel que :

$$|f(x) - f(0)| \leq M \cdot |x|^\lambda \quad \text{pour} \quad |x| \leq \delta$$

Cela est possible car $f \in C_\lambda$.

Nous pouvons écrire pour le dernier terme :

$$\left| \int_\delta^\infty K_{r(n)}(t) (f(t) - f(0)) dt \right| \leq \int_\delta^\infty |K_{r(n)}(t)| |f(t) - f(0)| dt$$

$$\begin{aligned} \text{" " } &\leq \frac{C}{r(n)} \|f\|_1 + C \int_{r(n) \cdot \delta}^\infty \frac{dt}{1+t^2} |f(0)| \\ &= O(r^{-1}(n)) \end{aligned}$$

$$\text{car } K_{r(n)}(t) \leq \frac{C r(n)}{\pi [(r(n) \times t^2 + 1)]}$$

Pour le terme suivant et pour $r(n) \geq \frac{1}{\delta}$,

$$\int_\delta^\infty \frac{1}{r(n)} |K_{r(n)}(t)| |f(t) - f(0)| dt \leq \frac{C}{r(n)} \int_\delta^\infty \frac{1}{r(n)} \frac{|f(t) - f(0)|}{t^2} dt$$

$$\text{" " } \leq \frac{C}{r(n)} \cdot M \int_\delta^\infty \frac{1}{r(n)} \frac{t^\lambda}{t^2} dt$$

$$\begin{aligned} \text{" " } &\leq \frac{C \cdot M}{r(n)} \left[\frac{\delta^{\lambda-1}}{\lambda-1} - \frac{[r(n)]^{\lambda-1}}{\lambda-1} \right] \\ &= O(r^{-1}(n)) + O(r^{-\lambda}(n)) \end{aligned}$$

Le troisième terme se majore immédiatement car :

$$\begin{aligned} \int_{-\frac{1}{r(n)}}^{+\frac{1}{r(n)}} |K_{r(n)}(t)| |f(t) - f(0)| dt &\leq C \cdot r(n) M \int_{-\frac{1}{r(n)}}^{+\frac{1}{r(n)}} |t|^\lambda dt \\ &= O(r(n)^{-\lambda}) \end{aligned}$$

d'où la conclusion.

On peut alors citer :

Proposition. - La famille d'estimateurs \hat{h}_n correspondant à une suite $(K_{r(n)})$ de type Fejer est asymptotiquement minimax sur C_λ^U pour la moyenne quadratique.

Démonstration :

Le corollaire établi en II.1. prouve que :

$$\inf_{\hat{h}_n} \sup_{f \in C_\lambda^U} R(\hat{h}_n, f(0)) \geq D_0 n^{-\frac{2\lambda}{2\lambda+1}} .$$

D'autre part, le théorème de II.2. montre que la vitesse maximale est atteinte pour \hat{h}_n associé à une suite de type Fejer sur C_λ et $C_\lambda^U \subset C_\lambda$; d'où le résultat d'après la définition de l'estimateur minimax asymptotique de l'introduction.

(B) Efficacité des estimateurs dans le cas de la perte quadratique intégrée.

• Dans toute la suite du travail, on utilisera la fonction de perte $L_2(\lambda)$, λ désigne une mesure finie ou σ finie.

• La fonction de risque pour un estimateur h_n de f sera alors notée $R_2(\hat{h}_n, f)$ (Notation du chapitre I).

D. Bosq [8] a obtenu une inégalité de Cramer-Rao pour un paramètre hilbertien, ce qui nous permettra d'avoir une minoration du risque.

I - Inégalité de Cramer-Rao pour un paramètre hilbertien.

1. - Notations et principaux résultats de [8] (1).

• Le modèle statistique étant $[E, \mathcal{B}, (P_\theta)_{\theta \in \Theta}]$.

Soit : $\Theta \xrightarrow{\mathcal{E}} H$, H hilbert séparable.

• On désigne par B^1 l'espace des variables aléatoires définies sur E et à valeurs dans H qui sont intégrables au sens de

(1) également d'après : C.R. Acad. Sci. Paris. ser. A, t. 276, p. 1361-1364. X. Milhaud.

Bochner pour tout $\theta \in \Theta$.

. On note B^2 le sous-espace de B^1 formé des variables aléatoires Y telles que : $\int ||Y||^2 dP_\theta < \infty$ pour tout θ de Θ .

. Un estimateur T de $g(\theta)$ est dit sans biais s'il appartient à B^1 et si

$$\forall \theta \in \Theta, E_\theta(T) = \int T dP_\theta = g(\theta).$$

. Soit Y une variable aléatoire telle que $\int ||Y||^2 dP_\theta < \infty$, on appelle opérateur de covariance de Y pour P_θ l'unique opérateur D_Y symétrique sur H et tel que

$$\langle D_Y(y), y \rangle = E_\theta [\langle Y, y \rangle^2], \quad y \in H$$

où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire dans H et $||\cdot||$ la norme de H .

. Soient T et S deux estimateurs de $g(\theta)$, appartenant à B^2 . On dira que T est préférable à S si

$$\forall \theta \in \Theta \quad D_{T-g(\theta)}^\theta \preceq D_{S-g(\theta)}^\theta$$

pour la relation d'ordre usuelle entre opérateurs symétriques de H c'est-à-dire si

$$\forall \theta \in \Theta \quad \forall y \in H \quad \langle D_{T-g(\theta)}^\theta(y), y \rangle \leq \langle D_{S-g(\theta)}^\theta(y), y \rangle.$$

On peut citer alors :

Proposition 1. voir [8].-

Soient S et T deux estimateurs de $g(\theta)$, appartenant à B^2 . Si T est préférable à S , on a aussi :

$$E_\theta (||T-g(\theta)||^2) \leq E_\theta (||S-g(\theta)||^2), \quad \theta \in \Theta.$$

Soit alors $U = U_{\theta, \theta'}(X)$ une variable aléatoire réelle, définie P_θ presque sûrement (indépendamment de θ'), fonction de X et indexée par un sous-ensemble G de $\Theta \times \Theta$.

On suppose que U possède les propriétés suivantes :

$$(C_1) \quad E_\theta [U_{\theta, \theta'}(X)] = 0 \quad ; \quad (\theta, \theta') \in G$$

$$(C_2) \quad 0 < E_\theta [U_{\theta, \theta'}^2(X)] < +\infty \quad ; \quad (\theta, \theta') \in G$$

$$(C_3) \quad \text{Si } S \text{ et } T \text{ sont deux estimateurs } B^2 \text{ de } g(\theta)$$

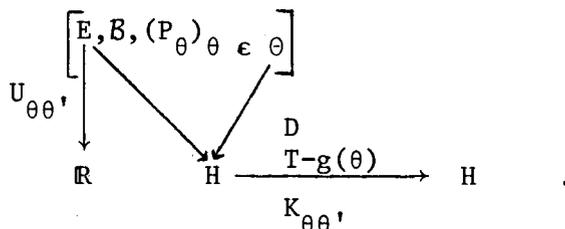
$$\text{tels que } m(\theta) = E_\theta(S) = E_\theta(T) \quad ; \quad \theta \in \Theta$$

$$\text{alors } E_\theta [U_{\theta, \theta'} \cdot T] = E_\theta [U_{\theta, \theta'} \cdot S] \quad (\theta, \theta') \in G$$

autrement dit :

$$E_\theta [U_{\theta, \theta'} \cdot T] = \psi_m(\theta, \theta') \quad (\theta, \theta') \in G .$$

On a le schéma suivant :



Dans ces conditions, on a le résultat suivant :

Proposition 2. [8]. -

Sous les hypothèses C_1, C_2, C_3 on a l'inégalité :

$$D_{T-g(\theta)} \supseteq K_{\theta, \theta'} \quad ; \quad (\theta, \theta') \in G$$

où $K_{\theta, \theta'}$ est l'opérateur sur H défini par

$$K_{\theta, \theta'}(y) = \frac{\langle \psi_m(\theta, \theta'), y \rangle}{E_\theta(U^2)} \cdot \psi_m(\theta, \theta') \quad ; \quad y \in H .$$

Soit la proposition suivante qui donne une condition nécessaire et suffisante pour que U vérifie C_3 .

Proposition 3. [8].-

Soit U vérifiant C_1 et C_2 . Une condition nécessaire pour que U vérifie C_3 est qu'elle soit de la forme $h_{\theta, \theta'}(S)$ où S est une statistique exhaustive. Si S est une statistique exhaustive complète, la condition est aussi suffisante.

Proposition 4. [8] (Inégalités du type Chapman-Robbins).

On suppose que $P_\theta = f_\theta \mu$, $\theta \in \Theta$, que l'application $\theta \rightarrow P_\theta$ est injective et que

$$(D_1) \quad \forall \theta \in \Theta, \quad M_\theta = \{\theta', \theta' \in \Theta - \{\theta\}, E_\theta \left[\left(\frac{f_{\theta'}}{f_\theta} \right)^2 \right] < +\infty\} \neq \emptyset.$$

Dans ces conditions si T est un estimateur de $g(\theta)$, élément de B^2 et tel que $E_\theta(T) = m(\theta)$, $\theta \in \Theta$, on a les inégalités suivantes :

$$D_{T-g(\theta)} \geq D_{g(\theta)-m(\theta)} + R_{\theta, \theta'}, \quad \theta \in \Theta \quad \theta' \in M_\theta$$

où $R_{\theta, \theta'}$ est l'opérateur défini sur H par

$$R_{\theta, \theta'}(y) = \frac{\langle m(\theta') - m(\theta), y \rangle}{E_\theta \left[\left(\frac{f_{\theta'}}{f_\theta} - 1 \right)^2 \right]} \cdot [m(\theta') - m(\theta)] \quad y \in H$$

et

$$E_\theta (\|T-g(\theta)\|^2) \geq \|g(\theta)-m(\theta)\|^2 + \sup_{\theta' \in M_\theta} \frac{ \|m(\theta') - m(\theta)\|^2 }{ E_\theta \left[\left(\frac{f_{\theta'}}{f_\theta} - 1 \right)^2 \right]}$$

$$\theta \in \Theta \quad \theta' \in M_\theta$$

2. - Application à l'estimation de la fonction de répartition.

Soit g l'application définie par $g(\theta) = F_\theta(\cdot)$ $\theta \in \Theta$

et $F_\theta(x) = \int_{-\infty}^x f(\theta, t) dt.$

On pose $G_\theta = \{g(\theta), \theta \in \Theta\}$

a. On peut alors énoncer :

Proposition.-

Soit λ une mesure finie telle que $G_\theta \subset L^2(\lambda)$.

Si la fonction de perte est $L_2(\lambda)$ et si les hypothèses des propositions 2 et 4 (chap. II.B.I.1.) sont vérifiées, alors \hat{F}_n désignant un estimateur de F_θ , et élément de B^2 , on a :

$$R_2(\hat{F}_n, F_\theta) \geq L_2(F_\theta, m(\theta)) + \sup_{\theta' \in M_\theta} \frac{L_2(m(\theta), m(\theta'))}{E_\theta(U_{\theta, \theta'}^2)} .$$

Démonstration :

Il suffit de remarquer que la fonction de perte $L_2(\lambda)$ est telle que :

$$\begin{aligned} L_2(m(\theta), m(\theta')) &= \int [m(\theta) - m(\theta')]^2 d\lambda \\ &= ||m(\theta) - m(\theta')||^2 \end{aligned}$$

et que tout élément de G_θ est borné, donc élément de l'espace de hilbert $L^2(\lambda)$.

Corollaire 1.-

F_n désignant la fonction de répartition empirique, on a :

$$R_2(F_n, F_\theta) \geq \sup_{\theta' \in M_\theta} \frac{L_2(F_\theta, F_{\theta'})}{E_\theta(U_{\theta, \theta'}^2)}$$

N.B. Un second corollaire sera donné au c.

Démonstration :

F_n est un estimateur sans biais de F_θ ; en effet, d'après [7], $\langle \dots \rangle$ désignant le produit scalaire de $L^2(\lambda)$, il suffit de montrer que $\forall \ell \in L^2(\lambda) : \langle F_n, \ell \rangle = \int \ell dF_n = \int \ell dF_\theta$ (car $F_n \in B^2 : ||F_n||^2 \leq 1$)

$$\langle \ell, \int_{E^n} F_n(x_1, \dots, x_n) dF_\theta(x_1) \dots dF_\theta(x_n) \rangle = \langle \ell, F_\theta \rangle .$$

$$\begin{aligned} \text{Or : } & \langle \ell, \int_{E^n} F_n(x_1, \dots, x_n) dF_\theta(x_1) \dots dF_\theta(x_n) \rangle \\ &= \int_{E^n} \langle \ell, F_n(x_1, \dots, x_n) dF_\theta(x_1) \dots dF_\theta(x_n) \rangle = \\ &= \int_{E^n} \int_E \ell(x) \cdot F_n(x_1, \dots, x_n)(x) d\lambda(x) dF_\theta(x_1) \dots dF_\theta(x_n) \end{aligned}$$

et d'après le théorème de Fubini, cette expression est encore égale à :

$$= \int_E \ell(x) \left[\int_{E^n} F_n(x_1, \dots, x_n)(x) dF_\theta(x_1) \dots dF_\theta(x_n) \right] d\lambda(x)$$

$$\text{et comme } \int_{E^n} F_n(x_1, \dots, x_n)(x) dF_\theta(x_1) \dots dF_\theta(x_n) = F_\theta(x),$$

on a le résultat.

Alors, $g(\theta) = m(\theta) = F_\theta$ (notations des propositions 2 et 4)

d'où la conclusion.

b. Nous allons donner 2 exemples illustrant le théorème et son corollaire, en utilisant deux estimateurs :

F_n estimateur sans biais

F_2^* estimateur minimax pour la fonction de perte $L_2(\lambda)$.

Exemple n° 1. - Estimation de la fonction de répartition

d'une loi gamma avec l'estimateur F_n

$$\text{pour } \left\{ \begin{array}{l} F_\theta(x) = 1 - e^{-\theta x} \\ f_\theta(t) = \theta e^{-\theta t} \end{array} \right. \quad \left| \begin{array}{l} E = \mathbb{R}^+ \\ \theta = \mathbb{R}^{**} \end{array} \right.$$

• λ désignera ici la mesure de probabilité de densité $f_1(t) = e^{-t}$.

• Soit $T_n = \sum_{i=1}^n X_i$, qui constitue une statistique exhaustive et complète. Soit alors :

$$U_{\theta, \theta'}(X_1 \dots X_n) = \left(\frac{\lambda'}{\lambda}\right)^n e^{(\lambda - \lambda') \cdot T_n - 1}$$

- $U_{\theta, \theta'}$ satisfait bien aux conditions C_1, C_2, C_3 , d'après la proposition 3.
- L'hypothèse $\theta \rightarrow P_\theta$ injective est vérifiée et la condition (D_1) de la proposition 4 du paragraphe 1. est satisfaite et on a même

$$M_\theta = \theta - \{\theta/2\} = \mathbb{R}^{*+} - \{\theta/2\} \neq \emptyset.$$

- Nous pouvons alors calculer les deux membres de l'inégalité de la proposition précédente.

• Calcul de $R_2(F_n, F_\theta)$.

$$\begin{aligned} R_2(F_n, F_\theta) &= \int_{(\mathbb{R}^+)^n} L_2[F_n(x_1, \dots, x_n)(x), F_\theta(x)] dF_\theta(x_1) \dots dF_\theta(x_n) \\ &= \int_{(\mathbb{R}^+)^n} \int_{\mathbb{R}^+} [F_n(x_1, \dots, x_n)(x) - F_\theta(x)]^2 dF_1(x) dF_\theta(x_1) \dots dF_\theta(x_n) \\ &= \int_{\mathbb{R}^+} \left[\int_{(\mathbb{R}^+)^n} [F_n(x_1, \dots, x_n)(x) - F_\theta(x)]^2 dF_\theta(x_1) \dots dF_\theta(x_n) \right] dF_1(x) \end{aligned}$$

en appliquant le théorème de Fubini

$$\text{et : } R_2(F_n, F_\theta) = \frac{1}{n} \int_{\mathbb{R}^+} F_\theta(x) [1 - F_\theta(x)] dF_1(x)$$

et dans le cas de la loi gamma :

$$\underline{R_2(F_n, F_\theta)} = \frac{1}{n} \int_0^\infty (e^{-\theta t} - e^{-2\theta t}) e^{-t} dt = \frac{1}{n} \left[\frac{1}{\theta+1} - \frac{1}{2\theta+1} \right]$$

• Calcul de $\sup_{\theta' \in M_\theta} \frac{L_2(F_{\theta'}, F_\theta)}{E_\theta[U_{\theta, \theta'}^2]}$

$$\cdot L_2(F_{\theta'}, F_\theta) = \int_0^\infty [F_{\theta'}(x) - F_\theta(x)]^2 dF_1(x)$$

$$'' = \int_0^{\infty} (e^{-\theta x} - e^{-\theta' x})^2 e^{-x} dx$$

$$'' = \int_0^{\infty} (e^{-2\theta x} - 2e^{-x(\theta+\theta')} + e^{-2\theta' x}) e^{-x} dx = \boxed{\frac{1}{2\theta+1} - \frac{2}{\theta+\theta'+1} + \frac{1}{2\theta'+1}}$$

. d'autre part, si $T_n = \sum_{i=1}^n X_i$, on obtient à l'aide du théorème d'inversion, que T_n admet comme densité

$$r_n(x) = \frac{\theta^n x^{n-1} e^{-\theta x}}{(n-1)!} .$$

Cela permet d'obtenir que

$$E_{\theta} [U_{\theta, \theta'}^2] = \int_0^{\infty} \left[\left(\frac{\theta'}{\theta} \right)^n e^{-(\theta' - \theta)t} \right]^2 \left[\frac{\theta^n t^{n-1} e^{-t}}{(n-1)!} \right] dt$$

$$'' = \boxed{\frac{\theta'^{2n} - \theta^n (2\theta' - \theta)^n}{\theta^n (2\theta' - \theta)^n}} .$$

$\frac{L_2(F_{\theta'}, F_{\theta})}{E_{\theta} [U_{\theta, \theta'}^2]}$ apparaît comme une fonction de θ' qui atteint

son maximum lorsque $\theta' \rightarrow \theta$ (on pose $\theta' = \theta + \frac{1}{p}$, $p \rightarrow \infty$).

$$\text{Alors } \sup_{\theta' \in M_{\theta}} \frac{L_2(F_{\theta'}, F_{\theta})}{E_{\theta} [U_{\theta, \theta'}^2]} = \frac{\theta}{2n(2\theta+1)^2}$$

(obtenu à l'aide d'arguments simples sur les limites de fonctions rationnelles).

. L'inégalité de Cramer-Rao s'écrit alors

$$\left| \forall \theta \in \mathbb{R}^{**} \quad \frac{\theta}{n(2\theta+1)(\theta+1)} \geq \frac{\theta}{2n(2\theta+1)^2} \right.$$

ce qui nous donne une efficacité égale à

$$\boxed{e = \frac{\theta+1}{2(2\theta+1)}} ,$$

Remarque : e est maximale au voisinage de 0.

Exemple n° 2.- Même situation que précédemment.

On utilise cette fois l'estimateur \hat{F}_2^* Minimax, déterminé au chapitre I. B.I.1. par :

$$\hat{F}_2^*(t) = \frac{1}{2(n^{1/2}+1)} + \frac{n^{1/2}}{n^{1/2}+1} F_n(t)$$

Il est immédiat que \hat{F}_2^* est un estimateur B^2 (car F_n en est un). La nature de \hat{F}_2^* est précisée dans le lemme suivant :



Lemme.- $\lim_{n \rightarrow \infty} \hat{F}_2^*(x) = F_n(x)$.

En effet :

$$\hat{F}_2^*(x) - F_n(x) = \frac{1}{2(n^{1/2}+1)} - \frac{F_n(x)}{n^{1/2}+1} \xrightarrow{n \rightarrow \infty} 0$$

car $F_n(x)$ est bornée pour tout n .

Le résultat est donc analogue à celui de Kronmal [36]

$$\lim_{n \rightarrow \infty} \hat{F}_n^\perp(x) = F_n(x)$$

\hat{F}_n^\perp est l'estimateur pour la méthode par projection et sera étudié au b.

. Calcul de $R_2(\hat{F}_2^*, F_\theta)$.

Ce calcul a été effectué au B. I.1. et on avait trouvé un risque constant sur G_θ

$$R_2(\hat{F}_2^*, F_\theta) = \frac{1}{4(n^{1/2}+1)^2} \int d\lambda(t)$$

$$" = \frac{1}{4(n^{1/2}+1)} \quad \text{car } \lambda \text{ est ici la mesure de probabilité de densité } f_1(t) = e^{-t}$$

• Calcul de $L_2[F_{\theta}, m(\theta)] + \sup_{\theta' \in M_{\theta}} \frac{L_2[m(\theta), m(\theta')]}{E_{\theta}[U_{\theta, \theta'}^2]}$.

Le premier terme est égal à :

$$\begin{aligned} L_2[F_{\theta}, m(\theta)] &= L_2\left[F_{\theta}, \frac{1}{2(n^{1/2}+1)} + \frac{n^{1/2}}{n^{1/2}+1} F_{\theta}\right] \\ &= \int_0^{\infty} \left[-\frac{1}{2(n^{1/2}+1)} + \frac{1}{n^{1/2}+1} F_{\theta}(t) \right]^2 dF_1(t) \\ &= \frac{1}{4(n^{1/2}+1)^2} - \frac{1}{(n^{1/2}+1)^2} \int_0^{\infty} (1-e^{-\theta t}) \cdot e^{-t} dt + \dots/\dots \\ &\dots/\dots + \frac{1}{(n^{1/2}+1)^2} \int_0^{\infty} [1 - 2e^{-t} + e^{-2\theta t}] e^{-t} dt \text{ et, après calculs :} \\ &= \frac{1}{4(n^{1/2}+1)^2} \left[1 - \frac{4\theta}{(\theta+1)(2\theta+1)} \right] \end{aligned}$$

d'autre part :

$$\begin{aligned} L_2[m(\theta), m(\theta')] &= \int_0^{\infty} \left[\frac{n^{1/2}}{n^{1/2}+1} \right]^2 [F_{\theta}(t) - F_{\theta'}(t)]^2 dF_1(t) \\ &= \frac{n}{(n^{1/2}+1)^2} \int_0^{\infty} [F_{\theta}(t) - F_{\theta'}(t)]^2 dF_1(t) \\ \text{et } \sup_{\theta' \in M_{\theta}} \frac{\int_0^{\infty} [F_{\theta}(t) - F_{\theta'}(t)]^2 dF_1(t)}{E_{\theta}(U_{\theta, \theta'}^2)} &= \frac{\theta}{2n(2\theta+1)^2} \end{aligned}$$

d'après le calcul effectué dans l'exemple n°1.

Alors :

$$\begin{aligned} L^2[F_{\theta}, m(\theta)] + \sup_{\theta' \in M_{\theta}} \frac{L_2[m(\theta), m(\theta')]}{E_{\theta}(U_{\theta, \theta'}^2)} &= \dots/\dots \\ \dots/\dots &= \frac{1}{4(n^{1/2}+1)^2} \left[1 - \frac{4\theta}{(\theta+1)(2\theta+1)} \right] + \frac{2\theta}{4(n^{1/2}+1)^2(2\theta+1)^2} \end{aligned}$$

. L'inégalité de Cramer-Rao s'écrit alors :

$$\frac{1}{4(n^{1/2}+1)^2} \geq \frac{1}{4(n^{1/2}+1)^2} \left[1 - \frac{2\theta(3\theta+1)}{(\theta+1)(2\theta+1)^2} \right]$$

ce qui nous donne une efficacité égale à e'

$$e' = 1 - \frac{2\theta(3\theta+1)}{(\theta+1)(2\theta+1)^2} .$$

Comparaison des efficacités entre F_n et \hat{F}_2^* .

$e' > e$ si

$$1 - \frac{4\theta(3\theta+1)}{2(\theta+1)(2\theta+1)^2} > \frac{(\theta+1)^2(2\theta+1)}{2(2\theta+1)^2(\theta+1)}$$

$$6\theta^3 - \theta^2 + 2\theta + 1 > 0 \quad \text{ce qui est toujours vrai } \theta \in \mathbb{R}^{**+} .$$

Alors | L'estimateur F_2^* Minimax est plus efficace que F_n
pour G_θ .

c. Un troisième estimateur de la fonction de répartition est celui obtenu à l'aide de la méthode par projection.

Notons \hat{F}_n^\perp cet estimateur.

Soit (e_j) une base de l'espace de hilbert séparable $L^2(\lambda)$, λ mesure finie.

Alors d'après Bosq [6] et Kronmal Tarter [36]

$$\hat{F}_n^\perp(t) = \sum_{j=1}^{q(n)} \hat{b}_j e_j(t) \quad \text{avec} \quad \left| \begin{array}{l} \hat{b}_j = \int e_j(t) F_n(t) d\lambda(t) \\ E_\theta(\hat{b}_j) = b_j^\theta \end{array} \right.$$

et $F_\theta(t) = \sum_{j=1}^{\infty} b_j^\theta e_j(t)$

Corollaire 2.- Avec les notations ci-dessus, on obtient une deuxième minoration du risque.

$$R_2(F_n, F) \geq \sum_{j>q(n)} \text{var } \hat{b}_j + \sup_{\theta' \in M_\theta} \sum_{j=1}^{q(n)} \frac{(b_j^\theta - b_j^{\theta'})^2}{E_\theta(U_{\theta, \theta'}^2)}$$

Démonstration :

$$\|\hat{F}_n^\perp\|^2 = \sum_{j=1}^{q(n)} (\hat{b}_j)^2 = \sum_{j=1}^{q(n)} \left[\int e_j(t) F_n(t) d\lambda(t) \right]^2$$

$$\leq \sum_{j=1}^{q(n)} \int e_j^2(t) d\lambda(t) \cdot \int F_n^2(t) d\lambda(t)$$

$$\leq q(n) \cdot \|F_n\|^2 \leq q(n).$$

\hat{F}_n^\perp est alors un estimateur de B^2 .

Nous pouvons donc appliquer la proposition précédente,

$$\text{d'où } R_2(\hat{F}_n^\perp, F_\theta) = \sum_{j=1}^{q(n)} \text{var } \hat{b}_j + \sum_{j>q(n)} b_j^2$$

$$\text{et } R_2(F_n, F_\theta) = \sum_{j=1}^{\infty} \text{var } \hat{b}_j \text{ d'où}$$

$$R_2(\hat{F}_n^\perp, F_\theta) = R_2(F_n, F_\theta) + \sum_{j>q(n)} [b_j^2 - \text{var } \hat{b}_j]$$

$$= R_2(F_n, F_\theta) + \sum_{j>q(n)} [2b_j^2 - E(\hat{b}_j^2)].$$

Remarque : d'après [36], avec un "bon choix" de $q(n)$:

$$R_2(\hat{F}_n^\perp, F_\theta) > R_2(F_n, F_\theta).$$

D'autre part :

$$L_2[\underline{F}_\theta, m(\theta)] = \int \left[\sum_{j>q(n)} b_j^\theta e_j(t) \right]^2 d\lambda(t)$$

$$\text{car : } \begin{cases} m(\theta) = E_\theta(\hat{F}_n) = \sum_{j=1}^{q(n)} b_j^\theta e_j(t) \text{ et} \\ F_\theta(t) = \sum_{j=1}^{\infty} b_j^\theta e_j(t) \end{cases}$$

$$L_2(F_\theta, m(\theta)) = \sum_{j>q(n)} (b_j^\theta)^2$$

et

$$L_2(m(\theta), m(\theta')) = \sum_{j=1}^{q(n)} (b_j^\theta - b_j^{\theta'})^2 \quad \text{d'où}$$

$$R_2(F_n, F_\theta) + \sum_{j>q(n)} [2(b_j^\theta)^2 - E_\theta(\hat{b}_j)^2] \geq \sum_{j>q(n)} (b_j^\theta)^2 + \sum_{j=1}^{q(n)} \frac{(b_j^\theta - b_j^{\theta'})^2}{E_\theta[U_{\theta, \theta'}^2]}$$

en transposant le terme :

$$\sum_{j>q(n)} [2(b_j^\theta)^2 - E_\theta(\hat{b}_j)^2] \quad \text{on a le résultat :}$$

Remarque :

L'inégalité est d'autant plus intéressante que l'expression

$$\sum_{j>q(n)} (\text{var } \hat{b}_j) + \sup_{\theta' \in M_\theta} \sum_{j=1}^{q(n)} \frac{(b_j^\theta - b_j^{\theta'})^2}{E_\theta(U_{\theta, \theta'}^2)} \quad \text{est grande.}$$

Par exemple, dans le cas λ mesure de Lebesgue sur $[-\pi, +\pi]$ et (e_k) le système trigonométrique orthonormal de $L^2[-\pi, +\pi]$, on peut utiliser [36] qui exprime $\text{var } \hat{b}_j$ en fonction de (b_k) .

Il suffit alors d'évaluer les coefficients de Fourier (b_k) .

3.- Application à l'estimation de la loi.

Soit $[E, B, (P_\theta) \quad \theta \in \Theta]$.

. On prendra $E = [0, 1]$ espace séparable, muni de la métrique de la valeur absolue.

- . $(P_\theta)_{\theta \in \Theta} \subset P$ est muni de la métrique de Prokhorov.
- . P désigne l'ensemble des mesures de probabilité.
- . M^* " " " " bornées positives.
- . M " " " " bornées à signes.
- . K désigne le noyau, tel que

$$\begin{aligned} & \cdot \mu \in M \quad \left| \quad \langle \mu, \nu \rangle_K = \int_{[0,1]^2} K(x,y) d\mu \otimes \nu(x,y) \right. \\ & \nu \in M \quad \left| \quad \text{est un produit scalaire sur } M. \right. \end{aligned}$$

• $\| \cdot \|_K$ la norme associée telle que $[M, \| \cdot \|_K]$ est un espace préhilbertien séparable.

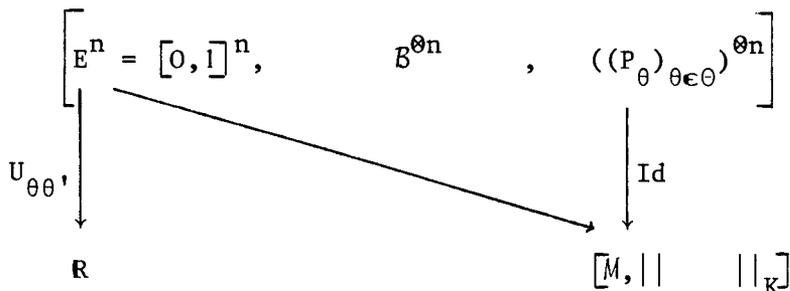
• On prendra ici $K(x,y) = \sum_{i=0}^{\infty} \left(\frac{x}{2}\right)^i \left(\frac{y}{2}\right)^i$

on a alors :

$$(\mu, \nu)_K = \sum_{i=0}^{\infty} \int_{[0,1]} \left(\frac{x}{2}\right)^i d\mu(x) \int_{[0,1]} \left(\frac{y}{2}\right)^i d\nu(y)$$

et $\| \mu \|_K^2 = \sum_{i=0}^{\infty} \left[\int_{[0,1]} \left(\frac{x}{2}\right)^i d\mu(x) \right]^2$.

Le schéma est alors le suivant :



On utilise l'estimateur T_n défini par

$$B \in \mathcal{B} \quad T_n(x_1, \dots, x_n)(B) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \in B\}}.$$

On peut alors énoncer :

Proposition.-

La fonction de perte étant ici L_K , définie par :

$$L_K(\mu, \nu) = \|\mu - \nu\|_K^2,$$

le risque associé se notant R_K , on a alors :

$$R_K(T_n, P_\theta) \geq \sup_{\theta' \in M_\theta} \frac{L_K(m(\theta), m(\theta'))}{E_\theta(U_{\theta, \theta'}^2)}$$

avec $m(\theta) = E_\theta(T_n)$

et les hypothèses des propositions 2 et 4 du paragraphe 1 qui sont vérifiées.

Démonstration :

$T_n \in B^2$. En effet :

Calcul de : $\|T_n(x_1, \dots, x_n)\|_K^2$

$$\begin{aligned} \|T_n(x_1, \dots, x_n)\|_K^2 &= \sum_{i=0}^{\infty} \left[\int_{[0,1]} \left(\frac{x}{2}\right)^i dT_n(x_1, \dots, x_n)(x) \right]^2 \quad \text{avec } \delta_{x_j}(x) = 1_{\{x_j=x\}} \\ &= \sum_{i=0}^{\infty} \left[\int_{[0,1]} \left(\frac{x}{2}\right)^i \frac{1}{n} \sum_{j=1}^n d\delta_{x_j}(x) \right]^2 \\ &= \sum_{i=0}^{\infty} \left[\frac{1}{n} \sum_{j=1}^n \int_{[0,1]} \left(\frac{x}{2}\right)^i d\delta_{x_j}(x) \right]^2 \\ &= \sum_{i=0}^{\infty} \left[\frac{1}{n} \sum_{j=1}^n \int_{[0,1]} \left(\frac{x}{2}\right)^i d[1_{\{x_j=x\}}] \right]^2 \\ &= \sum_{i=0}^{\infty} \left[\frac{1}{n} \sum_{j=1}^n \left(\frac{x_j}{2}\right)^i \right]^2, \quad \text{qui est une série convergente} \end{aligned}$$

car :

$$0 \leq x_j \leq 1 \quad \text{et} \quad 0 \leq \left(\frac{x_j}{2}\right)^i \leq \frac{1}{2^i}$$

$$\frac{1}{n} \sum_{j=1}^n \left(\frac{x_j}{2}\right)^i \leq \frac{1}{2^i} \quad \text{et} \quad \sum_{i=0}^{\infty} \frac{1}{2^i} \quad \text{est bien convergente}$$

et alors $T_n \in B^2$ car $\|T_n\|_K^2 \leq 2$.

T_n est sans biais. En effet :

Pour montrer que :

$$\int_{E^n} T_n(x_1, \dots, x_n) dP_\theta(x_1) \dots dP_\theta(x_n) = P_\theta, \text{ il suffit}$$

d'établir que :

$$\forall \mu \in M \quad \langle \mu, \int_{E^n} T_n(x_1, \dots, x_n) dP_\theta(x_1) \dots dP_\theta(x_n) \rangle_K = \langle \mu, P_\theta \rangle_K$$

car : T_n est élément de B^2 .

$$\text{On a : } \langle \mu, \int_{E^n} T_n(x_1, \dots, x_n) dP_\theta(x_1) \dots dP_\theta(x_n) \rangle_K$$

$$= \int_{E^n} \langle \mu, T_n(x_1, \dots, x_n) \rangle_K dP_\theta(x_1) \dots dP_\theta(x_n)$$

$$= \int_{E^n} \left\{ \int_{[0,1]^2} K(x,y) d(\mu \otimes T_n(x_1, \dots, x_n))(x,y) \right\} dP_\theta(x_1) \dots dP_\theta(x_n)$$

d'après le théorème de Fubini.

$$= \int_{[0,1]^2} K(x,y) d\mu(x) \int_{E^n} dT_n(x_1, \dots, x_n)(y) dP_\theta(x_1) \dots dP_\theta(x_n).$$

$$\text{Comme } \int_{E^n} T_n(x_1, \dots, x_n)(B) dP_\theta(x_1) \dots dP_\theta(x_n)$$

$$= \int_{E^n} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(B) dP_\theta(x_1) \dots dP_\theta(x_n)$$

$$= \frac{1}{n} \sum_{i=1}^n \int_{E^n} 1_{\{x_i \in B\}} dP_\theta(x_1) \dots dP_\theta(x_n)$$

$$= \int_E 1_{\{x \in B\}} dP_\theta(x) = \int_B dP_\theta(x)$$

$$= P_\theta(B).$$

$$\text{Alors : } \int_{E^n} \frac{1}{n} \sum_{i=1}^n d \delta_{x_i}(x) dP_\theta(x_1) \dots dP_\theta(x_n) = dP_\theta(x).$$

On obtient :

$$\int_{E^n} dT_n(x_1, \dots, x_n)(y) dP_\theta(x_1) \dots dP_\theta(x_n) = dP_\theta(y)$$

d'où :

$$\begin{aligned} & \langle \mu, \int T_n(x_1, \dots, x_n) dP_\theta(x_1) \dots dP_\theta(x_n) \rangle_K \\ &= \int_{[0,1]^2} K(x,y) d\mu(x) dP_\theta(y) = \langle \mu, P_\theta \rangle_K \end{aligned}$$

et T_n est bien sans biais.

Alors la proposition se déduit de la proposition 4.

Exemple : Soit la famille $(P_\theta)_{\theta \in \Theta}$ de probabilité sur $E = [0,1]$, de densité par rapport à la mesure de Lebesgue définie par :

$$f_\theta(x) = (\theta + 1)x^\theta$$

. Calcul de $R_K(T_n, P_\theta)$

$$R_K(T_n, P_\theta) = \int ||T_n - P_\theta||_K^2 dP_\theta^{\otimes n}$$

$$||T_n - P_\theta||_K^2 = \sum_{i=0}^{\infty} \left[\int_{[0,1]} \left(\frac{x}{2}\right)^i \{dT_n(x) - dP_\theta(x)\} \right]^2$$

$$" = \sum_{i=0}^{\infty} \left[\int_{[0,1]} \frac{1}{n} \sum_{j=1}^n \left(\frac{x}{2}\right)^i d\delta_{x_j}(x) - \int_{[0,1]} \left(\frac{x}{2}\right)^i (\theta+1)x^\theta dx \right]^2$$

$$= \sum_{i=0}^{\infty} \left[\frac{1}{n} \sum_{j=1}^n \left(\frac{x_j}{2}\right)^i - \frac{\theta+1}{2^i(\theta+i+1)} \right]^2$$

$$" = \sum_{i=0}^{\infty} \left\{ \frac{1}{n} \sum_{j=1}^n \left(\frac{x_j}{2}\right)^i \right\}^2 + \sum_{i=0}^{\infty} \frac{(\theta+1)^2}{2^{2i}(\theta+i+1)^2} - \frac{2(\theta+1)}{n} \sum_{i=0}^{\infty} \left\{ \frac{1}{2^{2i}(\theta+i+1)} \sum_{j=1}^n (x_j)^i \right\}.$$

On obtient 3 séries convergentes.

$R_K(T_n, P_\theta)$ se compose donc de trois termes.

□ 1er terme $\int_{E^n} \sum_{i=0}^{\infty} \left\{ \frac{1}{n} \sum_{j=1}^n \left(\frac{x_j}{2}\right)^i \right\}^2 dP_\theta(x_1) \dots dP_\theta(x_n)$

$$= \int_{E^n} \sum_{i=0}^{\infty} \frac{1}{2^{2i} n^2} \underbrace{\{(x_1)^{2i} + \dots + (x_n)^{2i}\}}_{(1)} + \underbrace{\{(\dots x_j^i x_\ell^i \dots)\}}_{(2) \quad j \neq \ell} dP_\theta(x_1) \dots dP_\theta(x_n).$$

On obtient deux intégrales : celle qui correspond aux termes (1) et celle qui correspond aux termes (2).

Soit : termes (1) $(x_1)^{2i} + \dots + (x_n)^{2i}$
 termes (2) $(\dots x_j^i x_\ell^i \dots)$, ce qui donne :

$$\begin{aligned} \text{1ère intégrale} &= \sum_{i=0}^{\infty} \frac{1}{2^{2i} n^2} \int_{[0,1]} x^{2i} dP_\theta(x) \\ &= \sum_{i=0}^{\infty} \frac{\theta+1}{2^{2i} n} \int_{[0,1]} x^{\theta+2i} dx = \sum_{i=0}^{\infty} \frac{\theta+1}{n 2^{2i}(\theta+2i+1)} \end{aligned}$$

$$\begin{aligned} \text{2ème intégrale} &= \sum_{i=0}^{\infty} \frac{n(n-1)}{2^{2i} n^2} \int_{[0,1]} x_1^i x_2^i dP_\theta(x_1) dP_\theta(x_2) \\ &= \sum_{i=0}^{\infty} \frac{n(n-1)}{2^{2i} n^2} \left\{ \int_{[0,1]} x^{i(\theta+1)} x^\theta d\theta \right\}^2 \\ &= \frac{n-1}{n} \sum_{i=1}^{\infty} \frac{(\theta+1)^2}{2^{2i}(\theta+i+1)^2} \end{aligned}$$

Le premier terme vaut donc :

$$\frac{1}{n} \sum_{i=0}^{\infty} \frac{\theta+1}{2^{2i}(\theta+2i+1)} + \frac{n-1}{n} \sum_{i=0}^{\infty} \left[\frac{\theta+1}{2^i(\theta+i+1)} \right]^2$$

□ 2ème terme

$$(\theta+1)^2 \int_{E^n} \sum_{i=0}^{\infty} \frac{1}{2^{2i}(\theta+i+1)^2} dP_\theta(x_1) \dots dP_\theta(x_n) = \sum_{i=0}^n \left[\frac{\theta+1}{2^i(\theta+i+1)} \right]^2$$

□ 3ème terme

$$- \frac{2(\theta+1)}{n} \int_{E^n} \sum_{i=0}^{\infty} \left\{ \frac{1}{2^{2i}(\theta+i+1)} \sum_{j=1}^n (x_j)^i \right\} dP_\theta(x_1) \dots dP_\theta(x_n)$$

$$= - \frac{2(\theta+1)}{n} \sum_{i=0}^{\infty} \frac{1}{2^{2i}(\theta+i+1)} \underbrace{\sum_{j=1}^n \int_{E^n} (x_j)^i dP_{\theta}(x_1) \dots dP_{\theta}(x_n)}_{(a)} .$$

$$\begin{aligned} \text{Or } (a) &= n \int_{[0,1]} x^i dP_{\theta}(x) = n \int_{[0,1]} x^{\theta+i}(\theta+1)dx. \\ &= \frac{n(\theta+1)}{\theta+i+1} \end{aligned}$$

et on obtient pour le troisième terme :

$$- 2 \sum_{i=0}^{\infty} \frac{(\theta+1)^2}{2^{2i}(\theta+i+1)^2}$$

donc

$$R_K(T_n, P_{\theta}) = \frac{1}{n} \left\{ (\theta+1) \sum_{i=0}^{\infty} \frac{1}{2^{2i}(\theta+2i+1)} - (\theta+1)^2 \sum_{i=0}^{\infty} \frac{1}{2^{2i}(\theta+i+1)^2} \right\}$$

. Calcul de $||m(\theta) - m(\theta')||_K^2 = L_K(m(\theta), m(\theta'))$

avec $m(\theta) = E_{\theta}(T_n) = P_{\theta}$.

$$\begin{aligned} L_K(m(\theta), m(\theta')) &= \sum_{i=0}^{\infty} \left[\int_{[0,1]} \left(\frac{x}{2}\right)^i d(P_{\theta'} - P_{\theta})(x) \right]^2 \\ &= \sum_{i=0}^{\infty} \left[\int_{[0,1]} \left(\frac{x}{2}\right)^i [(\theta'+1)x^{\theta'} - (\theta+1)x^{\theta}] dx \right]^2 \\ &= \sum_{i=0}^{\infty} \left[\frac{\theta'+1}{2^i(\theta'+i+1)} - \frac{\theta+1}{2^i(\theta+i+1)} \right]^2 \\ &= \sum_{i=0}^{\infty} \left[\frac{i(\theta' - \theta)}{2^i(\theta+i+1)(\theta'+i+1)} \right]^2 . \end{aligned}$$

. Calcul de $E_{\theta}(U_{\theta, \theta'}^2)$.

Comme dans les exemples (1) et (2) de b., avec :

$$U_{\theta, \theta'}(X_1 \dots X_n) = \left(\frac{\theta' + 1}{\theta + 1}\right)^n (X_1 \dots X_n)^{\theta' - \theta} - 1$$

on obtient :

$$E_{\theta}(U_{\theta, \theta'}^2) = \frac{(\theta' + 1)^{2n}}{(\theta + 1)^n (2\theta' - \theta + 1)^n} - 1.$$

L'inégalité de Cramer-Rao s'écrit alors :

$$\frac{1}{n} \sum_{i=0}^{\infty} \frac{i^2 (\theta + 1)}{2^{2i} (\theta + 2i + 1) (\theta + i + 1)^2} \geq \sup_{\theta' \in M_{\theta}} \frac{\sum_{i=0}^{\infty} \left[\frac{i(\theta' - \theta)}{2^i (\theta + i + 1) (\theta' + i + 1)} \right]^2}{\frac{(\theta' + 1)^{2n}}{(\theta + 1)^n (2\theta' - \theta + 1)^n} - 1}$$

Le second terme de l'inégalité atteint son maximum pour $\theta' \rightarrow \theta$. En posant $\theta' = \theta + \frac{1}{p}$, $p \rightarrow \infty$, on obtient après calculs

l'inégalité suivante ; entre séries convergentes :

$$\boxed{\frac{1}{n} \sum_{i=0}^{\infty} \frac{i^2 (\theta + 1)}{2^{2i} (\theta + 2i + 1) (\theta + i + 1)^2} \geq \frac{1}{2n} \sum_{i=0}^{\infty} \left[\frac{i(\theta + 1)}{2^i (\theta + i + 1)^2} \right]^2}$$

On peut d'ailleurs montrer que chaque terme de la première série majore chaque terme de la seconde série.

4. - Complément sur l'inégalité de Cramer-Rao, dans le cas de l'estimateur \hat{h}_n de la densité.

On considère le modèle statistique $[E, B, (P_f) f \in \tilde{F}]$ avec $E = [0, 1]$.

. Soit $f \in F$ telle que : $\inf_{t \in [0, 1]} f(t) > 0$.

On considère alors :

. $L^2(\frac{1}{f})$ l'espace de hilbert associé de norme $\| \cdot \|_{\frac{1}{f}}$.

. Alors si $g \in L^2(\frac{1}{f})$, on a : $\|g\|_{\frac{1}{f}}^2 = \int_0^1 \frac{g^2(t)}{f(t)} dt < \infty$.

La condition (D₁) de la proposition 4 (B.I.1.) est donc vérifiée. En prenant les notations de la proposition 4, on peut alors citer :

Proposition D. Bosq.-

La fonction de perte étant $L_2(\frac{1}{f})$

$$R_2(\hat{h}_n, f) \geq L_2[f, m(f)] + \sup_{\substack{g \neq f \\ g \in \mathcal{D}}} \frac{L_2[m(g), m(f)]}{\left[\|g\|_{\frac{1}{f}}^2 \right]^{n-1}}$$

$$\mathcal{D} = \{g \in F / \|g\|_{\frac{1}{f}}^2 < \infty\}.$$

Démonstration :

La vraisemblance est égale à $\frac{g^{\otimes n}}{f^{\otimes n}}$, l'expression analogue à $E_{\theta}(U_{\theta, \theta}^2)$ de la proposition 4 est ici : ($g \neq f$)

$$E_f \left[\left(\frac{g^{\otimes n}}{f^{\otimes n}} - 1 \right)^2 \right] = \int \frac{[g(x_1) \dots g(x_n)]^2}{f(x_1) \dots f(x_n)} dx_1 \dots dx_n - 2 \int g(x_1) \dots g(x_n) dx_1 \dots dx_n + 1$$

$$= \left[\int \frac{g^2(t)}{f(t)} dt \right]^n - 1.$$

En posant $a = \|g\|_{\frac{1}{f}}^2$, il vient :

$$a^n = \left(\|g\|_{\frac{1}{f}}^2 \right)^n = \left[\int_0^1 \frac{g^2(t)}{f(t)} dt \right]^n = E_f \left[\left(\frac{g^{\otimes n}}{f^{\otimes n}} - 1 \right)^2 \right] + 1 > 1$$

Corollaire 1. D. Bosq.-

Si \hat{h}_n est un estimateur sans biais et si \mathcal{D} est convexe,

alors :

$$E_f \int (\hat{h}_n - f)^2 dt \geq \frac{\inf f}{n}$$

En effet :

$$L_2[\underline{m}(g), m(f)] = \int \frac{[g(t) - f(t)]^2 dt}{f(t)}$$

$$" = \frac{\|g\|_1^2}{\bar{f}} - 1 = a - 1$$

et $R_2(\hat{h}_n, f) \geq \sup_{g \neq f} \frac{a-1}{a^n - 1}$

$$" \geq \sup_{g \neq f} \frac{1}{1 + a + \dots + a^{n-1}} \quad \text{avec } a^n > 1 \quad \text{et} \\ a > 1$$

. Si \mathcal{D} est convexe, on peut trouver g ($g \neq f$) tel que :

$$a = \frac{\|g\|_1^2}{\bar{f}} \text{ soit arbitrairement voisin de } 1.$$

En effet, soit une combinaison linéaire convexe de f et $g_0 \in \mathcal{D}$ et soit : $g = \alpha.f + (1-\alpha)g_0$

$$\implies \frac{\|g\|_1}{\bar{f}} \leq \alpha + (1-\alpha) \frac{\|g_0\|_1}{\bar{f}}$$

Si $\alpha \rightarrow 1$ alors $\frac{\|g\|_1}{\bar{f}} \rightarrow 1$ d'où le résultat.

. On obtient alors : $R_2(\hat{h}_n, f) = E_f \int \frac{(\hat{h}_n - f)^2}{f} dt \geq \frac{1}{n}$

d'où le résultat.

Corollaire 2 D. Bosq.-

Si \mathcal{D} est convexe et s'il existe un estimateur \hat{h}_n sans biais et à noyau borné, alors \mathcal{D} est borné dans $L^2(\lambda)$, λ étant cette fois la mesure de Lebesgue.

Démonstration :

$$\begin{aligned}
 E_f \int [\hat{h}_n(x^n, x) - f(x)]^2 dx &= \int [E_f [\hat{h}_n^2(x^n, x)] - f^2(x)] dx \\
 &= \int \left[E_f \left[\frac{1}{n} \sum_{i=1}^n K_{r(n)}(X_i, x) \right]^2 - f^2(x) \right] dx \\
 &= \int \left\{ E_f \left[\frac{1}{n^2} \sum_{i=1}^n K_{r(n)}^2(X_i, x) + \frac{1}{n^2} \sum_{i \neq j} K_{r(n)}(X_i, x) \cdot K_{r(n)}(X_j, x) \right] - f^2(x) \right\} dx \\
 &= \frac{1}{n} \left[\int K_{r(n)}^2(x, t) f(x) dx dt - \int f^2(t) dt \right] \quad \text{car}
 \end{aligned}$$

$E_f(\hat{h}_n) = \int K_{r(n)} K(x, t) f(x) dx = f(t)$; d'après l'inégalité obtenue au corollaire 1, on obtient :

$$\int K_{r(n)}^2(x, t) f(x) dx dt \geq \inf f + \int f^2(t) dt$$

d'où le résultat avec l'hypothèse $K_{r(n)}$ borné.

On peut donner aussi la minoration suivante :

$$\int K_{r(n)}^2(x, t) dx dt \geq \frac{\inf f + \int f^2(t) dt}{\sup f}$$

et

$$\int K_{r(n)}^2(x, t) dx dt \geq \frac{2 \inf f}{\sup f} .$$

Remarque : Boyd et Steele [12] ont énoncé un théorème analogue au corollaire 1.

Théorème [12].- Pour tout estimateur \hat{h}_n , il existe une densité f de carré intégrable et une constante $c > 0$ telle que :

$$E_f \left\{ \int_{-\infty}^{+\infty} [\hat{h}_n(x^n, x) - f(x)]^2 dx \right\} \geq \frac{c}{n} \quad \text{pour une infinité de } n.$$

Il n'existe donc pas d'estimateur pour lequel le MISE

(Mean integrated square error) est meilleur que $O(n^{-1})$.
f peut d'ailleurs être choisie comme une densité $N(0, \sigma)(x)$.

Exemple n° 1.-

Soit $F_\theta = \{N(\theta, 1(x))\}$ la famille de densités étudiée au chapitre I.B.II.4. On peut trouver dans [71] l'estimateur sans biais suivant : ($n \geq 2$), élément de B^2

$$\hat{h}_n^S(X_1 \dots X_n, x) = \left[2\pi \left(1 - \frac{1}{n}\right) \right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \frac{(x - \bar{X}_n)^2}{1 - \frac{1}{n}}\right\}$$

On utilise ici $L_2(\lambda)$ λ mesure de Lebesgue.

Nous allons donner l'efficacité d'un tel estimateur.

. Valeur de $R_2(\hat{h}_n^S, f_\theta)$.

Par un calcul analogue à celui effectué au B.II.4., on obtient

$$\text{var}(\hat{h}_n^S) = \frac{n}{2\pi(n^2-1)^{1/2}} e^{-\frac{n(x-\theta)^2}{n+1}} - \frac{1}{2\pi} e^{-(x-\theta)^2}$$

et

$$R_2(\hat{h}_n^S, f_\theta) = \frac{1}{2\sqrt{\pi}} \left[\left(\frac{n}{n-1}\right)^{\frac{1}{2}} - 1 \right].$$

. Valeur de $\sup_{\theta' \in M_\theta} \frac{||m(\theta') - m(\theta)||^2}{E_\theta(U_{\theta, \theta'}^2)}$.

On considère

$$U_{\theta, \theta'}(X_1 \dots X_n) = e^{-\left[(\theta' - \theta) \sum_{i=1}^n X_i + \frac{n}{2} (\theta'^2 - \theta^2) \right]} - 1.$$

Alors : $\sigma_\theta^2(U_{\theta, \theta'}) = e^{n(\theta' - \theta)^2} - 1.$

$$\text{D'autre part : } ||m(\theta') - m(\theta)||^2 = \frac{1}{\sqrt{\pi}} \left[1 - e^{-\frac{(\theta' - \theta)^2}{2}} \right]$$

$$\sup_{\theta' \in M_\theta} \frac{||m(\theta') - m(\theta)||^2}{E_\theta(U_{\theta, \theta'}^2)} \text{ est atteint pour } \theta' \rightarrow \theta$$

et on obtient l'inégalité suivante :

$$\frac{1}{2\sqrt{\pi}} \left[\left(\frac{n}{n-1} \right)^{\frac{1}{2}} - 1 \right] \geq \frac{1}{4\sqrt{\pi n}} \quad \text{inégalité vérifiée pour } n \geq 2.$$

Remarque : il est immédiat que $M_\theta = \theta = \mathbb{R}$.

$$\frac{1}{2\sqrt{\pi}} \left[\frac{1}{n-1+\sqrt{n(n-1)}} \right] \geq \frac{1}{4\sqrt{n} n} . \quad \text{L'efficacité est donc :}$$

$$e = \frac{n - 1 + \sqrt{n(n-1)}}{2n} \xrightarrow{n \rightarrow \infty} 1$$

Exemple n° 2.-

Même situation. Il a été démontré au chapitre I. B. II.4.

que l'estimateur minimax est

$$h^*(x^n, x) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+1}} \exp \left\{ -\frac{\frac{(x - \bar{X}_n)}{n}}{2 \cdot \frac{n+1}{n}} \right\} \quad h^* \in B^2$$

Dans le cas de la fonction de perte $L_2(\lambda)$ λ mesure de Lebesgue.

. Valeur de $R_2(h^*, f_\theta)$.

On obtient le risque suivant, dont la valeur est indépendante

de θ .

$$R_2(h^*, f_\theta) = \frac{1}{2\sqrt{\pi}} \left[1 - \sqrt{\frac{n}{n+1}} \right] .$$

. Valeur de $||m(\theta) - f(\theta)||^2$

Il a été calculé que :

$$m(\theta) = E_{\theta}(h^*) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+2}} \exp\left\{-\frac{(x-\theta)^2}{2 \cdot \frac{n+2}{n}}\right\} . \text{ On a donc :}$$

$$\begin{aligned} ||m(\theta) - N(\theta, 1)||^2 &= \int \left[\frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+2}} \exp\left\{-\frac{(x-\theta)^2}{2 \cdot \frac{n+2}{n}}\right\} - \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta)^2}{2}\right\} \right]^2 dx \\ " &= \frac{1}{2\sqrt{\pi}} \left[\sqrt{\frac{n}{n+2}} + 1 - 2 \sqrt{\frac{n}{n+1}} \right] . \end{aligned}$$

. Valeur de $||m(\theta') - m(\theta)||^2$.

$$\begin{aligned} ||m(\theta') - m(\theta)||^2 &= \int \left[\frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+1}} \exp\left\{-\frac{(x-\theta)^2}{2 \cdot \frac{n+2}{n}}\right\} - \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{n+2}} \exp\left\{-\frac{(x-\theta')^2}{2 \cdot \frac{n+2}{n}}\right\} \right]^2 dx \\ &= \frac{1}{\sqrt{\pi}} \sqrt{\frac{n}{n+2}} \left[1 - \exp\left\{-\frac{n}{n+2} \cdot \left(\frac{\theta-\theta'}{2}\right)^2\right\} \right] . \end{aligned}$$

D'autre part, on a, comme précédemment :

$$. E_{\theta}(U_{\theta, \theta'}^2) = e^{n(\theta'-\theta)^2} - 1$$

et

$$\sup_{\theta' \in M_{\theta}} \frac{\frac{1}{\sqrt{\pi}} \sqrt{\frac{n}{n+2}} \left[1 - \exp\left\{-\frac{n}{n+2} \left(\frac{\theta-\theta'}{2}\right)^2\right\} \right]}{\exp\{n(\theta'-\theta)^2\} - 1} \quad \text{correspond à}$$

$$\frac{1}{4\sqrt{\pi}} \sqrt{\frac{n}{n+2}} \cdot \frac{1}{n+2} .$$

L'inégalité est donc :

$$\frac{1}{2\sqrt{\pi}} \left[1 - \sqrt{\frac{n}{n+1}} \right] \geq \frac{1}{2\sqrt{\pi}} \left[\sqrt{\frac{n}{n+2}} + 1 - 2 \sqrt{\frac{n}{n+1}} \right] + \frac{1}{4\sqrt{\pi}} \left[\frac{1}{n+2} \sqrt{\frac{n}{n+2}} \right],$$

qui est bien toujours vérifiée.

D'où l'efficacité :

$$e' = \frac{\sqrt{\frac{n}{n+2}} + 1 - 2 \sqrt{\frac{n}{n+1}} + \frac{1}{2(n+2)} \sqrt{\frac{n}{n+2}}}{1 - \sqrt{\frac{n}{n+1}}} \xrightarrow{n \rightarrow \infty} 1.$$

Remarque (1). - Comparaison des risques.

On montre que : $\forall \theta \in \Theta = \mathbb{R}, \quad R_2(h_n^*, f_\theta) < R_2(\hat{h}_n^S, f_\theta)$,

dès que $n \geq 2$ (Condition d'existence de l'estimateur sans biais).

Remarque (2). - Comparaison des efficacités.

A l'aide de développements limités :

$$e \approx \frac{n}{n+1} \quad \text{et} \quad e' \approx \frac{n^2 + n + 6}{(n-1)(n+2)}$$

et $e' > e$ dès que $n \geq 2$.

II - Etude des vitesses de convergence d'estimateurs de la densité dans le cas de la perte quadratique intégrée.

Dans toute la suite du travail, on utilise la fonction de perte $L_2(\lambda)$ λ mesure de Lebesgue.

. Le risque associé est alors le M.I.S.E. (mean integrated square error).

. Les densités considérées seront supposées dérivables au sens de Lebesgue.

. L'espace de Hilbert $L^2(\lambda)$ sera noté ici L^2 .

On notera encore

- . W^m qui correspond à W_2^m , l'ensemble défini au début du Chapitre
- . W_L^m l'ensemble des densités $f/\mathcal{L}(f) \in L^2$, \mathcal{L} opérateur différentiel d'ordre m .
- . W_π^m l'ensemble des densités de $W^m/\text{supp } f \subset [-\pi, +\pi]$.
- . W_π^m l'ensemble des densités de $W_\pi^m/f^{(m)}$ soit continue.

1. Minoration du risque sur W^m .

Bretagnolle et Hubert [13] ont réalisé un travail analogue à celui de Wahba [59] pour l'estimation globale. La méthode est similaire, à cela près que la perturbation n'est plus locale et qu'au lieu de prendre un couple de densités dans l'ensemble considéré, on en prend un nombre croissant avec n . La minoration utilise l'information de Kullback.

. Soit la jauge ρ_m sur W^m (qui est d'ailleurs la jauge naturelle du problème) définie par :

$$f \in W^m \quad \rho_m(f) = \left\| f^{(m)} \right\|_2^{2/2m+1} \cdot \left\| f \right\|_1^{2m/2m+1}$$

On peut alors citer :

Théorème [13].-

Soit $m \geq 1$ et F_r l'ensemble des densités de probabilité sur \mathbb{R} qui vérifient $\rho_m(f) \leq r$.

Soit g_0 une densité auxiliaire de jauge $\rho_m(g_0) = r_0$,
 $r > r_0$.

Alors :
$$\liminf_n n^{\frac{2m}{2m+1}} \inf_{\hat{h}_n} \sup_{f \in F_r} E_f \left[\left\| \hat{f}_n - f \right\|_2^2 \right] \geq (2e)^{-4} \frac{r}{r_0} .$$

\hat{h}_n désignant un estimateur quelconque de f .

g_0 étant choisie telle que : g_0 positive, bornée par 1, nulle en dehors de : $\left[-\frac{3}{4}, +\frac{3}{4} \right]$, égale à 1 sur $\left[-\frac{1}{4}, +\frac{1}{4} \right]$, et de classe $C^{(m)}$.

2. Vitesse atteinte par l'estimateur \hat{h}_n^1 sur W^m .

Le premier résultat relatif à notre ensemble W^m est de

Nadaraya [42]. On peut citer :

Théorème d'après [42]. - Si K est un noyau de Parzen d'ordre m .

Alors \hat{h}_n^1 désignant l'estimateur associé, on a :

$$\forall f \in W^m \quad E_f \left(\left\| \hat{h}_n^1 - f \right\|_2^2 \right) = O(n^{-\frac{2m}{2m+1}}).$$

Remarque: [42] impose la condition supplémentaire f bornée, qui peut être supprimée car

$$\forall f^{(m)} \in L^2 \quad \text{pour } m \geq 1 \quad \Rightarrow \quad \|f\|_\infty < \infty$$

$$f \in L^2 \text{ d'après le lemme 1. A.I.3. chap. II.}$$

On obtient alors la proposition suivante :

Proposition [13]. - L'estimateur \hat{h}_n^1 déduit d'un noyau de Parzen de type m est asymptotiquement minimax sur W^m .

La démonstration est immédiate d'après la définition d'un estimateur asymptotiquement minimax et les résultats contenus dans les deux théorèmes précédents.

Remarque : Dans le cas de l'estimateur \hat{h}_n^1 la qualité de l'estimation dépend des propriétés de régularité de f et surtout de celles de K .

On pourra consulter pour cela [19] et [67]. On peut citer par exemple le résultat de Watson et Leadbetter [67], qui obtiennent un *mise minimum* J_n^* tel que : $J_n^* = O(\frac{\log n}{n})$ dans le cas d'une famille de densités telle que la fonction caractéristique $\psi_f(u)$ décroît exponentiellement et pour un choix de K tel que :

$$\psi_K(t) = \frac{n|\psi_f(t)|}{1+(n-1)|\psi_f(t)|^2} \quad \text{avec} \quad \psi_K(t) = \int K(y)e^{iuy}dy.$$

On a d'ailleurs vu que le *mise*, dans le cas de la famille $F_\theta = \{N(\theta,1)(x)\}$ (chapitre I. B.II.4. ; chapitre II.B.I.4) atteint la vitesse maximale $O(n^{-1})$.

3. Vitesse atteinte par $h_n^{2,H}$ sur $W_{\mathbb{E}_0}^m$.

Avec $\mathbb{E}_0(f) = (x-D)^m f$. \mathbb{E}_0 opérateur différentiel d'ordre m .

Citons d'abord deux théorèmes de Walter [63] qui améliorent les résultats de Schwartz [53].

Théorème 1. [63].-

Si $f \in W_{\mathbb{E}_0}^m$ et $q(n) = O(n^{1/m})$, alors :

$$E_f(||\hat{h}_n^{2,H} - f||_2^2) = O(n^{\frac{5}{6m} - 1}).$$

Théorème 2. [63].-

Si $f \in W^m$ et si, de plus, f est à support compact, et si $q(n) = O(n^{1/m})$, alors :

$$E_f(||\hat{h}_n^{2,H} - f||_2^2) = O(n^{\frac{1}{2m} - 1}).$$

Remarque :

On peut améliorer le résultat du théorème 2 avec $q(n) = O(n^{\frac{2}{2m+1}})$

et, cette fois, le mise est : $J_n = O(n^{-\frac{2m}{2m+1}})$.

. Nous allons maintenant, dans le cas de l'ensemble $W_{\tau_0}^m$ donner une amélioration du théorème 1 pour $m \geq 2$.

. (e_j) désigne ici la base de $L^2(\mathbb{R})$ constituée par les fonctions d'hermite. On rappelle qu'elles satisfont aux relations suivantes :

$$(x^2 - D^2)e_j = (2j+1)e_j \quad j = 0, 1, 2, \dots$$

Elles peuvent être exprimées à l'aide de polynômes d'hermite H_j

$$e_j(x) = \frac{H_j(x)e^{-x^2/2}}{(2^j j! \pi^{1/2})^{1/2}} \quad j = 0, 1, 2, \dots$$

Elles satisfont à la formule de récurrence : (voir [56], p. 106)

$$xe_j = \frac{j}{2} e_{j-1} + ((j+1)/2)^{1/2} e_{j+1} \quad j = 1, 2, \dots$$

et
$$De_j = \frac{j}{2} e_{j-1} - ((j+1)/2)^{1/2} e_{j+1} \quad j = 1, 2, \dots$$

Elles satisfont les inégalités suivantes (voir [56], p. 242) :

$$|e_j(x)| \leq C_\infty (j+1)^{-1/12} \quad x \in (-\infty, \infty), \quad j = 0, 1, \dots$$

et
$$|e_j(x)| \leq C_M (j+1)^{-1/4} \quad x \in (-M, M), \quad j = 0, 1, \dots$$

La constante C_M satisfait à : $C_M \leq 1 + \frac{1}{2} M^{5/2}$.

De façon analogue à B.I.2., on a :

$$\hat{h}_n^{2,H} = \sum_{k=1}^{q(n)} \hat{a}_k e_k(x) \quad \text{avec}$$

$$\hat{a}_k = \frac{1}{n} \sum_{j=1}^n e_k(x_j) \quad \forall k = 1, \dots, q(n)$$

$$E_f(\hat{a}_k) = a_k = \int e_k(n) f(x) dx.$$

Comme $\hat{h}_n^{2,H} - E_f(\hat{h}_n^{2,H})$ et $E_f(\hat{h}_n^{2,H}) - f$ sont orthogonaux, on a :

$$J_n = E_f(\|\hat{h}_n^{2,H} - f\|_2^2) = E_f(\|\hat{h}_n^{2,H} - E_f(\hat{h}_n^{2,H})\|_2^2) + \|E_f(\hat{h}_n^{2,H}) - f\|_2^2$$

$$J_n = \frac{1}{n} \sum_{k=1}^{q(n)} \left[\int e_k^2(x) f(x) dx - a_k^2 \right] + \sum_{k>q(n)} a_k^2$$

$$J_n < \frac{1}{n} \sum_{k=1}^{q(n)} \int e_k^2(x) f(x) dx + \sum_{k>q(n)} a_k^2$$

Avant d'énoncer le théorème 3., nous aurons besoin d'un lemme préparatoire.

Lemme. -

f étant une densité, s'il existe $\gamma > \frac{5}{3} / x^\gamma \quad f(x) \rightarrow 0$
 $|x| \rightarrow \infty$

Alors

$$\int e_j^2(x) f(x) dx < A j^{-1/2}$$

Démonstration :

L'hypothèse sur f peut encore s'écrire

$$\forall \varepsilon > 0 \quad \exists x_0(\varepsilon, \gamma) / \forall x > x_0 \quad f(x) < \frac{\varepsilon}{x^\gamma}$$

$$\text{Alors : } \int e_j^2(x) f(x) dx = \underbrace{\int_{-x_0}^{x_0} e_j^2(x) f(x) dx}_{(1)} + \underbrace{\int_{|x|>x_0} e_j^2(x) f(x) dx}_{(2)}$$

Pour (1), on utilise la majoration des fonctions d'hermite, valable sur tout compact

$$\int_{-x_0}^{+x_0} e_j^2(x) f(x) dx < C_{x_0}^2 j^{-1/2} \int_{-x_0}^{+x_0} f(x) dx < C_{x_0}^2 j^{-1/2} .$$

Pour (2)

$$\int_{|x| > x_0} e_j^2(x) f(x) dx < C^2 \varepsilon j^{-1/2} \int_{|x| > x_0} |x|^{2/3-\gamma} dx$$

car Swego p. 242

$$\text{soit } a > 0 \quad \left| \begin{array}{l} e_j(x) \leq C j^S x^{-\lambda} \quad \text{pour } |x| \geq a \\ \text{avec } S = \max\left(+\frac{\lambda}{2} - \frac{1}{12}, -\frac{1}{4}\right) \\ \text{appliqué à } \lambda = -\frac{1}{3} \end{array} \right.$$

$$\int_{x_0}^{\infty} |x|^{2/3-\gamma} dx \text{ sera convergente si } \gamma - \frac{2}{3} > 1 \text{ et } \gamma > \frac{5}{3} .$$

On obtient alors :

Théorème 3.-

Si $f \in W_{\mathbb{E}_0}^m$ et $q(n) = O(n^{\frac{2}{2m+1}})$ $m \geq 2$, alors

$$E_f(\|\hat{h}_n^{2,H} - f\|_2^2) = O(n^{-\frac{2m}{2m+1}}).$$

Ce théorème montre que, sur un ensemble de densités $W_{\mathbb{E}_0}^m \subset W^m$, l'estimateur $\hat{h}_n^{2,H}$ a la même vitesse que \hat{h}_n^1 sur W^m ; et comme l'hypothèse "f à support compact" n'influe pas sur la vitesse de \hat{h}_n^1 (d'après [19]), on peut donc conclure que sur $W_{\mathbb{E}_0}^m$, \hat{h}_n^1 et $\hat{h}_n^{2,H}$ convergent à la même vitesse (si f à support compact, $f \in W^m \implies f \in W_{\mathbb{E}_0}^m$).

Démonstration :

Soit b_j le coefficient de Fourier de $(x-D)^m f$ alors, d'après [63] : $b_k = \int (x-D)^m f \cdot e_k = \int f(x+D)^m \cdot e_k$
 $= (2k)^{1/2} (2k-2)^{1/2} \dots (2k-2m+2)^{1/2} a_{k-m}.$

A l'aide d'intégrations par parties et en utilisant les propriétés de (e_j) , on obtient :

$$\sum_{k>q(n)} a_k^2 \leq \sum_{k>q(n)} b_{k+m}^2 (2k)^{-m}$$

$$\leq \frac{1}{2^m (q+1)^m} \sum_{k>q(n)} b_k^2$$

et $\sum_{k>q(n)} b_k^2 < \infty$

d'où $\sum_{k>q(n)} a_k^2 < \frac{C}{q^m(n)}$ C constante (1)

(Schwartz [53] obtient une majoration en $\frac{C}{q^{m-1}(n)}$)

D'après le lemme précédent, si $\gamma > \frac{5}{3}$, on aura :

$$\frac{1}{n} \sum_{k=1}^{q(n)} \int e_k^2(x) f(x) dx < \frac{A}{n} \sum_{k=1}^{q(n)} k^{-1/2} < \frac{A \cdot q(n)^{1/2}}{n} \quad (2)$$

En prenant $q(n) = O(n^{2/2m+1})$, on a le résultat avec (1) et (2).

Si $m \geq 2$ $(x-D)^m f \in L^2 \implies x^m f \in L^2$ et $x^m f(x) \rightarrow 0$
 $|x| \rightarrow \infty$ CQFD.

Remarque (1).- Une autre inégalité valable sur tout \mathbb{R} (cf. [63], p. 1263) : $|e_n(x)| \leq (n+1)^{-1/4} (1 + \frac{|x|^{5/2}}{2})$ conduit à $\gamma > 6$.

L'inégalité de Swego donne donc de meilleurs résultats.

Remarque (2).- Si $f \in L^2([0, +\infty[)$, les fonctions de Laguerre ψ_n

$$\psi_n(x) = \left[\frac{n!}{\Gamma(n+\alpha+1)} \right]^{1/2} e^{-\frac{x}{2}} x^{\frac{\alpha}{2}} L_n^\alpha(x)$$

avec :

$$L_n^\alpha(x) = e^x \frac{x^{-\alpha}}{n!} \frac{d^n}{dx^n} e^{-x} x^{n+\alpha},$$

forment un système orthonormal de $L^2([0, +\infty[)$ qui ne semble cependant pas plus intéressant à utiliser.

En effet, d'après [38],

$$\psi_n(x^2) = \left[\frac{n!}{\Gamma(n + \frac{1}{2})} \right]^{\frac{1}{2}} e^{-\frac{x^2}{2}} x^{-\frac{1}{4}} L_n^{-\frac{1}{2}}(x^2) \text{ dans le cas}$$

$$\alpha = -\frac{1}{2}.$$

$$\Gamma(n + \frac{1}{2}) = \frac{(2n)! \sqrt{\pi}}{2^{2n} n!}. \text{ On obtient donc,}$$

$$\text{pour } x \in [a, b] \quad a > 0 \quad |\psi_n(x^2)| < \frac{B}{n^{1/4}}.$$

Les hypothèses sont plus restrictives que dans le cas hermite et la majoration obtenue n'est pas meilleure.

Remarque (3).- On a évidemment la version améliorée du théorème 2.

$$\text{Si } f \in W^m \text{ et si, de plus, il existe } \gamma > \frac{5}{3} / x^\gamma f(x) \rightarrow 0, \quad |x| \rightarrow \infty$$

alors :

$$E_f(\|\hat{h}_n^{2,H} - f\|_2^2) = O(n^{-2m/2m+1}) \text{ pour } q(n) = O(n^{2/2m+1}).$$

4. Vitesse atteinte par l'estimateur $\hat{h}_n^{2,T}$ sur W_π^m .

. On suppose cette fois que les densités sont à support inclus dans $[-\pi, +\pi]$.

. (e_k) désignant le système orthonormal de $L^2[-\pi, +\pi]$ constitué par les fonctions trigonométriques, il existe des éléments

$f \in W_{\pi}^m$ tels que :

$$|a_k| = O\left(\frac{1}{k^m}\right) \quad \text{avec} \quad a_k = \int e_k(x) f(x) dx.$$

Soit

$$J_n = E_f \left[\left\| \hat{h}_n^{2,T} - E_f(\hat{h}_n^{2,T}) \right\|_2^2 \right] + \left\| E_f(\hat{h}_n^{2,T}) - f \right\|_2^2$$

en notant $D_q(y) = \frac{\sin(q+1)\frac{y}{2}}{\sin\frac{y}{2}}$ le noyau de Dirichlet.

Alors $\left\| f * D_q - f \right\|_2^2 = \left\| E_f(\hat{h}_n^{2,T}) - f \right\|_2^2$

" $= \sum_{|k| > q(n)} |a_k|^2$

" $= O\left(\frac{1}{q(n)^{2m-1}}\right)$ d'après la formule de Parseval.

. L'autre partie du Mise se majore en remarquant que, dans le cas trigonométrique, il existe M , constante telle que :

$\int e_k^2(x) f(x) dx \leq M$ d'après [6]. Alors, on a la majoration de l'aléa :

$$E_f \left[\left\| \hat{h}_n^{2,T} - E_f(\hat{h}_n^{2,T}) \right\|_2^2 \right] \leq M \frac{q(n)}{n}$$

Pour $q(n) = O\left(n^{\frac{1}{2m}}\right)$, on a donc

$$J_n = O\left(n^{-\frac{2m-1}{2m}}\right)$$

. Remarque 1. - On comparera ce résultat avec celui obtenu au Chap. III.III.3.

. Remarque 2. - F_q , noyau de Fejer a des propriétés plus fortes que le noyau de Dirichlet. Cependant, si :

$$\sigma_q f(x) = \frac{1}{2\pi} \int f(y) F_q(x-y) dy = F_q * f(x)$$

$$S_q f(x) = \frac{1}{2\pi} \int f(y) D_q(x-y) dy = D_q * f(x).$$

Les résultats ne sont pas meilleurs pour les sommes de Cesaro car S_q est la meilleure approximation dans L^2 :

$$\|f - S_q f\|_2 \leq \|f - \sigma_q f\|_2.$$

. Nous avons, si $f \in W_\pi^m$.

- . $\text{supp } f \subset [-\pi, +\pi]$
- . f continue ... $f^{(m-1)}$ continue
- . $f^{(m)}$ existe
- . $f^{(m)} \in L^2[-\pi, +\pi]$.

On aura alors pour $f \in W_\pi^m$. ($f^{(m)}$ est ici continue)

$$f(-\pi_+) = f(-\pi_-) = f'(-\pi_+) = f'(-\pi_-) = \dots$$

$$\dots = f^{(m)}(-\pi_+) = f^{(m)}(-\pi_-) = 0$$

En considérant \tilde{f} l'extension 2π périodique de f appartenant à W_π^m , nous allons pouvoir donner une meilleure majoration du biais de J_n . (L'aléa étant toujours majoré par $\frac{M_q(n)}{n}$).

Lemme d'après [39].- T_q désignant un polynôme trigonométrique de degré q , alors $\forall m \geq 1 \quad \exists M_m$ constante / $\min \| \tilde{f} - T_q \|_\infty$

$\leq \frac{M_m}{q^m} \omega(\tilde{f}^{(m)}, \frac{1}{q})$ avec $\omega(\tilde{f}, h) = \max_{x,t \mid |t| \leq h} | \tilde{f}(x+t) - \tilde{f}(x) |.$

Démonstration :

Nous utiliserons le noyau de Jackson :

$$L_q(t) = \lambda_q^{-1} \frac{(\sin qt/2)^4}{\sin q/2} \quad \text{tel que : } \int_{-\pi}^{+\pi} L_q(t) dt = 1$$

qui se généralise en :

$$L_{q,r}(t) = \lambda_{q,r}^{-1} \left(\frac{\sin qt/2}{\sin t/2} \right)^{2r} \quad \text{tel que} \quad \int_{-\pi}^{+\pi} L_{q,r}(t) dt = 1.$$

On peut en déduire un noyau :

$$K_{q,r}(t) = L_{q,r}(t) \quad q' = \left[\frac{q}{r} \right] + 1$$

positif, pair, de degré q et tel que :

$$\int_{-\pi}^{+\pi} K_{q,r}(t) dt = 1 \quad \int_0^{\pi} t^k K_{q,r}(t) dt \approx q^{-k} \quad k = 0, \dots, 2r-2.$$

Soit l'opérateur

$$I_q(x) = I_{q,m}(\tilde{f}, x) = - \int_{-\pi}^{+\pi} K_{q,r}(t) \sum_{k=1}^{m+1} (-1)^k \binom{m+1}{k} \tilde{f}(x+kt) dt.$$

r étant le plus petit nombre tel que $r \geq \frac{1}{2} (m+3)$.

Cette intégrale est une combinaison linéaire de :

$$\int_{-\pi}^{+\pi} \tilde{f}(x+kt) \cos \ell t dt \quad \ell = 0, \dots, n \quad k = 1, \dots, m+1.$$

(comme fonction de t , $\tilde{f}(x+kt)$ a comme période $\frac{2\pi}{k}$).

Remarque : Soit Φ continue, périodique (ayant comme période $\frac{2\pi}{k}$), si ℓ n'est pas divisible par k , alors on aura :

$$\int_{-\pi}^{+\pi} \Phi(t) \cos \ell t dt = \int_{-\pi}^{+\pi} \Phi(t) \sin \ell t dt = 0$$

$$\text{en fait : } \int_0^{2\pi} \Phi(t) e^{i\ell t} dt = \int_{2\pi/k}^{2\pi+2\pi/k} \Phi(t) e^{i\ell t} dt = e^{2\pi i \ell / k} \int_0^{2\pi} \Phi(t) e^{i\ell t} dt.$$

. Si ℓ est divisible par k , en écrivant $u = x + kt$, on obtient que : $\int_{-\pi}^{+\pi} \tilde{f}(x+kt) \cos \ell t dt$ est un polynôme trigonométrique en x de degré $\frac{\ell}{k}$.

Alors $I_q(x)$ est un polynôme trigonométrique de degré q
 et $|\tilde{f}(x) - I_q(x)| = \left| \int_{-\pi}^{+\pi} K_{q,r}(t) \Delta_t^{m+1} \tilde{f}(x) dt \right| \leq \int_{-\pi}^{+\pi} K_{q,r}(t) \omega_{m+1}(f, |t|) dt$

avec

$$\left| \begin{aligned} \Delta_t^r \tilde{f}(x) &= \sum_{k=0}^r (-1)^{r-k} \binom{r}{k} \tilde{f}(x+kt) \\ \omega_r(\tilde{f}, h) &= \max_{x,t} \left| \Delta_t^r \tilde{f}(x) \right| \quad r = 1, 2, \dots \quad 0 \leq h \leq \frac{\ell}{r} \end{aligned} \right.$$

Comme $\omega_{m+1}(h) \leq (qh+1)^{m+1} \cdot \omega_{m+1}\left(\frac{1}{n}\right)$

$$\omega_r(f, h) \leq h^r \sup_x |f^{(r)}(x)|$$

$$\omega_{r+s}(f, h) \leq h^r \omega_s(f^{(r)}, h)$$

$$\omega_r(f, nh) \leq n^r \omega_r(f, h)$$

$$\omega_r(f, \lambda h) \leq (\lambda+1)^r \omega_r(f, h).$$

A l'aide des propriétés de $K_{q,r}(t)$ si $r \geq \frac{1}{2}(m+3)$

$$\begin{aligned} |\tilde{f}(x) - I_q(x)| &\leq 2\omega_{m+1}\left(\frac{1}{n}\right) \int_0^\pi (qt+1)^{m+1} K_{q,r}(t) dt \\ &\leq M_m \omega_{m+1}\left(\frac{1}{q}\right) \text{ d'après les propriétés de } \omega_r \end{aligned}$$

$\forall x \in \mathbb{R}$

$$\boxed{|\tilde{f}(x) - I_q(x)| \leq M_m \frac{1}{q^m} \omega\left(f, \frac{1}{q}\right)} .$$

On peut alors énoncer :

Théorème. -

1. Si $f \in W_\pi^m$ pour $q(n) = O(n^{1/2m+1})$

$$\text{Alors : } E_f = (|\hat{h}_n^{2,T} - f|_2^2) = O(n^{-2m/2m+1})$$

2. Si de plus $\omega(f^{(m)}, \frac{1}{q(n)}) \leq M_{(m)} \cdot \frac{1}{q^\alpha(n)}$

$$\text{pour } q(n) = O(n^{1/2\beta+1}) \quad 0 < \alpha < 1.$$

$$\text{Alors : } E_f (|\hat{h}_n^{2,T} - f|_2^2) = O(n^{-2\beta/2\beta+1}) \text{ avec } \beta = m + \alpha.$$

Démonstration :

Nous pouvons utiliser la majoration donnée dans le lemme pour \tilde{f} à une seule période.

$$\text{Alors } |f(x) - I_q(x)| \leq M_m \frac{1}{q^m} \omega(f^{(m)}, \frac{1}{q}).$$

Cette majoration étant uniforme sur $[-\pi, +\pi]$, on a donc :

$$\|f - I_q\|_2^2 \leq M_m'^2 \frac{1}{q^{2m}} \omega^2(f^{(m)}, \frac{1}{q}).$$

D'autre part, d'après l'inégalité de Parseval

$$\|f * D_{q(n)} - f\|_2^2 = \|E_f(h_n^{2,T}) - f\|_2^2 \leq \|f - I_q\|_2^2 \leq \frac{C}{q^{2m}} \|f^{(m)}\|_\infty \text{ si}$$

$f \in W_\pi^m$ pour tout polynôme I_q de degré $\leq q$.

$$\text{et } \|E_f(h_n^{2,T}) - f\|_2^2 \leq \frac{C'}{q^{2(m+\alpha)}} \text{ si, de plus } \omega(f^{(m)}, \frac{1}{q}) \leq M_m \frac{1}{q^\alpha}.$$

Comme l'aléa tend vers 0 avec $\frac{q(n)}{n}$, on obtient le résultat.

Remarque : La première hypothèse "f à support compact" n'influe pas sur la vitesse de convergence de \hat{h}_n^1 (voir par exemple [19]) ; la seconde non plus ($f^{(m)}$ continue). Il semble donc que dans cette situation l'estimateur $\hat{h}_n^{2,T}$ converge plus rapidement que \hat{h}_n^1 .

On peut citer un résultat pour la convergence en moyenne uniforme qui utilise le théorème de Lebesgue.

Théorème [39] p. 54.-

Soit \tilde{f} continue périodique. En désignant par $S_q(\tilde{f})$ la somme partielle de la série de Fourier, alors $\exists M$ constante telle que :

$$\|\tilde{f} - s_q(\tilde{f})\|_\infty \leq M E_q^*(\tilde{f}) \text{Log } q \quad \text{avec la notation précédente}$$

$$\text{où } E_q^*(\tilde{f}) = \min_{T_q} \|\tilde{f} - T_q\|_\infty.$$

On obtient alors :

Proposition.-

Soit $f \in W_{\pi}^{m}$ telle que $\omega(f^{(m)}, \frac{1}{q}) < M_m \cdot \frac{1}{q^{\alpha}}$ et telle que la série de Fourier de f converge uniformément sur $[-\pi, +\pi]$. Alors pour $q(n) = O(n^{1/m+\alpha+1})$.

$$E \left[\sup_{x \in [-\pi, +\pi]} |\hat{h}_n^{2,T}(x) - f(x)| \right] = O\left(\frac{\text{Log } n}{\frac{m+\alpha}{n^{m+\alpha+1}}}\right).$$

Démonstration :

$E_q^*(\hat{f})$ est majoré d'après le lemme précédent et, à l'aide du théorème de Lebesgue, on obtient alors :

$$\|\hat{f} - s_q(\hat{f})\|_{\infty} \leq B_m \cdot \frac{\text{Log } q}{q^{m+\alpha}}.$$

On applique cette majoration à une période en utilisant le fait que l'aléa tend vers 0 avec $q(n) \cdot \frac{\text{Log } n}{n}$ d'après [5]. Alors :

$$E(\|\hat{h}_n^{2,T} - f\|_{\infty}) \leq E(\|\hat{h}_n^{2,T} - s_q(f)\|_{\infty}) + \|s_q(f) - f\|_{\infty}$$

et
$$E(\|\hat{h}_n^{2,T} - f\|_{\infty}) = O\left(\frac{\text{Log } n}{\frac{m+\alpha}{n^{m+\alpha+1}}}\right).$$

Remarque (1).- Si la densité à estimer est par exemple uniforme sur $[0, 1]$ l'estimation par la méthode du noyau - cas du noyau uniforme - donne de mauvais résultats.

L'estimateur $\hat{h}_n^{2,T}$ ne fonctionne pas bien non plus ($f(0) = f(1) \neq 0$) mais l'estimateur $\hat{h}_n^{2,H}$ donne de bons résultats.

Remarque (2).- On peut utiliser une remarque de Neyman [37], reprise par Fellner [23]. Dans le cas d'une densité positive sur R , le système trigonométrique n'est évidemment pas approprié. Cependant, on peut aisément se ramener à une densité ayant comme support $[0, 1]$ (et appliquer le système trigonométrique).



Soit Z ayant une densité inconnue f sur \mathbb{R} .

Soit f_0 une première estimation de f et soit

$$X = F_0(Z) \quad \text{avec} \quad F_0(Z) = \int_{-\infty}^Z f_0(t) dt.$$

Alors X est à valeurs dans $[0,1]$ et sa densité est donnée par :

$$p(x) = \frac{f(F_0^{-1}(x))}{f_0(F_0^{-1}(x))}$$

ayant estimé p par \hat{p} l'estimateur \hat{f} de f sera donné par

$$\hat{f}(Z) = \hat{p}[F_0(Z)] f_0(Z) .$$

Cette remarque sera utilisée dans le chapitre III pour les estimateurs de Whittle-Brunk et Whittle-Wahba.

III - Détermination d'une famille d'estimateurs, asymptotiquement minimax pour le Mise.

. On a vu au paragraphe II.2 que l'estimateur \hat{h}_n^1 associé à un noyau de Parzen de type m est asymptotiquement minimax sur W^m .

. Ayant observé que de nombreux noyaux $K_{r(n)}(x_i, x)$ peuvent se décomposer en la somme de 2 noyaux $H(x_i - x)$ et $\varepsilon_r(x_i, x)$ il est naturel de donner, dans un premier temps, les conditions sur ε_r pour que l'estimateur \hat{h}_n associé à $K_{r(n)}$ soit asymptotiquement minimax sur W^m , avec

$$\left| \begin{array}{l} K_{r(n)}(x_i, x) = H_{r(n)}(x_i - x) + \varepsilon_r(x_i, x) \\ H_{r(n)} \quad \text{Noyau de Parzen de type } (m). \end{array} \right.$$

Dans un second temps, nous donnerons des conditions générales sur $K_{r(n)}$.

1. Première détermination d'une famille asymptotiquement minimax sur W^m pour le Mise.

Notations particulières :

$$\hat{h}_n(x) = \frac{1}{n} \sum_{i=1}^n K_r(x_i, x) = \frac{1}{n} \sum_{i=1}^n H_r(x-x_i) + \frac{1}{n} \sum_{i=1}^n \epsilon_r(x, x_i)$$

$$\hat{h}_n(x) = \hat{\mu}_n * H_r(x) + \frac{1}{n} \sum_{i=1}^n \epsilon_r(x, x_i) \quad \text{où } \hat{\mu}_n \text{ désigne la mesure empirique}$$

et * la convolution.

$$h_n(x) = E_f(\hat{h}_n) = f * H_r + \int \epsilon_r(x, t) f(t) dt$$

On obtient alors la proposition suivante :

Proposition. -

Soit H un noyau de Parzen de type m

Soit ϵ_r tel que $\int \epsilon_r^2(x, t) dt \leq c(x) r^{-\alpha} \quad \alpha > 0.$
avec $c \in L^1 \cap L^2.$

Alors l'estimateur \hat{h}_n de noyau $K_r(x_i, x) = H_r(x-x_i) + \epsilon_r(x, x_i)$ est tel que $f \in W^m, E_f(\|\hat{h}_n - f\|_2^2) = O(n^{-2m/2m+1}). \quad m \leq \alpha$

Il est, de plus asymptotiquement minimax sur $W^m.$

Démonstration :

On utilise la majoration du Mise

$$E_f(\|\hat{f}_n - f\|_2^2) \leq 2[E_f(\|\hat{f}_n - f_n\|_2^2) + \|f_n - f\|_2^2]$$

et nous allons majorer séparément les termes de l'aléa et du biais.

. Majoration de $\|f_n - f\|_2^2$

$$\|f_n - f\|_2 = \|f * H_r - f + \int \epsilon_r(x, t) f(t) dt\|_2$$

$$\leq \underbrace{\|f * H_r - f\|_2}_{(1)} + \underbrace{\left\| \int \epsilon_r(x,t) f(t) dt \right\|_2}_{(2)}$$

. Pour majorer (1) : sous réserve que H est un noyau de Parzen de type m, on obtient :

$$(1) \leq \frac{1}{r^m} \|f^{(m)}\|_2 \cdot \|H_m\|_1 \quad \text{d'après [13].}$$

H_m noyau déduit de H

$$H_m(x) = (-1)^m \int_x^\infty \frac{(y-x)^{m-1}}{(m-1)!} H(y) dy \quad x > 0 \quad \text{et} \\ H_m(-x) = (-1)^m \cdot H_m(x).$$

. Pour majorer (2) :

$$(2)^2 = \int \left[\int \epsilon_r(x,t) f(t) dt \right]^2 dx \quad \text{à l'aide de l'inégalité de Schwartz}$$

$$" \leq \int \left\{ \int \epsilon_r^2(x,t) dt \right\} \left\{ \int f^2(t) dt \right\} dx$$

$$" \leq \|f\|_2^2 r^{-\alpha} \int C(x) dx \quad \text{avec l'hypothèse faite sur } \epsilon_r.$$

Il est donc nécessaire que C soit intégrable et suivant les valeurs de α de l'hypothèse, on ne pourra obtenir des résultats que sur un ensemble de densités W^m tel que $m \leq \alpha$.

. Majoration de $E_f(\|\hat{f}_n - f_n\|_2^2)$

$$(\hat{f}_n - f_n)(x) = \frac{1}{n} \sum_{i=1}^n \{H_r(x-x_i)\} - f * H_r + \frac{1}{n} \sum_{i=1}^n \left\{ \epsilon_r(x,x_i) - \int \epsilon_r(x,x_i) f(x_i) dx_i \right\}$$

$$\|\hat{f}_n - f_n\|_2 \leq \|\hat{u}_n * H_r - f * H_r\|_2 + \left\| \frac{1}{n} \sum_{i=1}^n \left[\epsilon_r(x,x_i) - \int \epsilon_r(x,x_i) f(x_i) dx_i \right] \right\|_2$$

et

$$E_f \left\| \|\hat{f}_n - f_n\|_2^2 \right\| \leq \underbrace{2 \{ E_f(\|\hat{u}_n * H_r - f * H_r\|_2^2) \}}_{(1)} + \underbrace{\left\| \frac{1}{n} \sum_{i=1}^n \left[\epsilon_r(x,x_i) - \int \epsilon_r(x,x_i) f(x_i) dx_i \right] \right\|_2^2}_{(2)}$$

. Pour majorer (1) : on utilise l'inégalité de Khintchine [13] ;
Avec H noyau de Parzen (H_r borné à r fixé), on a alors une
majoration du type $O(\frac{r}{n})$ pour (1)².

. Pour majorer (2) :

$$E_f \left[\left\| \frac{1}{n} \sum_{i=1}^n \epsilon_r(x, x_i) - \int \epsilon_r(x, x_i) f(x_i) dx_i \right\|_2^2 \right] = \frac{1}{n} \int \text{var}(\epsilon_r(x, t)) dt$$

et

$$\begin{aligned} \frac{1}{n} \int \text{var}(\epsilon_r(x, t)) dt &\leq \frac{1}{n} \int E_f[\epsilon_r^2(x, t)] dt \\ &\leq \frac{1}{n} \iint \epsilon_r^2(x, t) f(x) dx dt. \end{aligned}$$

Avec l'hypothèse précédente sur ϵ_r , on obtient :

$$\frac{1}{n} \iint \epsilon_r^2(x, t) f(x) dx dt \leq \frac{1}{n} r^{-\alpha} \int C(x) f(x) dx.$$

Si C est de carré intégrable,

$$\frac{1}{n} \iint \epsilon_r^2(x, t) f(x) dx dt \leq \frac{r^{-\alpha}}{n} \|C\|_2 \|f\|_2.$$

D'après les majorations précédentes, nous pouvons conclure que :

$$\begin{aligned} J_n &\leq 2\{2\|f * H_r - f\|_2^2 + 2\|\int \epsilon_r(x, t) f(t) dt\|_2^2 + \dots / \dots \\ &\dots / \dots + 2E_f\|\hat{\mu}_n * H_r - f * H_r\|_2^2 + 2E_f(\|\frac{1}{n} \sum \epsilon_r - \int \epsilon_r f\|_2^2)\}. \end{aligned}$$

Avec $\|f * H_r - f\|_2^2 = O(\frac{1}{r(n)^{2m}})$

$$\|\int \epsilon_r(x, t) f(t) dt\|_2^2 = O(\frac{1}{r^{2\alpha}})$$

$$E_f(\|\hat{\mu}_n * H_r - f * H_r\|_2^2) = O(\frac{r}{n})$$

$$E_f(\|\frac{1}{n} \sum \epsilon_r - \int \epsilon_r f\|_2^2) = O(\frac{1}{nr^\alpha}).$$

En prenant $r(n) = O(n^{1/2m+1})$, il est immédiat que la vitesse maximale est atteinte sur W^m , $m \leq \alpha$, et on peut conclure grâce à la proposition sur la minoration du risque obtenue en B.II.1.

2. Conditions sur le noyau $K_{r(n)}(x,t)$ pour que la
famille d'estimateurs \hat{h}_n soit asymptotiquement
minimax pour le mise.

a. Cas E borné.

On a vu au chapitre II.A.I.3. que si $K_{r(n)}$ est un noyau satisfaisant aux hypothèses H.1. à H.7., pour $f \in F \subset Z_2^m$, alors la convergence en moyenne quadratique est uniforme sur E.

Il nous suffit de modifier l'hypothèse H.4. afin du "régler" convenablement la vitesse du noyau $K_{r(n)}$.

On prendra donc l'hypothèse H'_4 .

H'_4. Identique à H.4. mais :

$\exists \beta$ tel que $0 < \beta \leq m$.

Alors on peut énoncer :

Proposition. -

Soit $F \subset Z_2^m$ un ensemble de densités et

$K_{r(n)}$ un noyau satisfaisant aux hypothèses H.1. à H.3, H.'₄, et H.5. à H.7.

Alors l'estimateur associé \hat{h}_n est tel que

$$f \in F \quad E_f(\|\hat{h}_n - f\|_2^2) = O(n^{-\frac{2m}{2m+1}}).$$

Démonstration : - -

Elle est immédiate avec B.II.1.

b. Cas E non borné.

On prend $E = \mathbb{R}$.

Proposition.

Soit $W_{\mathcal{L}}^m = \{f \in W^m / \mathcal{L}(f) \in L^2\}$.

\mathcal{L} étant un opérateur différentiel d'ordre m .

Soit $K_{r(n)}$ un noyau vérifiant H.1. et H.2. et tel que :

H.8. $\cdot K_{r(n)} \in L^2(\mathbb{R} \times \mathbb{R})$

H.9. $\cdot \|K_{r(n)}\|_2^2 = O(r(n))$

H.10 $\cdot \exists \beta / 0 < \beta \leq m$ tel que

$$\|\mathcal{L}^{-1}(K_{r(n)}(x,t)) - \mathcal{L}^{-1}\delta(x-t)\|_2^2 = O(r^{-\beta}(n)).$$

Alors l'estimateur \hat{h}_n associé est tel que :

(avec $r(n) = O(n^{1/2m+1})$)

$$\forall f \in W_{\mathcal{L}}^m \quad E_f(\|\hat{h}_n - f\|_2^2) = O(n^{-\frac{2m}{2m+1}}).$$

Il est, de plus, asymptotiquement minimax sur W^m si

$$W_{\mathcal{L}}^m \supset W_m.$$

Démonstration :

Le Mise est égal à :

$$J_n = \int \frac{1}{n} \left[\int K_{r(n)}^2(x,t)f(t)dt - \left(\int K_{r(n)}(x,t)f(t)dt \right)^2 \right] dx + \int \left[\int K_{r(n)}(x,t)f(t)dt - f(x) \right]^2 dx.$$

La première partie du Mise :

$$\begin{aligned} \frac{1}{n} \int \left[\int K_{r(n)}^2(x,t)f(t)dt \right] dx &\leq \frac{1}{n} \cdot \|f\|_{\infty} \|K_{r(n)}\|_2^2 \\ &= O\left(\frac{r(n)}{n}\right) \text{ avec l'hypothèse H.9.} \end{aligned}$$

La deuxième partie du Mise :

$$\frac{1}{n} \int \left[\int K_{r(n)}(x,t) f(t) dt \right]^2 dx \leq \frac{1}{n} \int \left\{ \int K_{r(n)}^2(x,t) dt \cdot \int f^2(t) dt \right\} dx$$

" d'après l'inégalité de Schwartz

$$\leq \frac{1}{n} \|f\|_2^2 \cdot \|K_{r(n)}\|_2^2$$

La troisième partie du Mise :

$$\int \left[\int K(x,t) f(t) dt - f(x) \right]^2 dx = \int \left[\int (K(x,t) - \delta(x-t)) f(t) dt \right]^2 dx.$$

On intègre et :

$$\int (K(x,t) - \delta(x-t)) f(t) dt = \int [L^{-1}(K(x,t)) - L^{-1}(\delta(x-t))] [L(f)(t)] dt.$$

On applique l'inégalité de Schwartz et :

$$\int \left[\int K(x,t) f(t) dt - f(x) \right]^2 dx \leq \|L(f)\|_2^2 \|L^{-1}(K(x,t)) - L^{-1}(\delta(x-t))\|_2^2$$

" = $O(r_{(n)}^{-2\beta})$, en considérant $r(n) = O(n^{1/2m+1})$.

$$\text{On a alors : } \forall f \in W_{\mathbb{E}}^m, E_f(\|\hat{h}_n - f\|_2^2) = O(n^{-2m/2m+1})$$

et on peut conclure, avec le paragraphe B.II.1. au minimax asymptotique, si $W_{\mathbb{E}}^m$ contient W^m .

Exemple 1.- $\mathbb{E} = D^m$, K noyau de Parzen de type m .

Exemple 2.- $\mathbb{E} = (x^2 - D^2)^{m/2}$ m pair.

$$K_{r(n)}(x,t) = \sum_{i=0}^{r(n)} e_i(x) e_i(t), \quad e_i \text{ fonctions d'hermite.}$$

Dans ce second cas, on ne peut cependant pas conclure au minimax asymptotique car : $W_{\mathbb{E}}^m \not\subset W^m$, sauf si f est à support compact, ce qui modifie peut-être le résultat obtenu en B.II.1.

CHAPITRE III

ETUDE DE LA FAMILLE D'ESTIMATEURS DE NOYAUX

$$K_r(x, t) = \sum_{i=1}^{\infty} \lambda_{i,n} e_i(x) e_i(t) .$$

L'étude faite au chapitre II permet une comparaison de la rapidité de convergence des estimateurs \hat{h}_n^1 et \hat{h}_n^2 sur certains ensembles de densités.

On peut rappeler par exemple deux résultats significatifs :

. $\hat{h}_n^{2,T}$ converge à la même vitesse que \hat{h}_n^1 sur W_{π}^m

. $\hat{h}_n^{2,H}$ converge à la même vitesse que \hat{h}_n^1 sur W^m , dans

le cas où f est à support compact (il converge probablement aussi à la même vitesse sur bon nombre d'espaces W_L^m , L étant un opérateur différentiel d'ordre m).

Cependant, chacune de ces deux méthodes possède des insuffisances, que nous allons rappeler, au début de ce chapitre. Ensuite nous présenterons, dans le cas de la méthode par projection, une légère amélioration qui permettra de supprimer quelques uns des inconvénients de \hat{h}_n^2 . A la fin du chapitre, nous introduirons un nouveau critère de comparaison d'estimateurs qui avantage l'estimateur \hat{h}_n^2 sous réserve d'une information sur la dépendance entre F l'ensemble des densités à estimer et (e_i) la base de L^2 ($F \subset L^2$).

I - Comparaison des estimateurs \hat{h}_n^1 et \hat{h}_n^2

. L'estimateur \hat{h}_n^1

- . De nombreuses hypothèses sur le noyau K sont nécessaires pour assurer une bonne vitesse de convergence de l'estimateur \hat{h}_n^1 , et l'on est en droit de se poser le problème de la justification de telles hypothèses (cf. chap. II A.I.1.).
- . Si l'on se place d'un point de vue bayésien, ces hypothèses "a priori" peuvent être considérées comme très exigeantes et sous entendent, avant même l'estimation, une connaissance parfaite de la fonction à estimer.
- . Il est certes raisonnable d'utiliser un noyau, pair, borné, intégrable, mais supposer une condition en rapport avec la dérivée $m^{\text{ième}}$ de la densité, est difficilement admissible.
- . D'autre part l'estimateur est peut être strict, mais pour le praticien cette méthode est compliquée (il n'y a aucune relation simple entre \hat{h}_n^1 et \hat{h}_{n+1}^1), et on peut noter le manque de robustesse des estimateurs optimaux. Quant aux courbes, elles sont souvent en dent de scie.

. L'estimateur \hat{h}_n^2

Cet estimateur apparaît comme beaucoup plus "honnête" (et avec $K_n(x,t) = \sum_{i=1}^{r(n)} e_i(n) e_i(t)$), les relations simples entre les noyaux

K_n et K_{n+1} facilitent le calcul de \hat{h}_{n+1}^2 en fonction de \hat{h}_n^2 . L'estimateur \hat{h}_n^2 a une précision qui diminue peu quand la dimension de E augmente.

Cependant il n'est pas strict, et un même ensemble de densités F peut être inclus dans plusieurs espaces de hilbert ; le choix du bon espace n'est pas évident. La même remarque est à faire pour le choix de

la base et de la règle de tronquature. D'autre part cet estimateur est "aveugle" car on projette toujours de la même façon sans tenir compte de l'ensemble des densités à estimer.

Alors que sans tomber dans l'excès précédent, on dispose toujours d'idées a priori qu'il est bon d'utiliser.

Cette information apparaîtra sous forme de coefficients attachés aux éléments (e_i) de la base dans l'amélioration de \hat{h}_n^2 , que nous allons présenter.

Cette amélioration répond à bon nombre d'objections et constitue une sorte de bayésification de la méthode par projection.

N.B. - On trouvera dans V quelques unes des courbes obtenues par

Ouledcheikh [45] pour l'estimateur \hat{h}_n^2 .

II - Effet d'un changement de base sur l'estimateur \hat{h}_n^2 .

Un espace F peut être inclus dans plusieurs espaces de hilbert du type $L^2(\lambda)$ [λ mesure de Lebesgue, ou mesure qui admet une densité par rapport à la mesure de Lebesgue] ou d'un autre type comme $H_1(\mathbb{R})$.

On peut se demander, pour un ensemble F considéré, quel est le meilleur espace de hilbert adapté à F (ou quelle est la meilleure norme, ou quelle est la meilleure base).

Soit par exemple $F \subset W_2^m$. Au chapitre II, nous avons utilisé l'espace de Hilbert $L^2(\lambda)$ (λ mesure de Lebesgue sur \mathbb{R}). Nous allons ici utiliser $H_1(\lambda)$ et sa base, et comparer les résultats avec les précédents H_1 désigne l'ensemble des f , éléments de L^2 tels que $f^{(1)}$ appartient aussi à L^2 .

On note :

. $\hat{h}_n^{2,\psi}$ l'estimateur correspondant au système orthonormal

(ψ_i) de H_1 , espace de Hilbert.

. $h_n^{2,\psi} = E_f(\hat{h}_n^{2,\psi})$.

. $\langle f, g \rangle_{(1)}$ le produit scalaire et $\| \cdot \|_{(1)}$ la norme correspondante de H_1 , avec :

$$\langle f, g \rangle_{(1)} = \langle f, g \rangle + \langle f', g' \rangle$$

(où $\langle . \rangle$ est le produit scalaire de L^2).

. La fonction de perte sera donc égale à :

$$L(f, g) = \|f - g\|_{(1)}^2 = L_2(f, g) + L_2(f', g')$$

. G désigne l'espace vectoriel engendré par $(e_{3p+1})_{p \in \mathbb{N}}$ où e_p désigne la p -ième fonction d'hermite.

Lemme 1.-

Soit $(\psi_i)_{i \in \mathbb{N}}$ un système orthonormal et total de H_1 . Soit F un ensemble de densités, $F \subset H_1$, tel que

$$\forall f \in F \quad \forall i \in \mathbb{N} \quad \left[\overline{f(t) \cdot \psi_i'(t)} \right]_{-\infty}^{+\infty} = 0.$$

Alors $\hat{a}_i = \frac{\sum_{j=1}^n \frac{\psi_i(x_j) - \psi_i'(x_j)}{n}}$ est un estimateur sans

biais de a_i (où a_i désigne le i -ème coefficient de Fourier).

Démonstration :

$$a_i = \langle f, \psi_i \rangle_{(1)} = \langle f, \psi_i \rangle + \langle f', \psi_i' \rangle$$

et

$$\langle f', \psi_i' \rangle = \int f'(t) \psi_i'(t) dt = \left[f(t) \cdot \psi_i'(t) \right]_{-\infty}^{+\infty} - \int f(t) \psi_i''(t) dt$$

Avec l'hypothèse on a donc :

$$\langle f', \psi_i' \rangle = - \int f(t) \psi_i''(t) dt \quad \text{d'où}$$

$a_i = \langle f(t), \psi_i - \psi_i'' \rangle$; il est donc immédiat que :

$$E_f(\hat{a}_i) = a_i .$$

Lemme 2.-

Si il existe une constante M telle que :

$$\forall i \in \mathbb{N} \quad \int \left[\psi_i(x) - \psi_i''(x) \right]^2 f(x) dx \leq M \cdot i^{1/2} < \infty ,$$

lorsque f est une densité telle que : $\exists \gamma > \frac{5}{3} / x f(x) \rightarrow 0$
 $|x| \rightarrow \infty$

Alors pour $r(n) \rightarrow \infty$ quand $n \rightarrow \infty$
 $r(n)$ tel que : $\frac{r^{3/2}(n)}{n} \rightarrow 0$ quand $n \rightarrow \infty$

On a : $E_f \left[\left\| \hat{h}_n^{2,\psi} - f \right\|_{(1)}^2 \right] \xrightarrow{n \rightarrow \infty} 0 .$

Démonstration : Elle est immédiate à l'aide de la majoration suivante du risque :

$$E_f[|\hat{h}_n^{2,\psi} - f|_{(1)}^2] = E_f[|\hat{h}_n^{2,\psi} - h_n|_{(1)}^2] + ||h_n - f||_{(1)}^2$$

car $\hat{h}_n^{2,\psi} - h_n$ et $h_n - f$ sont orthogonaux .

$$E_f[|\hat{h}_n^{2,\psi} - f|_{(1)}^2] = \sum_{i=1}^{r(n)} E_f(\hat{a}_i - a_i)^2 + \sum_{i>r(n)} a_i^2$$

$$E_f[|\hat{h}_n^{2,\psi} - f|_{(1)}^2] = \sum_{i=1}^{r(n)} [E_f(\hat{a}_i^2) - a_i^2] + \sum_{i>r(n)} a_i^2$$

$$E_f[|\hat{h}_n^{2,\psi} - f|_{(1)}^2] < \frac{1}{n} \sum_{i=1}^{r(n)} \int [\psi_i(x) - \psi_i''(x)]^2 f(x) dx + \sum_{i>r(n)} a_i^2$$

$$" < \frac{M \cdot r^{3/2}(n)}{n} + \sum_{i>r(n)} a_i^2 , \text{ d'où le résultat .}$$

Lemme 3.- Soit F un ensemble de densités tel que, pour

$f \in F$:

$$\exists \gamma > \frac{5}{3} / x^\gamma f(x) \rightarrow 0 \\ |x| \rightarrow \infty$$

$$. F \subset H_1 \cap G .$$

Alors la suite $(\psi_p)_{p \in \mathbb{N}}$, définie par :

$$\psi_p = \alpha_{3p+1} \cdot e_{3p+1} \quad (e_i \text{ est la } i\text{-ème fonction d'hermite}).$$

$$\text{avec } \alpha_{3p+1} = \sqrt{\frac{2}{6p+5}}$$

forme une suite orthonormale dans H_1 (et totale dans $H_1 \cap G$).

De plus, la suite $(\psi_p)_{p \in \mathbb{N}}$ vérifie les conditions des lemmes 1 et 2.

Démonstration :

$$e_i'(x) = \sqrt{\frac{i+1}{2}} e_{i+1}(x) - \sqrt{\frac{i}{2}} e_{i-1}(x) . \quad \text{Il est donc immédiat}$$

que : $\forall p \in \mathbb{N}$

$$p \neq q \quad \langle \varphi_p, \varphi_q \rangle_{(1)} = 0 .$$

$\forall q \in \mathbb{N}$

D'autre part, $\|\varphi_p\|_{(1)}^2 = \|\varphi_p\|_2^2 + \|\varphi_p'\|_2^2 = \alpha_{3p+1}^2 \|e_{3p+1}\|_{(1)}^2$ et

un calcul simple montre que : $\|e_{3p+1}\|_{(1)}^2 = \frac{6p+5}{2}$.

. Comme f est une densité sur \mathbb{R} et d'après les propriétés des fonctions d'hermite :

$\forall i \in \mathbb{N}$

$$\forall f \in F \quad [f(x) \cdot e_i(x)]_{-\infty}^{+\infty} = 0 .$$

on a de même

$$[f(x) \cdot \varphi_i(x)]_{-\infty}^{+\infty} = 0 .$$

. Enfin on obtient :

$$e_i''(x) = \frac{[i(i-1)]^{1/2}}{2} e_{i-2}(x) - \frac{2i+1}{2} e_i(x) + \frac{[(i+1)(i+2)]^{1/2}}{2} e_{i+2}(x)$$

et

$$\varphi_i(x) - \varphi_i''(x) = \frac{2i+3}{\sqrt{2(6i+5)}} e_i(x) - \sqrt{\frac{i(i-1)}{6i+5}} e_{i-2}(x) - \sqrt{\frac{(i+1)(i+2)}{6i+5}} e_{i+2}(x)$$

D'autre part :

Si $x^Y f(x) \rightarrow 0$ alors, d'après le lemme du chap. II B.II.3.,

$$|x| \rightarrow \infty$$

$$\int e_i^2(x) f(x) dx < A i^{-1/2} \quad A \text{ constante positive finie}$$

et il est immédiat que :

$$\exists M < \infty \text{ t.q. } \int (\psi_i(x) - \psi_i''(x))^2 f(x) dx \leq M \cdot i^{1/2}$$

Alors on peut énoncer :

Proposition. -

Soit $f \in F \subset W^m \cap G$ ($m > 2$).

On suppose qu'il existe $\gamma > \frac{5}{3}$ t.q. $x^\gamma f(x) \rightarrow 0$
 $|x| \rightarrow \infty$

Alors pour $r(n) = O(n^{2/2m-1})$, on a :

$$E_f \left[\left\| \hat{h}_n^2 - f \right\|_{(1)}^2 \right] = O(n^{-2m-4/2m-1})$$

Démonstration : Il suffit d'évaluer les coefficients de Fourier (a_i) et d'utiliser la majoration donnée dans la démonstration du lemme 2.

$$\begin{aligned} f(x) &= \sum_{i=0}^{\infty} a_i \psi_i(x) = \sum_{i=0}^{\infty} a_i \alpha_{3i+1} e_{3i+1}(x) \\ &= \sum_{i=0}^{\infty} b_{3i+1} e_{3i+1}(x) \text{ avec } b_{3i+1} = a_i \alpha_{3i+1} \end{aligned}$$

On sait que si : $f \in W^m$, $b_i^2 = O(\frac{1}{i^m})$ d'où :

$$\begin{aligned} a_i^2 &= \frac{b_{3i+1}^2}{\alpha_{3i+1}^2} = \frac{6i+5}{2} \cdot b_{3i+1}^2 \\ &= O\left(\frac{1}{i^{m-1}}\right) \end{aligned}$$

et, en utilisant la majoration $\sum_{i>r(n)} \frac{1}{i^{m-1}} < \int_{r(n)}^{\infty} \frac{dx}{x^{m-1}}$, on obtient le résultat pour $r(n) = O(n^{2/2m-1})$.

Conclusion :

. Le risque associé à cette nouvelle fonction de perte a une vitesse de convergence inférieure à celle correspondant à la fonction de perte $L_2(\lambda)$ pour le même ensemble de densités.

. D'autre part dans le cas de l'estimateur \hat{h}_n^1 , on peut montrer que la vitesse de convergence du risque correspondant à la nouvelle fonction de perte est supérieure à celle de \hat{h}_n^2 , d'après :

$$E_f(\|\hat{h}_n^1 - f\|_{(1)}^2) = E_f[\|\hat{h}_n^1 - f\|_2^2] + E_f[\|(\hat{h}_n^1)' - f'\|_2^2],$$

et donc l'espace H^1 , sa norme et son système orthonormal ne sont pas adaptés au problème puisque l'utilisation de L^2 et de sa base donne de meilleurs résultats.

III - Etude de la famille d'estimeurs \hat{h}_n de noyaux

$$K_n(x,t) = \sum_{i=0}^{\infty} \lambda_i(n) e_i(x) e_i(t) .$$

Soit $f \in F \subset L^2$ (L^2 muni de la base (e_i)).

. L'idée de ce paragraphe est d'utiliser l'information que l'on possède afin de pondérer les éléments de la base (e_i) et d'obtenir ainsi une variante de l'estimateurs \hat{h}_n^2 , plus adaptée à F , (ensemble des densités à estimer), ou tout simplement positive (au IV l'information modifiera la fonction de perte et non l'estimateur).

. Il est intéressant de rappeler les résultats anciens sur ce sujet, avant d'étudier de façon plus précise deux résultats récents dûs à Brunk [14] et [15] et Wahba [60].

1. Rappels bibliographiques.

. L'idée d'attacher des coefficients aux éléments de la base du Hilbert était déjà contenue dans Cencov [17] en 1962 et dans Kronmal et Tarter [36] en 1968.

. Watson [65] en 1969 a déterminé dans cet esprit un noyau généralisé :

$$K_n(x,t) = \sum_{i=0}^{\infty} \lambda_i(x) e_i(x) e_i(t) ;$$

(qui converge vers la fonction de Dirac de représentation $\sum_{i=0}^{\infty} e_i(x) e_i(t)$).

$$\text{Dans le cas particulier où : } \left| \begin{array}{l} \lambda_i(x) = 1 \quad \text{si } i \leq r(n) \\ \lambda_i(x) = 0 \quad \text{si } i > r(n) \end{array} \right.$$

il retrouve bien entendu le noyau de l'estimateur \hat{h}_n^2 . Un bon choix des $\lambda_i(n)$ permet d'avoir un mise minimum.

$$\text{En effet, } \left| \begin{array}{l} \hat{h}_n(x^n, x) = \frac{1}{n} \sum_{j=1}^n K_n(x_j, x) \\ \hat{h}_n(x^n, x) = \sum_{i=0}^{\infty} \lambda_i(n) \hat{a}_i e_i(x) \\ f(x) = \sum_{i=0}^{\infty} a_i e_i(x) \quad \text{avec} \quad \hat{a}_i = \frac{1}{n} \sum_{j=1}^n e_i(x_j) \end{array} \right.$$

$$\begin{aligned}
 J_n &= E_f [\| \hat{h}_n - f \|]^2 = E_f \left[\sum_{i=0}^{\infty} (\lambda_i(n) \hat{a}_i - a_i)^2 \right] \\
 &= \sum_{i=0}^{\infty} \left[a_i^2 (1 - \lambda_i(n))^2 + \frac{1}{n} \lambda_i^2(n) \text{var} [e_i(x)] \right] \\
 &= \sum_{i=0}^{\infty} \left[\lambda_i^2(n) \left[a_i^2 + \frac{1}{n} \text{var}(e_i(x)) \right] - 2 \lambda_i(n) a_i^2 + a_i^2 \right]
 \end{aligned}$$

Le minimum est atteint pour :

$$\lambda_i(n) = \frac{a_i^2}{a_i^2 + \frac{1}{n} \text{var}(e_i(x))}$$

ou

On a alors : $\lambda_i(n) \rightarrow 1$ et $\lambda_i(n) \rightarrow 0$
 $n \rightarrow \infty$ $i \rightarrow \infty$

$$\lambda_i(n) = \frac{n}{(n-1) + E[e_i^2(x)]/a_i^2}$$

et :

$$\min_{\lambda_i(n)} J = \sum_{i=0}^{\infty} \frac{a_i^2 \text{var}[e_i(x)]}{\text{var}[e_i(x)] + n a_i^2} .$$

Les $\lambda_i(n)$ sont des coefficients adaptés à F . Cependant les quantités $E[e_i^2(x)]/a_i^2$ sont rarement connues. (Dans le cas trigonométrique,

Tarter et Raman [57] estiment $\lambda_i(n)$ par $\hat{\lambda}_i(n) = \frac{n}{n-1} - \frac{1}{(n-1) (\hat{a}_i)^2}$)

Exemple de calcul de $\min J (e_i)$ système trigonométrique.

Soit $F \subset W_{\pi}^m$ et on suppose qu'il existe A et B , constantes positives et finies telles que $\frac{A}{i^{2m}} \leq a_i^2 \leq \frac{B}{i^{2m}}$.

On sait qu'il existe C et D constantes positives et finies

telles que $C \leq E[e_i^2(x)] \leq D$. Alors :

$$\min J \leq \sum_{i=0}^{\infty} \frac{a_i^2 (D - a_i^2)}{C + (n-1) a_i^2}$$

$$\min J \leq \sum_{i=0}^{\infty} \frac{D a_i^2}{C + (n-1) a_i^2} - \sum_{i=0}^{\infty} \frac{a_i^4}{C + (n-1) a_i^2}$$

Les deux séries sont convergentes.

La fonction de n qui tend vers zéro le moins rapidement est donnée par :

$$\sum_{i=1}^{\infty} \frac{D a_i^2}{C + (n-1) a_i^2} \leq \sum_{i=1}^{\infty} \frac{\frac{D \cdot B}{i^{2m}}}{C + \frac{(n-1)A}{i^{2m}}} = \sum_{i=1}^{\infty} \frac{D B}{C \cdot i^{2m} + (n-1) A}$$

On majore alors la série par l'intégrale :

$$\sum_{i=1}^{\infty} \frac{D a_i^2}{C + (n-1) a_i^2} < D B \int_0^{\infty} \frac{1}{C \cdot x^{2m} + (n-1) A} dx = \dots/\dots$$

$$\dots/\dots = \frac{D B}{(n-1)A} \int_0^{\infty} \frac{1}{\frac{C}{(n-1)A} \cdot x^{2m} + 1} dx \quad . \quad \text{Soit } X = x \left[\frac{C}{(n-1)A} \right]^{1/2m}$$

$$= \frac{DB (n-1)^{1/2m} A^{1/2m}}{(n-1) A C^{1/2m}} \int_0^{\infty} \frac{1}{X^{2m} + 1} dX$$

convergente

et $\min J = O(n^{-2m-1/2m})$.

Il n'y a donc pas eu d'accélération de la convergence pour le Mise de cet estimateur par rapport à celui de $\hat{h}_n^{2,T}$ dans la même situation.

. Whittle [69] en 1958, en considérant cette fois des arguments bayésiens obtient un estimateur analogue à celui de Watson.

. f_0 étant une densité proche de la densité f à estimer, (f_0 correspond à une première estimation de f), soit alors $\{e_i(x)\}$ une suite orthonormale par rapport à f_0 .

. On suppose que $e_0(x) = 1$

Soit alors la suite (a_i) telle que :

$$f(x) = f_0(x) \sum_{i=0}^{\infty} a_i e_i(x)$$

($a_0 = 1$ pour que f soit une densité).

Whittle considère les estimateurs du type $\hat{h}_n^{(\lambda_i)}$ définis par :

$$\left[\begin{array}{l} \hat{h}_n^{(\lambda_i)}(x^n, x) = f_0(x) \left[1 + \sum_{i=1}^n \lambda_i(n) \hat{a}_i e_i(x) \right] \\ \hat{a}_i = \left[\frac{1}{n} \sum_{j=1}^n e_i(X_j) \right] \\ \lambda_i(n) = \frac{n}{(n-1) + \pi_i} \end{array} \right.$$

. π_i est une précision a priori $\pi_i = [\text{var}(\tilde{a}_i)]^{-1}$.

. On note \tilde{a}_i l'élément a_i considéré cette fois comme aléatoire, c'est-à-dire que l'on considère une distribution a priori \mathcal{P} sur $\{\tilde{f}(x) \mid x \in E\}$.

Whittle considère l'existence des deux premiers moments de la distribution a priori \mathcal{P} sur $\{\tilde{f}(x) \mid x \in E\}$.

La fonction de risque de son problème d'estimation ponctuelle est :

$$r(x) = E_P E_f (\hat{h}_n(x) - f(x))^2$$

. Fellner [23] en 1974 améliore le travail de Whittle en ce sens qu'il ne suppose cette fois l'existence que du premier moment $f_0(x)$ de la loi a priori P :

$$f_0(x) = E_P [\tilde{f}(x)]$$

Le moment d'ordre 2 sera estimé et dépend de λ , une constante qui permet ensuite de fixer la tronquature de la série.

. On trouvera dans le V les courbes obtenues par Fellner pour différentes valeurs de λ ; la courbe la plus intéressante est celle où il utilise la transformation décrite dans la remarque 2, à la fin du B. II.4, chap. II. Cependant la détermination de λ et de la tronquature font que cette méthode garde les inconvénients de l'estimateur \hat{h}_n^2 .

2. Estimateurs de Whittle - Brunk [14] [15].

. La distribution P sur $\{\tilde{f}(x), x \in E\}$ est donnée ici au moyen d'une distribution a priori sur (\tilde{a}_i) . Cette distribution sera prise de telle façon que $\tilde{a}_1, \tilde{a}_2, \dots$ soient indépendants. D'autre part, on aura :

$$E_P(\tilde{f}(x)) = f_0(x) \quad \forall x \in E \quad \text{si} \quad E_P(\tilde{a}_i) = 0 \quad i = 1, 2, \dots$$

. La spécification des moments d'ordre 1 et 2 sera établie en donnant la variance de (\tilde{a}_i) . On pose :

$$\sigma_i^2 = \text{var}(\tilde{a}_i) \quad i = 1, 2, \dots \quad \text{variances telles que} \quad \sigma_i^2 \rightarrow 0 \quad \text{si} \quad i \rightarrow \infty .$$

. Soit alors

$$\left| \begin{array}{l} a_0 = 1 \quad E(\tilde{a}_i) = 0 \quad i = 1, 2, \dots \\ \text{cov}(\tilde{a}_r, \tilde{a}_s) = \frac{\delta_{r,s}}{r} \quad r, s = 1, 2, \dots \\ \pi_r = \frac{1}{\sigma_r^2} \quad (\delta_{rr} = 1 ; \delta_{rs} = 0 \quad r \neq s). \end{array} \right.$$

. Cas particulier.

L'estimateur \hat{h}_n^2 de Kronmal-Tarter s'obtient avec $\sigma_i^2 = 1$ pour $i = 1, 2, \dots, r(n)$ et $\sigma_i^2 = 0$ pour $i > r(n)$.

$$\hat{h}_n^2(x^n, x) = f_0(x) \sum_{i=0}^{r(n)} \tilde{a}_i e_i x.$$

On peut cependant espérer une meilleure information.

. Dans le cas où f_0 est proche de f à estimer et si une grande précision π_i peut être donnée, on peut compter sur des estimateurs positifs.

. Au lieu de donner une distribution a priori pour chaque $\tilde{\beta}_i$, on peut en donner une pour l'hyperparamètre $(\tilde{\beta}_1 \dots \tilde{\beta}_i \dots)$.

On est amené à des spécifications du type :

$\text{var } \tilde{a}_i = c \cdot \rho^{-i} \quad (i = 1, 2, \dots ; c \text{ constante})$

Avant de donner un moyen de déterminer, à l'aide des données le paramètre ρ (ce qui constitue l'originalité du travail de Brunk), nous allons comparer les valeurs du mise de l'estimateur \hat{h}_n^2 et de ce nouvel estimateur $\hat{h}_n^{(\lambda_i)}$ pour différentes valeurs de ρ .

a. Comparaison du mise entre \hat{h}_n^2 et $\hat{h}_n^{(\lambda_i)}$.

On rappelle que :

$$\hat{h}_n^2(x^n, x) = f_0(x) \sum_{i=0}^{r(n)} [\tilde{a}_i e_i(x)]$$

$$\hat{h}_n^{(\lambda_i)}(x^n, x) = f_0(x) \sum_{i=0}^{\infty} \lambda_i(n) [\hat{a}_i e_i(x)]$$

$$\text{avec } \left\{ \begin{array}{l} \lambda_i(n) = \frac{n}{n-1+\pi_i} = \frac{n}{n-1+1/\sigma_i^2} = \frac{n}{n-1+c^{-1} \rho^i} \\ \hat{a}_i = \frac{1}{n} \sum_{j=1}^n e_i(X_j). \end{array} \right.$$

L'étude est faite dans le cas trigonométrique sur $E = [0, 1]$, dans le cadre suivant.

- f_0 sera prise égale à 1 sur $[0, 1]$.
- On prendra différentes valeurs de $r(n)$ correspondant à $n = 20, 50, 100$.
- On prendra aussi différentes valeurs de ρ , en considérant :

$$\text{var}(a_i) = \rho^{1-k} \quad \text{pour } i = 2k \text{ ou } i = 2k-1 \\ k = 1, \dots$$

- f à estimer est égale à :

$$f(x) = \frac{8}{\sqrt{2\pi}} \exp \left\{ -32 \left(x - \frac{1}{2} \right)^2 \right\} \quad \text{si } x \in]0, 1[$$

$$f(x) = 0 \quad \text{sinon.}$$

Valeurs du
mise de $\hat{h}_n^{2,T}$

r(n)	n = 20	n = 50	n = 100
2	0,473	0,443	0,432
4	0,381	0,251	0,188
6	0,487	0,308	0,218
8	0,581	0,367	0,260

Valeurs du
mise de $\hat{h}_n^{(\lambda_i)}$

ρ	$n = 20$	$n = 50$	$n = 100$
3	0,445	0,327	0,255
5	0,373	0,271	0,210
8	0,344	0,244	0,188
10	0,338	0,238	0,182
15	0,337	0,236	0,178
20	0,343	0,239	0,180
50	0,383	0,277	0,208

Conclusions :

* Les résultats sont meilleurs pour le second estimateur $\hat{h}_n^{(\lambda_i)}$.

* Le fait d'ajouter ou de retirer un terme de $\hat{h}_n^{2,T}$ peut changer la valeur du mise de façon appréciable, tandis qu'un changement substantiel de ρ dans $\hat{h}_n^{(\lambda_i)}$ ne donne qu'un changement minime du mise.

* On peut toujours se ramener à $E = [0,1]$ en utilisant la transformation décrite dans le chapitre II, B.II.4.

b. Utilisations des données afin de déterminer les variances

a priori $\sigma_i^2 = 1/\pi_i$ (dans le cas du système orthonormal trigonométrique sur $[0,1]$).

. On se place dans le cas où les \tilde{a}_i sont indépendants, et tels que :

- . $E(\tilde{a}_i) = 0$
- . $\text{var}(\tilde{a}_i) = \sigma_i^2 = \frac{1}{\pi_i}$.

. Comme $\hat{a}_i = \frac{1}{n} \sum_{j=1}^n \psi_i(X_j)$ est asymptotiquement distribué comme un loi normale de moyenne a_i et de variance $\frac{1}{n}$. On a :

$$\left| \begin{array}{ll} \text{cov}(\hat{a}_i, \hat{a}_j) = \frac{1}{n} + \frac{n-1}{n} \sigma_i^2 & \text{si } i = j \\ \text{cov}(\hat{a}_i, \hat{a}_j) = 0 & \text{si } i \neq j \end{array} \right.$$

Le symbole " \sim " indique que l'élément est cette fois aléatoire.

. Soit $\gamma_i = \sigma_{2i-1}^2 = \sigma_{2i}^2 \quad i = 1, 2, \dots$

. Alors il est immédiat que :

$$\tilde{\xi}_i = \left\{ \frac{1}{2} \left[(\hat{a}_{2i-1})^2 + (\hat{a}_{2i})^2 \right] - \frac{1}{n} \right\}$$

est un estimateur sans biais de $\frac{n-1}{n} \gamma_i$.

Remarques :

1. D'après [15], l'utilisation de précisions a priori $\pi_i = 5^i$ donne des estimateurs assez réguliers.

2. D'autre part, si $\pi_i \rightarrow \infty$ rapidement, on peut en déduire d'après [4] l'existence d'un unique mode pour l'estimateur.

c. Estimation de la hauteur des vagues [14].

. Les hauteurs de vagues successives sont indépendantes et forment un échantillon d'une variable aléatoire ayant une distribution continue à estimer.

. Les données s'ajustent à une distribution de Weibull. En utilisant la méthode du maximum de vraisemblance, on obtient que F_0 est égale à :

$$F_0(Z) = 1 - \exp \left\{ - \left(\frac{Z}{0,2075} \right)^{2,36} \right\},$$

où F_0 désigne la fonction de répartition correspondant à f_0 ,
première estimation de f .

. On transforme alors les données à l'aide de F_0 (ce qui nous ramène à une estimation sur $[0,1]$, où l'on peut utiliser le système trigonométrique) et on utilise l'estimateur $\hat{h}_n^{(\lambda_i)}$ du a . (ρ étant déterminé par la méthode du b).

. γ_i est calculé à l'aide de $\tilde{\xi}_i$ et on en déduit la valeur de ρ car $\frac{\gamma_i}{\gamma_{i+1}} = \rho$.

. On peut donner le tableau suivant de valeurs de γ_k , pour différentes valeurs de k et pour les calculs menés sous gamma 1, 2, 3, 4.

k	gamma 1	gamma 2	gamma 3	gamma 4
1	0,2	0,3	0,3	0,3
2	0,1379	0,15	0,1	0,06
3	0,0951	0,075	0,03333	0,012
4	0,0656	0,0375	0,01111	0,0024
5	0,0452	0,0188	0,0037	0,0005
6	0,0312	0,0094	0,00123	0,0001
7	0,0215	0,0047	0,00041	
8	0,0148	0,0023	0,0001	
9	0,0102	0,0012		
10	0,0071	0,0006		
11	0,0049	0,0003		
12	0,0034	0,0001		
13	0,0023	0,0001		
14	0,0016			
15	0,0011			
16	0,0008			
17	0,0005			
18	0,0004			
19	0,0002			
20	0,0001			

. Pour gamma 1, on obtient $\rho = 1,45$. La courbe obtenue (voir V) oscille trop pour que l'on ait un estimateur réaliste de la vraie densité.

. Pour gamma 2, on obtient $\rho = 2$ et cela correspond à une hypothèse plus sévère de régularité.

. Pour gamma 3, on obtient $\rho = 3$. Les oscillations ont pratiquement disparu.

. Pour gamma 4, on obtient $\rho = 5$, ce qui nous donne l'estimateur unimodal souhaité.

Remarque 1. - Une étude analogue a été faite par Hominal [32] avec l'estimateur \hat{h}_n^1 dans le cas du noyau unité.

L'estimateur de Whittle-Brunk donne de meilleurs résultats, car l'estimateur est positif dans le cas d'une bonne précision π_i , ce qui n'est pas le cas de [32].

Remarque 2. - On pourra trouver dans [15] d'autres courbes construites à l'aide de $\hat{h}_n^{(\lambda_i)}$ avec $\pi_i = 5^i$.

Une comparaison avec les courbes de [45] est difficile, cependant il semble que dans le cas de la courbe 1. (Estimation de $f(x) = \frac{1}{2} e^{-x/2}$ par les fonctions d'hermite). L'estimateur de Whittle-Brunk aurait donné une courbe plus régulière, les $\lambda_i(n)$ apparaissant comme des coefficients qui atténuent les oscillations.

3. Estimateurs de Whittle-Wahba, d'après [60], [61].

L'étude est faite sur $E = [0,1]$ et le système orthonormal utilisé est celui des fonctions trigonométriques, avec

$$e_\nu(t) = e^{2i\pi\nu t} ;$$

et soit la suite (a_ν) , telle que : $f(t) = 1 + \sum_{\nu=-\infty}^{+\infty} a_\nu e_\nu(t)$, où f est une densité.

Les estimateurs considérés par Wahba sont définis par :

$$\hat{h}_n^{\lambda, m}(x^n, x) = 1 + \sum_{\nu} \frac{\hat{a}_\nu}{[1 + \lambda(2\pi\nu)^{2m}]^*} \cdot e_\nu(x)$$

avec

$$\sum_{\nu} = \sum_{\substack{\nu=-n/2 \\ \nu \neq 0}}^{+n/2} \quad (n \text{ pair})$$

$$\hat{a}_\nu = \frac{1}{n} \sum_{j=1}^n e_\nu^*(X_j) \quad (e^* \text{ désigne le conjugué})$$

Ces estimateurs correspondent donc à une variante des estimateurs précédents $(\hat{h}_n^{(\lambda, 1)})$, à la différence près que la série est tronquée et que les termes qui servent à pondérer les éléments $e_\nu(x)$ dépendent de 2 paramètres, λ et m .

Avant de préciser la détermination de λ et m , nous allons énoncer un théorème qui montre que cette nouvelle famille d'estimateurs converge plus rapidement que $\hat{h}_n^{2, T}$ sur $W^{m'}$, $m' \geq 1$.

a. Vitesse de convergence au sens du Mise de $\hat{h}_n^{\lambda, m}$.

Théorème [60], [61].-

Soit $f \in W^{m'}$ et telle que son support soit $[0, 1]$.

Alors pour tout m tel que $1 \leq m \leq m'$ si $\lambda = O(n^{-2m/2m+1})$

$$E_f \left[\int_0^1 \left[\hat{h}_n^{\lambda, m}(x^n, x) - f(x) \right]^2 dx \right] = O(n^{-\frac{2m}{2m+1}}).$$

Démonstration :

Ce résultat peut s'étendre aux fonctions de $W^{m'}$ telles que :

$$f^{(\nu)}(0) = f^{(\nu)}(1) \quad \nu = 0, \dots, m'-1.$$

La condition $\text{supp } f = [0,1]$ est plus restrictive mais permet une comparaison avec les résultats du chapitre II. B.II.4.

Soit :

$$\int_0^1 \left[\widehat{h}_n^{\lambda, m}(x^n, x) - f(x) \right]^2 dx = \sum_* \left| \frac{\lambda_\nu}{\lambda_\nu + \lambda} \widehat{a}_\nu - a_\nu \right|^2 + \sum_{|\nu| > \frac{n}{2}} \frac{n}{2} |a_\nu|^2 \quad \text{où}$$

$$\lambda_\nu = (2\pi\nu)^{-2m}.$$

Comme $E_f[\widehat{a}_\nu - a_\nu] = 0$
 $E[|\widehat{a}_\nu - a_\nu|^2] = \frac{1}{n} (1 - |a_\nu|^2).$

On obtient :

$$\begin{aligned} E_f \int_0^1 \left[\widehat{h}_n^{\lambda, m}(x^n, x) - f(x) \right]^2 dx &= \sum_* \left(\frac{\lambda_\nu}{\lambda_\nu + \lambda} \right)^2 \cdot E_f(|\widehat{a}_\nu - a_\nu|^2) + \sum_* \left(\frac{\lambda}{\lambda_\nu + \lambda} \right)^2 |a_\nu|^2 + \sum_{|\nu| > \frac{n}{2}} \frac{n}{2} |a_\nu|^2 \\ &= \sum_* \left\{ \left(\frac{\lambda}{\lambda_\nu + \lambda} \right)^2 - \frac{1}{n} \left(\frac{\lambda_\nu}{\lambda_\nu + \lambda} \right)^2 \right\} |a_\nu|^2 + \frac{1}{n} \sum_* \left(\frac{\lambda_\nu}{\lambda_\nu + \lambda} \right)^2 + \sum_{|\nu| > \frac{n}{2}} \frac{n}{2} |a_\nu|^2 \end{aligned}$$

et $\sum_* \left(\frac{\lambda}{\lambda_\nu + \lambda} \right)^2 |a_\nu|^2 \leq \lambda \sum_{-\infty}^{\infty} \frac{|a_\nu|^2}{\lambda_\nu} = \lambda \theta_m$

car : si $f \in W^{m'}$ et $f^{(\nu)}(0) = f^{(\nu)}(1) \quad \nu = 0, \dots, m'-1$

alors

$$\sum_{\nu=-\infty}^{+\infty} (2\pi\nu)^{2m'} \cdot |a_\nu|^2 = \theta_m < \infty$$

$$\frac{1}{n} \sum_* \left(\frac{\lambda_\nu}{\lambda_\nu + \lambda} \right)^2 = \frac{1}{n} \sum_* \frac{1}{[1 + \lambda(2\pi\nu)^{2m}]^2}$$

$$\leq \frac{2}{n} \int_0^\infty \frac{dx}{[1 + \lambda(2\pi x)^{2m}]^2} = \frac{km}{n\lambda^{1/2m}}$$

avec $k_m = \frac{1}{\pi} \int_0^\infty \frac{dx}{(1+x)^{2m}^2}$

$$|v| \sum_{\frac{n}{2}} |a_v|^2 \leq \frac{1}{(\pi n)^{2m}} |v| \sum_{\frac{n}{2}} (2\pi v)^{2m'} |a_v|^2 \leq \frac{\theta m'}{(\pi n)^{2m}}$$

d'où le résultat.

Remarque : L'estimateur $\hat{h}_n^{2,T}$, dans les mêmes conditions, ne réalise que :

$$E_f \left[\int_0^1 (\hat{h}_n^{2,T}(x^n, x) - f(t))^2 dt \right] = O(n^{-\frac{2m-1}{2m}})$$

Voir Chap. II, B.II.4.

b. Détermination de λ et m .

λ et m sont obtenues en minimisant l'expression suivante :

$$\hat{T}_{n,m}(\lambda) = \frac{n}{n-1} \sum_* \left\{ \left(\frac{\lambda}{\lambda_v + \lambda} \right)^2 - \frac{1}{n} \left(\frac{\lambda_v}{\lambda_v + \lambda} \right)^2 \right\} \cdot |\hat{a}_v|^2 - \frac{1}{n-1} \sum_* \left\{ \left(\frac{\lambda}{\lambda_v + \lambda} \right)^2 - \left(\frac{\lambda_v}{\lambda_v + \lambda} \right)^2 \right\}.$$

En effet, en posant :

$$T_{n,m}(\lambda) = \int_0^1 (\hat{h}_n^{\lambda,m}(x^n, x) - f(x))^2 dx,$$

on obtient le résultat suivant :

Théorème.-

Soit $f \in W^{m'}$ telle que $\text{supp } f = [0, 1]$. Alors

$$E_f [\hat{T}_{n,m}(\lambda)] = E_f [T_{n,m}(\lambda)] + O\left(\frac{1}{n^{2m'}}\right) \quad \left| \begin{array}{l} \lambda > 0 \\ m \geq 1 \end{array} \right.$$

Démonstration :

Elle est analogue à la précédente. On utilise aussi le fait que :

$$\begin{aligned} E(|\hat{a}_v|^2) &= E\left(\frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n e_v(X_j) e_v^*(X_k)\right) \\ &= \frac{n}{n^2} \int_0^1 f(x) dx + \frac{n^2 - n}{n^2} \int_0^1 e_v(x) f(x) dx \int_0^1 e_v^*(x) f(x) dx \end{aligned}$$

$$= \frac{1}{n} + \frac{n-1}{n} |a_v|^2 .$$

Conclusion :

. Cette variante de l'estimateur de Whittle possède de très bonnes propriétés de convergence au sens du Mise.

Le problème de la tronquature est supprimé ; (il y a cependant une série de n termes à calculer au lieu de $r(n)$ dans le cas de l'estimateur \hat{h}_n^2) et d'autre part, il y a la détermination de λ et m qui nécessite un calcul assez délicat.

4. Méthode d'estimation préconisée.

Les estimateurs \hat{h}_n de noyaux :

$$K_n(x,t) = \sum_{i=0}^{\infty} \lambda_i(n) e_i(x) e_i(t)$$

possèdent donc de très bonnes propriétés pour différentes suites $(\lambda_i(n))$. Plus particulièrement, les estimateurs $\hat{h}_n^{(\lambda_i)}$ et $\hat{h}_n^{\lambda,m}$, obtenus en 2. et 3. constituent des améliorations de l'estimateur \hat{h}_n^2 .

La seule objection qui reste à propos de ces méthodes est l'importance du choix de f_0 , première estimation de f (f_0 densité).

On pourra suivre les étapes suivantes :

1ère étape : Choix de f_0 à l'aide de l'estimateur \hat{h}_n^1 ou \hat{h}_n^2 , ou de l'estimateur par la méthode du maximum de vraisemblance.

2ème étape : Transformation des données à l'aide de F_0 (*)

3ème étape :

. Détermination des précisions π_i et utilisation de l'estimateur $\hat{h}_n^{(\lambda_i)}$ de Brunk ou :

. Détermination des paramètres λ et m et utilisation de l'estimateur $\hat{h}_n^{\lambda,m}$ de Wahba.

(*) F_0 fonction de répartition correspondant à f_0 .

IV - Détermination d'une nouvelle fonction de perte,
qui tient compte de l'information sur F.

. Soit \hat{h}_n un estimateur de $f \in F$ ($F \subset L^2$). Nous avons principalement utilisé dans le chapitre II la fonction de perte L_2 , avec :

$$L_2(\hat{h}_n, f) = \|\hat{h}_n - f\|_2^2 = \sum_{i=1}^{\infty} (\hat{a}_i - a_i)^2$$

avec (e_i) base de L^2

$$\left| \begin{array}{l} \hat{a}_i = \int \hat{h}_n(x^n, x) e_i(x) dx \\ a_i = \int f(x) e_i(x) dx \end{array} \right.$$

. Cette fonction de perte dépend directement de l'espace de Hilbert.

Comme chaque e_i est plus ou moins important dans l'estimation on convient de lui attacher un coefficient λ_i ($0 < \lambda_i \leq 1$).

. Ce coefficient ou poids λ_i étant plus petit que l'élément, est "intéressant".

Par exemple, si f est supposée régulière, (c'est-à-dire ne comporte pas trop d'oscillations) et si e_i est aussi régulière, alors λ_i sera petit.

Soit alors :

$$L_{(\lambda_i)}(\hat{h}_n, f) = \sum_{i=1}^{\infty} \lambda_i (\hat{a}_i - a_i)^2$$

. Dans le cas où tous les $\lambda_i = 1 \forall i \in \mathbb{N}$, on retrouve la perte classique qui correspond à un problème sans information.

1. Etude dans le cas de l'estimateur $\hat{h}_n^{2,T}$.

$$\hat{h}_n^{2,T}(x^n, x) = \sum_{i=1}^{\infty} \hat{a}_i e_i(x) \quad \text{avec} \quad \begin{cases} \hat{a}_i = \frac{1}{n} \sum_{j=1}^n e_i(x_j) & \text{pour } i \leq r(n) \\ a_i = 0 & \text{pour } i > r(n) \end{cases}$$

Donc
$$L_{(\lambda_i)}(\hat{h}_n^{2,T}, f) = \sum_{i=1}^{r(n)} \lambda_i (a_i - a_i)^2 + \sum_{i>r(n)} \lambda_i a_i^2 .$$

Dans le cas trigonométrique, on peut supposer cf [6] :

$$\exists M \text{ Constante t.q. } \int e_i^2(x) f(x) dx \leq M < \infty .$$

Alors :

$$\begin{aligned} E_f \left[L_{(\lambda_i)}(\hat{h}_n^{2,T}, f) \right] &= \frac{1}{n} \sum_{i=1}^{r(n)} \lambda_i \left[\int e_i^2(x) f(x) dx - a_i^2 \right] + \sum_{i>r(n)} \lambda_i a_i^2 \\ &< \frac{1}{n} \sum_{i=1}^{r(n)} \lambda_i \int e_i^2(x) f(x) dx + \sum_{i>r(n)} \lambda_i a_i^2 \\ &< \frac{M}{n} \sum_{i=1}^{r(n)} \lambda_i + \sum_{i>r(n)} \lambda_i a_i^2 . \end{aligned}$$

Exemple 1.

$$f \in F \subset W_{\pi}^m \quad \text{et} \quad \lambda_i = \frac{1}{i^{1/2}} .$$

Alors :

$$E_f \left[L_{(\lambda_i)}(\hat{h}_n^{2,T}, f) \right] < \frac{M}{n} \sum_{i=1}^{r(n)} \frac{1}{i^{1/2}} + \sum_{i>r(n)} \frac{a_i^2}{i^{1/2}} .$$

D'autre part, si $f \in W_{\pi}^m$ $(a_i)^2 = O(\frac{1}{i^{2m}})$ et, en majorant les 2 séries par les intégrales correspondantes, on obtient, avec

$$r(n) = O(n^{1/2m}) : \quad \boxed{E_f L_{(\lambda_i)}(\hat{h}_n^{2,T}, f) = O(n^{-\frac{4m-1}{4m}})} .$$

Exemple 2.

$$f \in F \subset W^m \quad \text{et} \quad \lambda_i = \frac{i-1}{i+1} .$$

Par un calcul analogue à celui de l'exemple 1, on obtient :

$$E_f [L_{(\lambda_i)}(\hat{h}_n^{2,T}, f)] = O(n^{-\frac{2m-1}{2m}}) \quad \text{pour} \quad r(n) = O(n^{1/2m}).$$

Vitesse égale à celle obtenue pour le mise dans la même situation.

Exemple 3.

$$f \in F \subset W^m \quad \text{et} \quad \lambda_i = \frac{1}{i^\alpha} \quad \alpha > 1,$$

information qui apparaît idéale. En effet :

Dans le cas de cette information, on a, pour $r(n) = O(n^{1/2m+\alpha-1})$ la vitesse maximale $O(n^{-1})$. En effet :

$$\sum_{i=1}^{r(n)} \lambda_i = \sum_{i=1}^{r(n)} \frac{1}{i^\alpha} < 1 + \int_1^{r(n)} x^{-\alpha} dx = 2 - \frac{1}{r^{\alpha-1}(n)}$$

$$\text{et :} \quad \sum_{i>r(n)} \frac{1}{i^{2m+\alpha}} < \int_{r(n)}^{\infty} \frac{1}{x^{2m+\alpha}} dx = \frac{1}{r^{2m+\alpha-1}(n)} \quad \text{C.Q.F.D.}$$

N.B. Les résultats sont aussi satisfaisants pour l'estimateur $\hat{h}_n^{2,H}$.

2. Etude dans le cas de l'estimateur \hat{h}_n^1 .

$$\text{Avec} \quad \hat{h}_n^1(x^n, x) = \frac{1}{n \cdot h(n)} \sum_{i=1}^n K\left(\frac{x-x_i}{h(n)}\right) .$$

Nous allons considérer le cas du noyau unité avec :

$$K(y) = 1 \quad \text{si} \quad |y| < \frac{1}{2} .$$

C'est un noyau de Parzen d'ordre 2.

donc :

$$h \int_{-\frac{1}{2}}^{+\frac{1}{2}} e_k(uh+x_i) du = \frac{h}{\sqrt{\pi}} \sin kx_i \frac{\sin kh/2}{kh/2}$$

On peut conclure que :

$$\hat{a}_{2k-1} = \frac{1}{nh} \sum_{i=1}^n \frac{h}{\sqrt{\pi}} \cos kx_i \frac{\sin kh/2}{kh/2} = \frac{1}{n\sqrt{\pi}} \frac{\sin kh/2}{kh/2} \sum_{i=1}^n \cos kx_i$$

$$\hat{a}_{2k} = \frac{1}{nh} \sum_{i=1}^n \frac{h}{\sqrt{\pi}} \sin kx_i \frac{\sin kh/2}{kh/2} = \frac{1}{n\sqrt{\pi}} \frac{\sin kh/2}{kh/2} \sum_{i=1}^n \sin kx_i$$

b. Soit à calculer : $E(\hat{a}_k)$, 2 cas.

$$E(\hat{a}_{2k-1}) = \frac{1}{n} \frac{\sin kh/2}{kh/2} \sum_{i=1}^n \int \frac{\cos kx_i}{\sqrt{\pi}} f(x_i) dx_i$$

$$E(\hat{a}_{2k-1}) = \frac{\sin kh/2}{kh/2} a_{2k-1}$$

où : a_{2k-1} est le coefficient de Fourier de f correspondant à $e_{2k-1}(t)$.

$$E(\hat{a}_{2k}) = \frac{\sin kh/2}{kh/2} a_{2k}$$

c. Soit calculer $E(\hat{a}_k^2)$, 2 cas.

$$E(\hat{a}_{2k-1}^2) = \frac{1}{n^2} \left(\frac{\sin kh/2}{kh/2} \right)^2 E \left[\sum_{i,j=1}^n \frac{\cos kx_i}{\sqrt{\pi}} \frac{\cos kx_j}{\sqrt{\pi}} \right]$$

$$E(\hat{a}_{2k-1}^2) = \frac{1}{n^2} \left(\frac{\sin kh/2}{kh/2} \right)^2 \left[n \int \frac{\cos^2 kx f(x) dx}{\pi} + n(n-1) a_{2k-1}^2 \right]$$

$$E(\hat{a}_{2k-1}^2) = \left(\frac{\sin kh/2}{kh/2} \right)^2 \left[\int \frac{\cos^2 kx f(x) dx}{n\pi} + \frac{n-1}{n} a_{2k-1}^2 \right]$$

de la même façon :

$$E(\hat{a}_{2k}^2) = \left(\frac{\sin kh/2}{kh/2} \right)^2 \left[\int \frac{\sin^2 kx f(x) dx}{n\pi} + \frac{n-1}{n} a_{2k}^2 \right]$$

d. Calcul de $E_f[\bar{L}(\lambda_i)(\hat{h}_n^1, f)]$.

$$E_f[\bar{L}(\lambda_i)(\hat{h}_n^1, f)] = \sum_{k=1}^{\infty} \lambda_k [E_f(\hat{a}_k^2) - 2a_k E_f(\hat{a}_k) + a_k^2]$$

puis, à l'aide de b. et c. :

$$\begin{aligned} " &= \sum_{k=1}^{\infty} \lambda_{2k} \left[\left(\frac{\sin kh/2}{kh/2} \right)^2 \left\{ \frac{\int \sin^2 kx f(x) dx}{n} + \frac{n-1}{n} a_{2k}^2 \right\} - 2 \frac{\sin kh/2}{kh/2} a_{2k} + a_{2k}^2 \right] + \\ &\quad \sum_{k=1}^{\infty} \lambda_{2k-1} \left[\left(\frac{\sin kh/2}{kh/2} \right)^2 \left\{ \frac{\int \cos^2 kx f(x) dx}{n} + \frac{n-1}{n} a_{2k-1}^2 \right\} - 2 \frac{\sin kh/2}{kh/2} a_{2k-1} + a_{2k-1}^2 \right]. \end{aligned}$$

$$\begin{aligned} E_f[\bar{L}(\lambda_i)(\hat{h}_n^1, f)] &= \sum_{k=1}^{\infty} \lambda_{2k} a_{2k}^2 \left[\left(\frac{\sin kh/2}{kh/2} \right)^2 \cdot \frac{n-1}{n} - 2 \left(\frac{\sin kh/2}{kh/2} \right) + 1 \right] + \\ &\quad \sum_{k=1}^{\infty} \lambda_{2k-1} a_{2k-1}^2 \left[\left(\frac{\sin kh/2}{kh/2} \right)^2 \cdot \frac{n-1}{n} - 2 \left(\frac{\sin kh/2}{kh/2} \right) + 1 \right] + \\ &\quad \sum_{k=1}^{\infty} \left(\frac{\sin kh/2}{kh/2} \right)^2 \left[\lambda_{2k} \frac{\alpha_k}{n} + \lambda_{2k-1} \frac{\beta_k}{n} \right] \end{aligned}$$

en posant : $\alpha_k = \int \frac{\sin^2 kx}{\pi} f(x) dx$; $\beta_k = \int \frac{\cos^2 kx}{\pi} f(x) dx$.

$$\begin{aligned} E_f[\bar{L}(\lambda_i)(\hat{h}_n^1, f)] &= \underbrace{\sum_{k=1}^{\infty} \lambda_{2k} a_{2k}^2 \left[\frac{\sin kh/2}{kh/2} - 1 \right]^2}_{(1)} + \underbrace{\sum_{k=1}^{\infty} \lambda_{2k-1} a_{2k-1}^2 \left[\frac{\sin kh/2}{kh/2} - 1 \right]^2}_{(2)} + \\ &\quad \underbrace{\sum_{k=1}^{\infty} \left[\lambda_{2k} \left(\frac{\alpha_k}{n} - \frac{a_{2k}^2}{n} \right) + \lambda_{2k-1} \left(\frac{\beta_k}{n} - \frac{a_{2k-1}^2}{n} \right) \right] \left(\frac{\sin kh/2}{kh/2} \right)^2}_{(3)}. \end{aligned}$$

Exemple : $f \in W^m$ et $\lambda_i = \frac{1}{i^{1/2}}$ $m \geq 2$.

On fait alors l'hypothèse qu'il existe une constante C positive et finie telle que : $|a_i|^2 < \frac{C}{i^{2m}}$. On obtient que :

$$E_f[\bar{L}(\lambda_i)(\hat{h}_n^1, f)] \leq (1) + (2) + (3) \quad \text{avec}$$

$$\textcircled{1} = \sum_{k=1}^{\infty} \frac{c}{(2k)^{2m+1/2}} \left[\frac{\sin kh/2}{kh/2} - 1 \right]^2 \quad \text{pour } k=1 \text{ et } h \rightarrow 0,$$

le premier terme de la série est nul, donc ,

$$\textcircled{1} \leq \int_1^{\infty} \frac{c'}{x^{2m+1/2}} \left[\frac{\sin hx/2}{xh/2} - 1 \right]^2 dx \quad \text{avec } hx = X.$$

$$\textcircled{1} \leq \int_h^{\infty} \frac{c'h^{2m - \frac{1}{2}}}{x^{2m+1/2}} \left[\frac{\sin X/2}{X/2} - 1 \right]^2 dX$$

$$\textcircled{1} \leq \left(\frac{\sin h/2}{h/2} - 1 \right)^2 \int_h^{\infty} \frac{c'h^{2m - \frac{1}{2}}}{x^{2m+1/2}} dX \quad \text{d'après la formule de la moyenne}$$

Pour $h \rightarrow 0$ $\sin h = h - \frac{h^3}{6} \dots$ et on obtient alors :

$$\textcircled{1} = O(h^4) \quad \text{car } \int_h^{\infty} \frac{c'h^{2m - \frac{1}{2}}}{x^{2m+1/2}} dX \text{ est convergente}$$

$$\textcircled{2} = O(h^4) \quad \text{de la même façon.}$$

Pour $\textcircled{3}$, on a en fait 4 séries convergentes.

Celle qui donnera la partie du risque la moins rapidement convergente est la suivante :

$$\sum_{k=1}^{\infty} \frac{(\sin kh/2)^2}{kh/2} \frac{\lambda 2k^{\alpha_k}}{n}$$

$$\text{Comme } \alpha_k = \int \frac{\sin^2 kx}{\pi} \cdot f(x) dx \leq M,$$

cette série est majorée par : (B étant une constante)

$$\frac{B}{n} + \frac{M}{n \cdot \pi} \int_1^{\infty} \left(\frac{\sin xh/2}{xh/2} \right)^2 \frac{1}{(2x)^{1/2}} dx .$$

Avec le changement de variable $X = xh$; on obtient :

$$\int_1^{\infty} \left(\frac{\sin xh/2}{xh/2} \right)^2 \frac{1}{x^{1/2}} dx = \int_h^{\infty} \left(\frac{\sin X/2}{X/2} \right)^2 \frac{1}{X^{1/2}} \frac{dX}{h^{1/2}} .$$

Alors :
$$\sum_{k=1}^{\infty} \left(\frac{\sin kh/2}{kh/2} \right)^2 \cdot \frac{\lambda 2k^\alpha k}{n\pi} < \frac{M'}{n h^{1/2}} \int_h^{\infty} \left(\frac{\sin X}{X} \right)^2 \frac{1}{X^{1/2}} dX + \frac{B}{n}$$

d'après la deuxième formule de la moyenne :

$$" \quad " \quad < \frac{M'}{nh} \int_h^{\infty} \left(\frac{\sin X}{X} \right)^2 dX + \frac{B}{n} \quad \begin{matrix} h > 0 \\ h \rightarrow 0 \end{matrix}$$

et $\int_h^{\infty} \left(\frac{\sin X}{X} \right)^2 dX$ est convergente.

On obtient alors, en prenant $h(n) = O(n^{-1/5})$

$$\boxed{E_f \left[L_{(\lambda_i)}(\hat{h}_n^1, f) \right] = O(n^{-\frac{4}{5}}) .}$$

$f \in F \subset W_{\pi}^m$ en supposant qu'il existe de plus une constante C telle que $a_i^2 \leq \frac{C}{i^{2m}}$.

Conclusion.

. Pour le mise classique, nous avons obtenu pour le même ensemble de densités W_{π}^2 ($m = 2$)

$$E_f \left[L(\hat{h}_n^1, f) \right] = O(n^{-\frac{4}{5}}) \quad (\hat{h}_n^1 \text{ associé au noyau nuité})$$

$$E_f \left[L(\hat{h}_n^2, f) \right] = O(n^{-\frac{3}{4}}) .$$

. Pour le nouveau risque, nous obtenons, pour $\lambda_i = i^{-\frac{1}{2}}$:

$$E_f \left[\bar{L}(\lambda_i) (\hat{h}_n^1, f) \right] = O(n^{-\frac{4}{5}})$$
$$E_f \left[\bar{L}(\lambda_i) (\hat{h}_n^1, f) \right] = O(n^{-\frac{7}{8}}).$$

Avec ce nouveau critère de comparaison sur W_π^2 , l'estimateur $\hat{h}_n^{2,T}$ converge plus rapidement que \hat{h}_n^1 pour K noyau unité (de Parzen d'ordre 2).

L'estimateur \hat{h}_n^1 n'a pas utilisé l'information et sa vitesse est restée bloquée.

Cette nouvelle fonction de perte possède cependant 2 défauts.

- Autant le calcul est aisé pour des estimateurs \hat{h}_n à noyaux de la forme :

$$K_{r(n)}(x, t) = \sum_{j=1}^{r(n)} \alpha_j e_j(x) e_j(t),$$

autant il est compliqué pour d'autres noyaux.

- La détermination des λ_i ne semble pas facile.

V - Courbes.

On trouvera :

• Les courbes 1 à 7 qui illustrent la méthode des fonctions orthogonales : cas des fonctions d'hermite et des fonctions de Laguerre d'après [45].

• Les courbes 8 qui illustrent l'estimateur de Whittle-Fellner, dans le cas où seul le premier moment de la loi a priori est connu.

La figure 1 suggère 3 modes qui se retrouvent dans la figure 4.

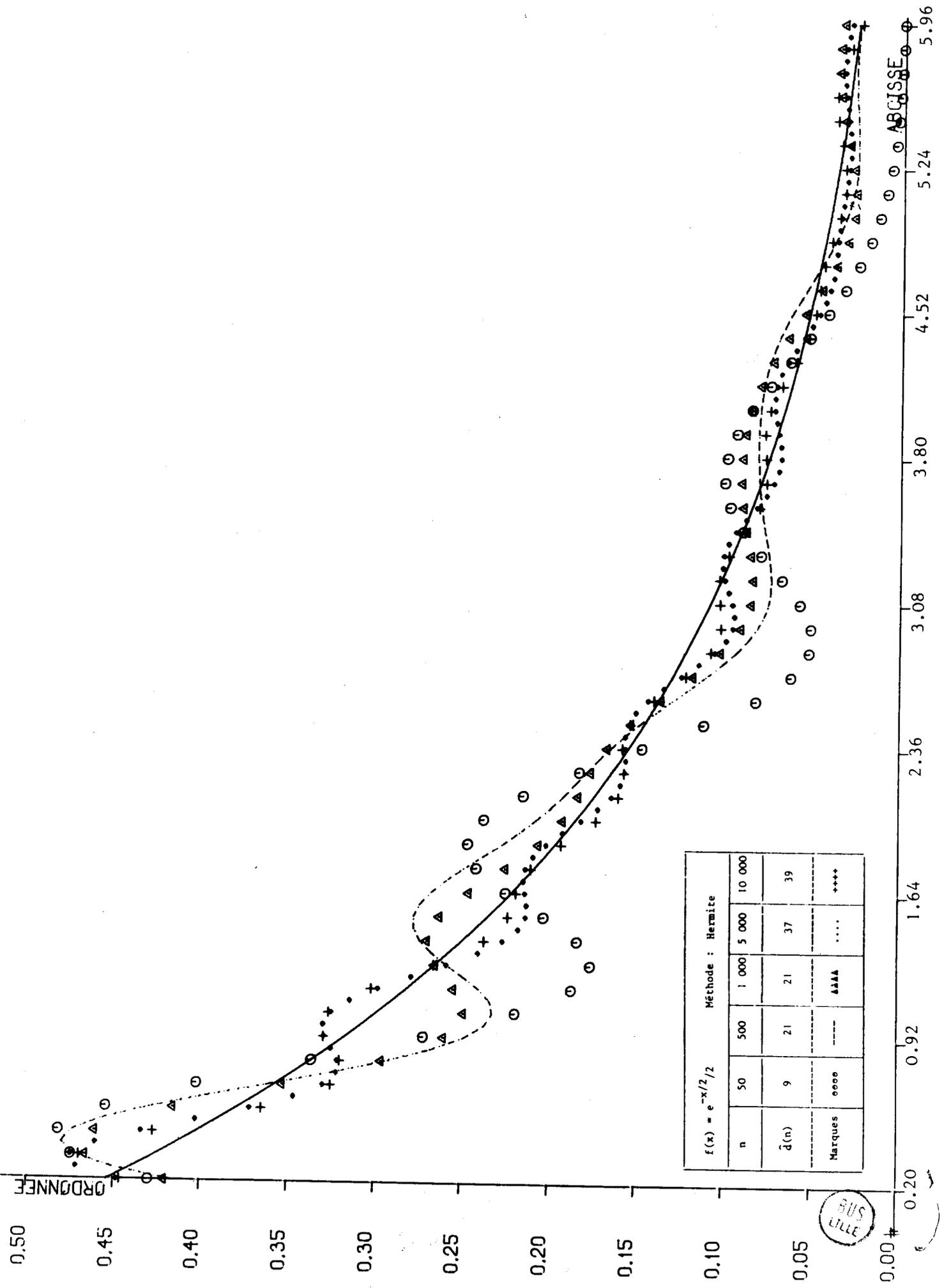
On remarquera dans les figures 2 et 3 des régions où l'estimateur est négatif.

La ligne en pointillés représente $f_0(x) = E_p[\hat{f}(x)]$.

La transformation appropriée de la figure 4 est la transformation par la fonction $F_0(Z) = \int_{-\infty}^Z f_0(x) dx$ d'après [23].

• Les courbes 9 et 10 illustrent l'estimateur de Whittle-Brunk d'après [14] étudié dans le chapitre III, III.

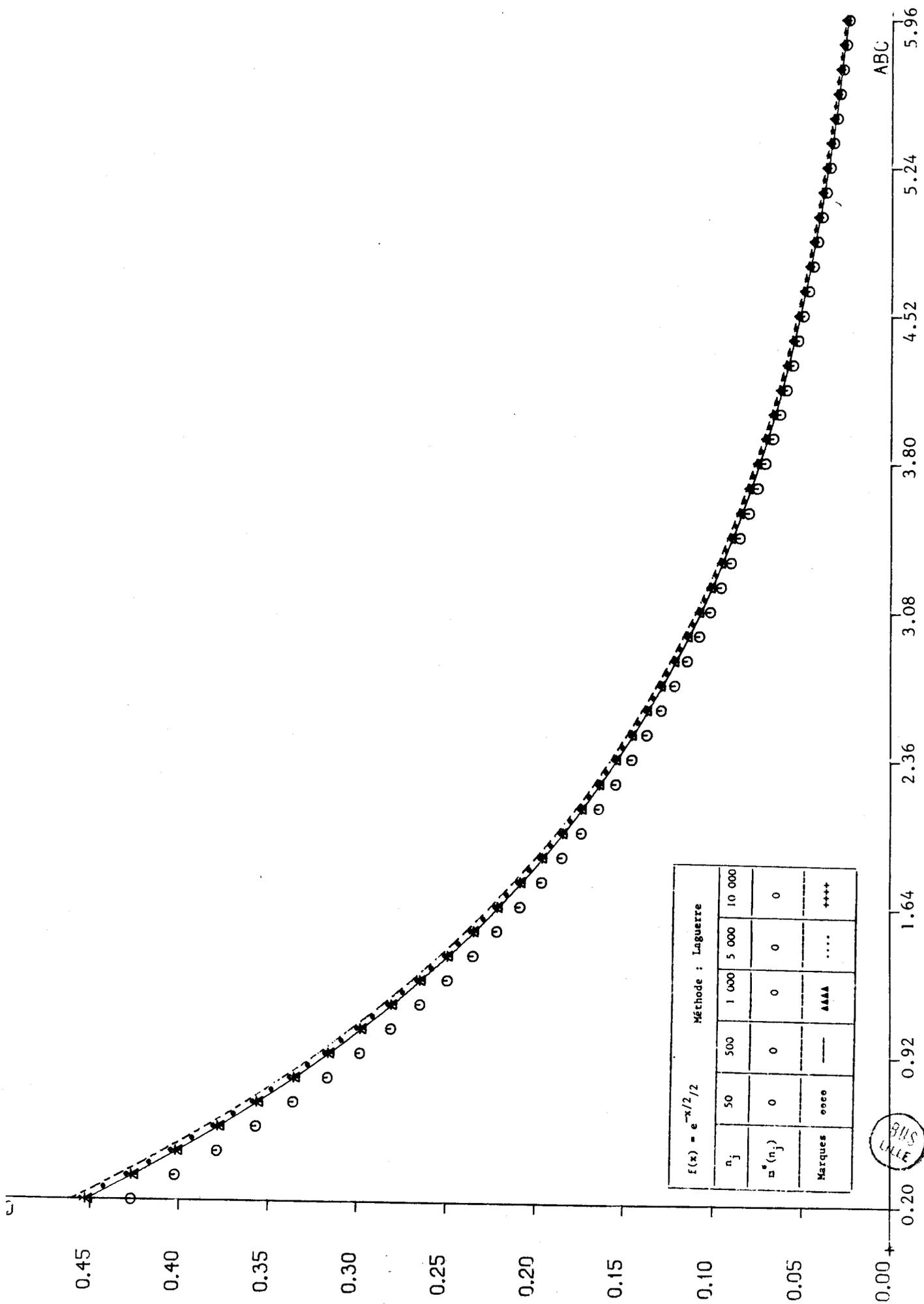
Courbe 1.



$f(x) = e^{-x^2/2}$		Méthode : Hermite			
n	50	500	1 000	5 000	10 000
$\hat{d}(n)$	9	21	21	37	39
Marques	oooo	----	AAAA	++++

BUS LILLE

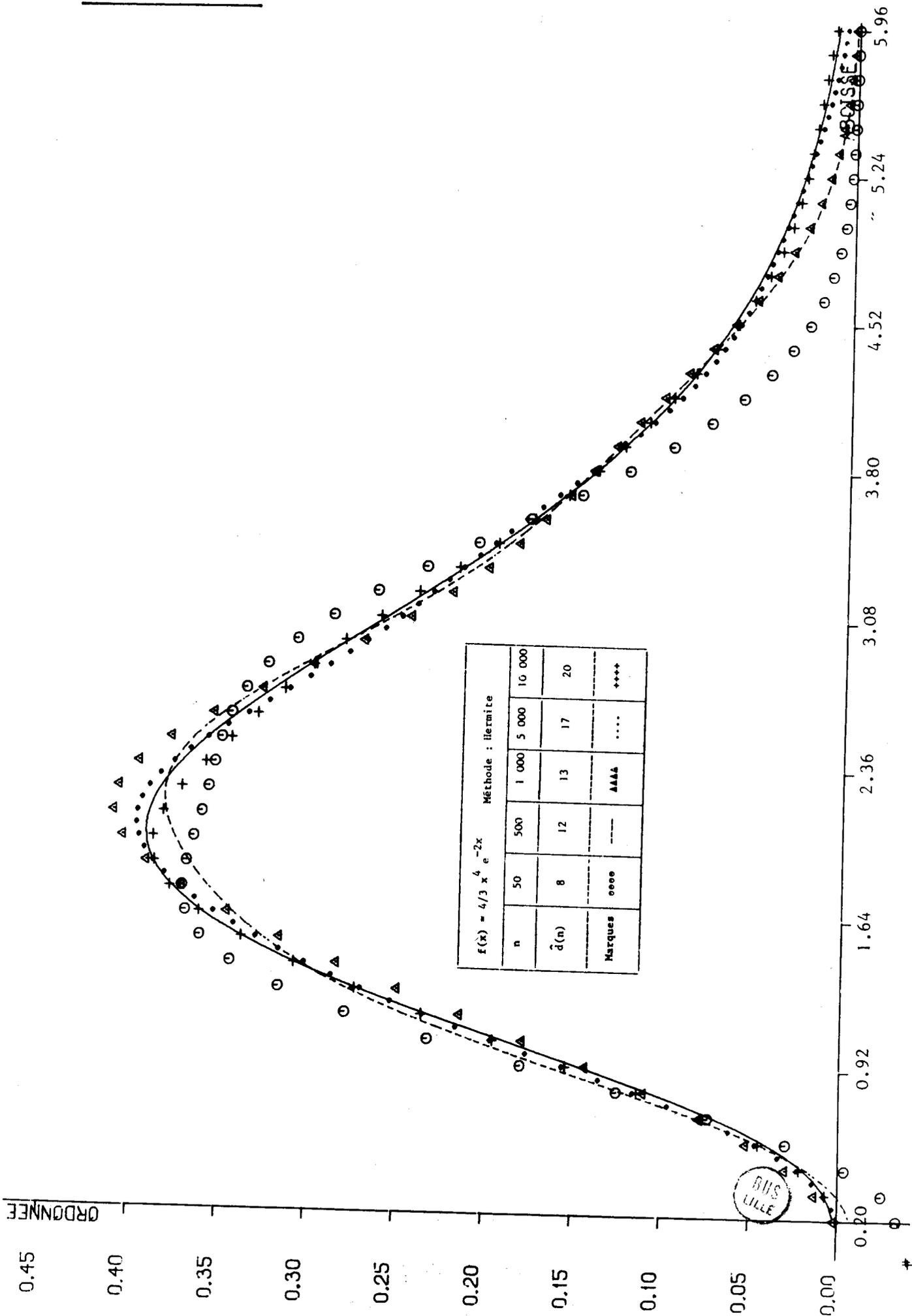
Courbe 2



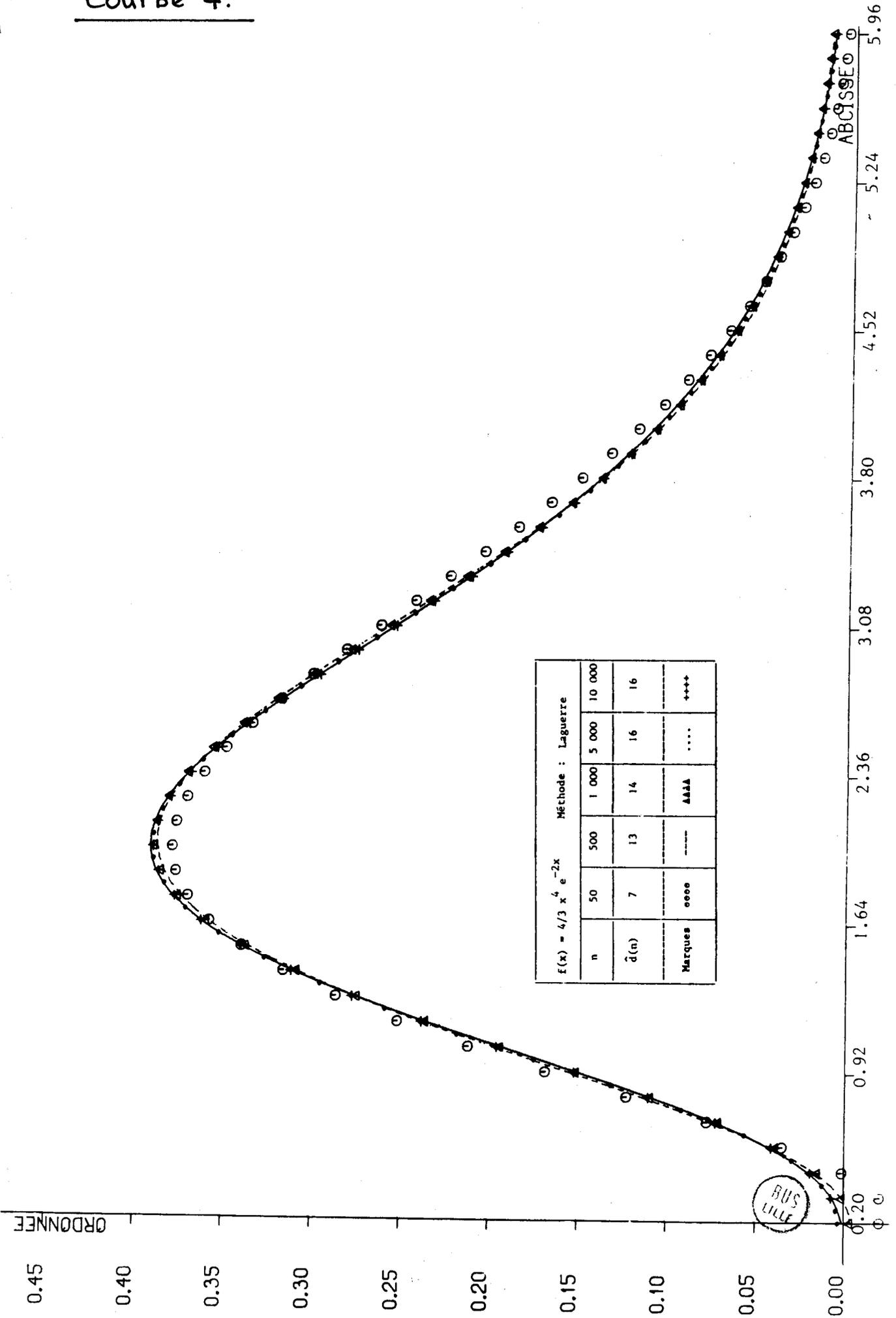
$f(x) = e^{-x/2}$		Méthode : Laguerre			
n_j	50	1 000	5 000	10 000	
$m^*(n_j)$	0	0	0	0	0
Marques	oooo	AAAA	++++	++++



Course 3.

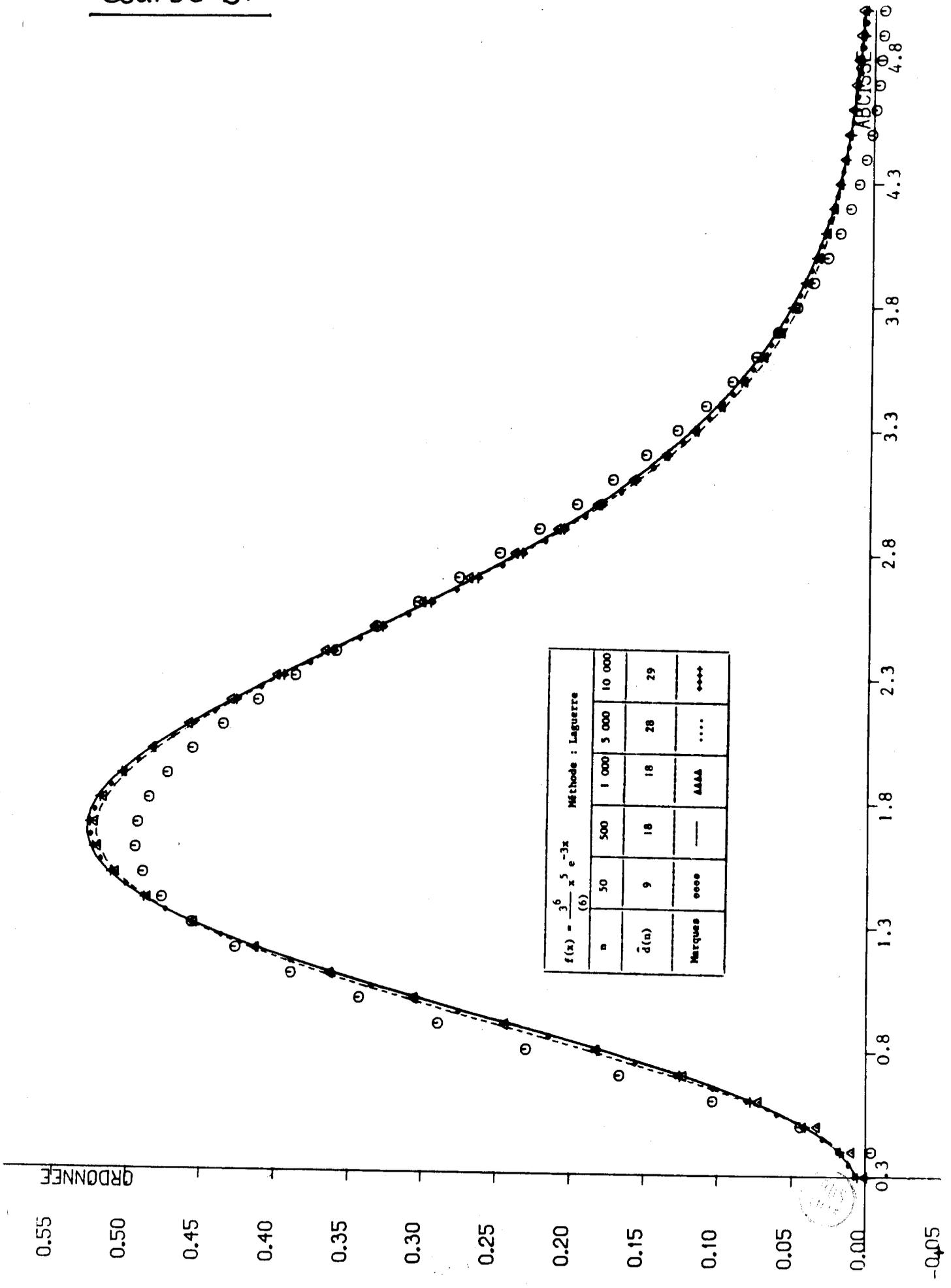


Courbe 4.



BUS LILLE

Courbe 5.

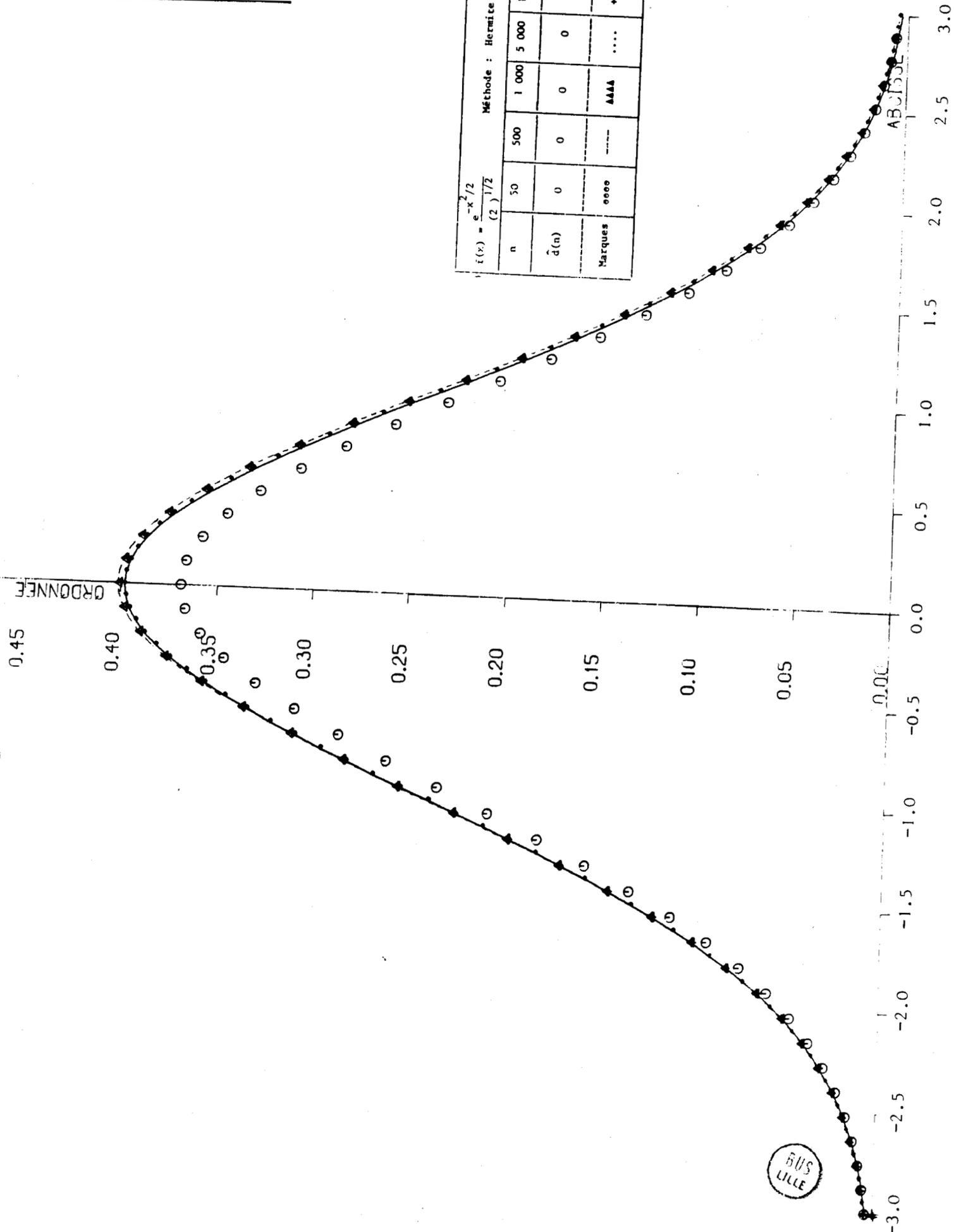


$f(x) = \frac{3^6 \cdot 5}{(6)} x^5 e^{-3x}$ Méthode : Laguerre

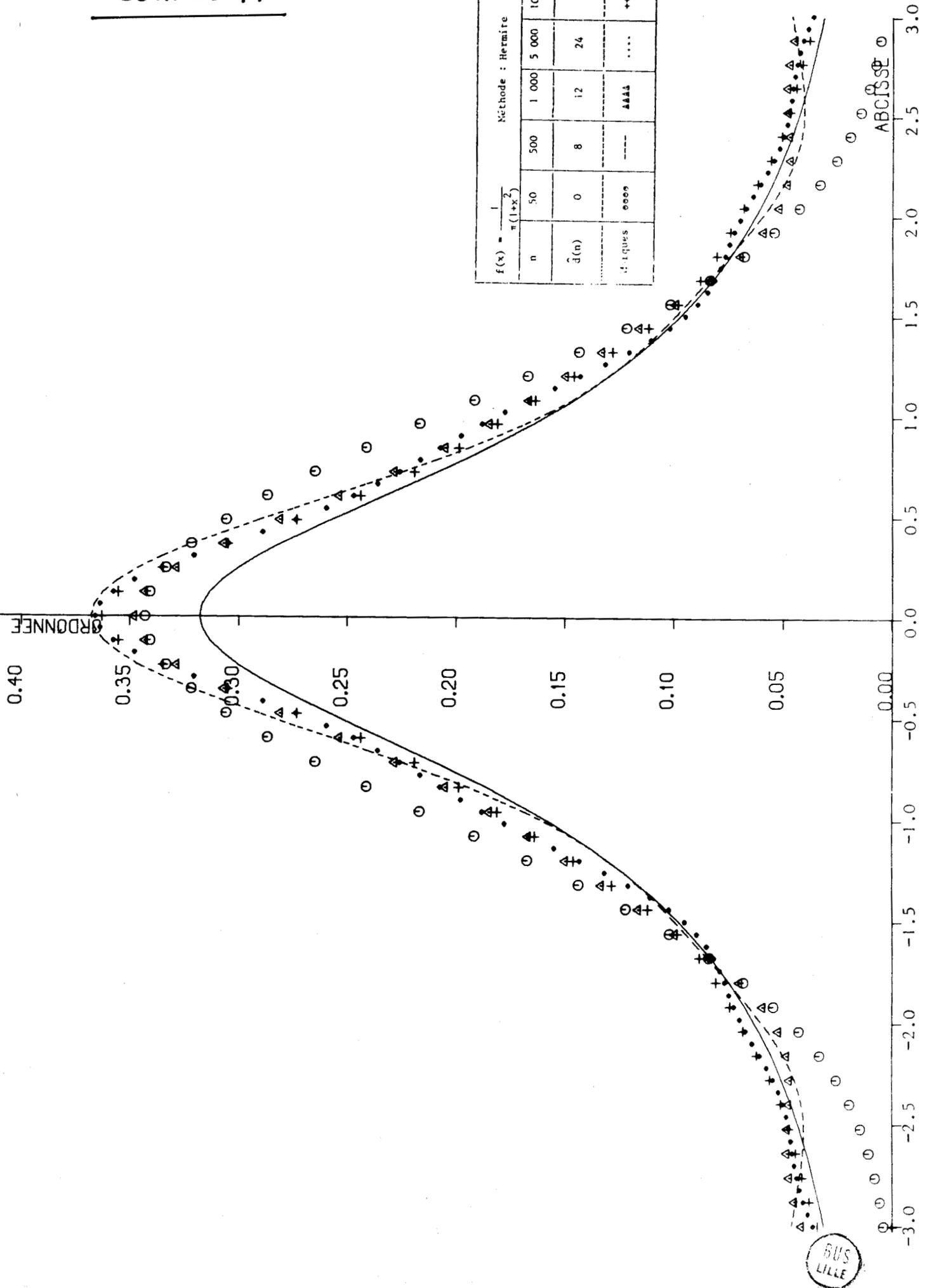
n	50	500	1 000	5 000	10 000
$\bar{d}(n)$	9	18	18	28	29
Marques	oooo	—	AAAA	++++

Courbe 6.

$f(x) = \frac{e^{-x^2/2}}{(2)^{1/2}}$		Méthode : Hermite			
n	50	500	1 000	5 000	10 000
d(n)	0	0	0	0	0
Marques	oooo	----	AAAA	++++



Courbe 7.



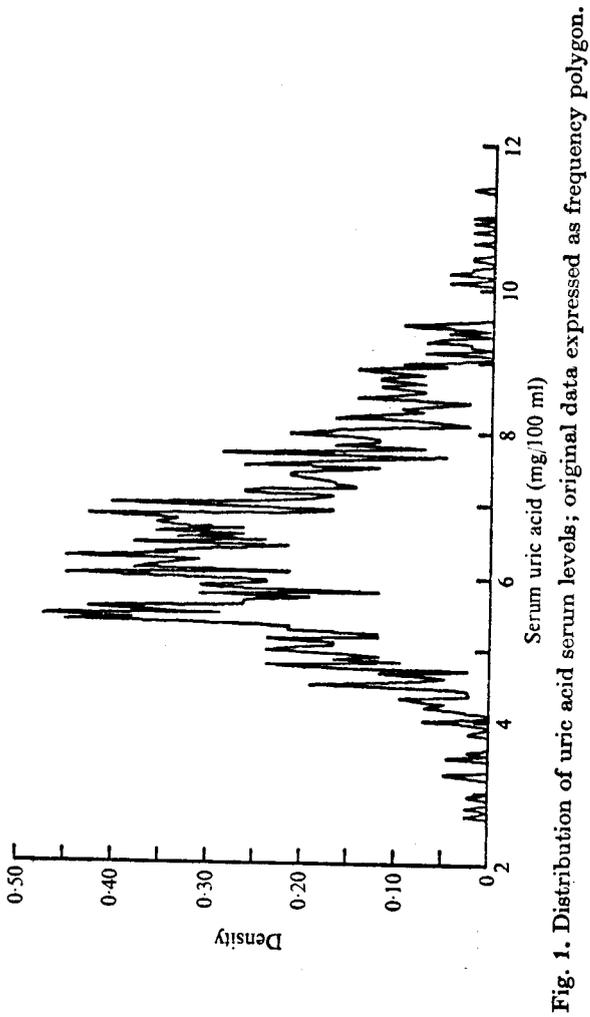


Fig. 1. Distribution of uric acid serum levels; original data expressed as frequency polygon.

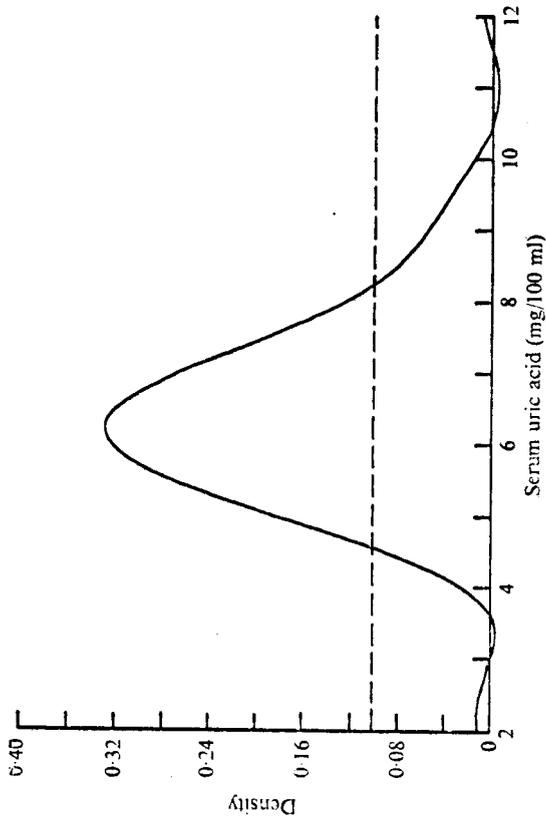


Fig. 3. Distribution of uric acid serum levels; $\lambda = 3.2$.

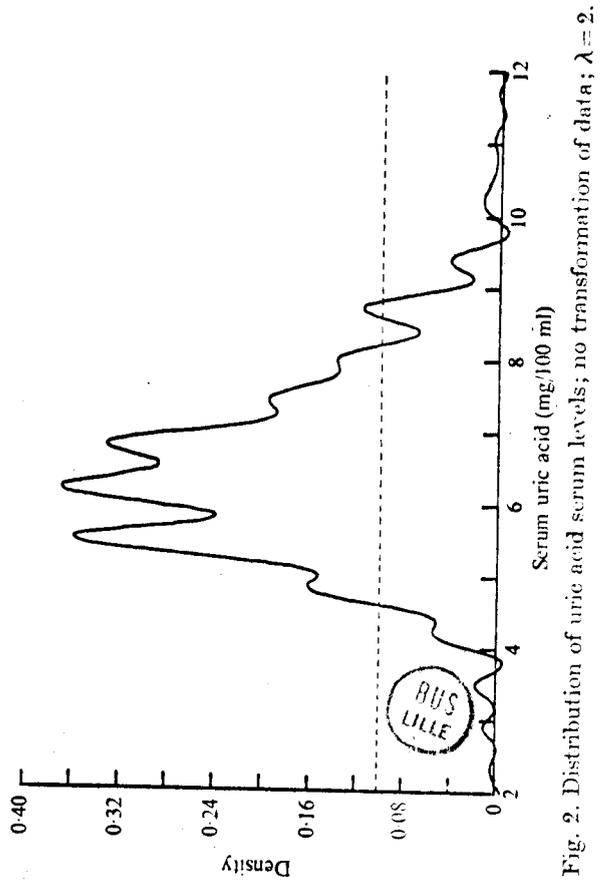


Fig. 2. Distribution of uric acid serum levels; no transformation of data; $\lambda = 2$.

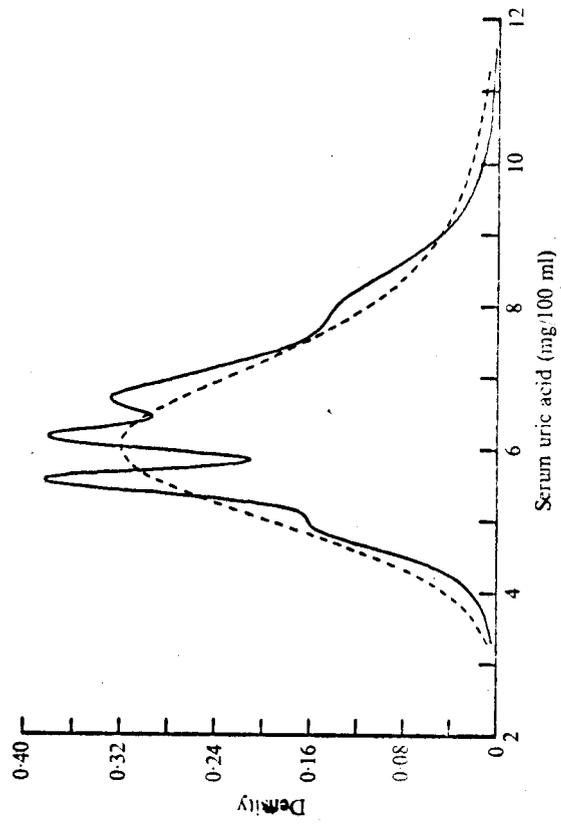
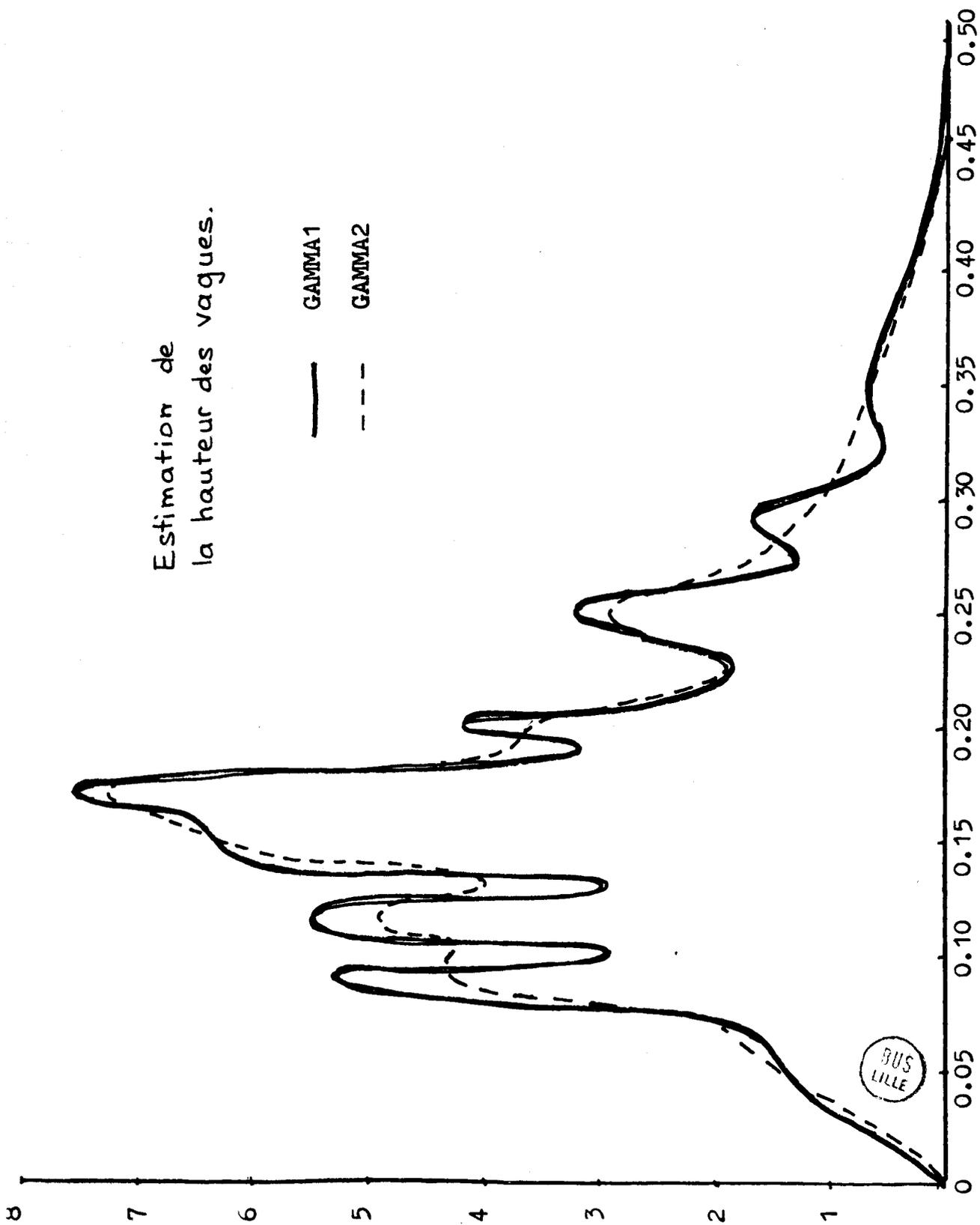


Fig. 4. Distribution of uric acid serum levels; with appropriate transformation; $\lambda = 3.2$.

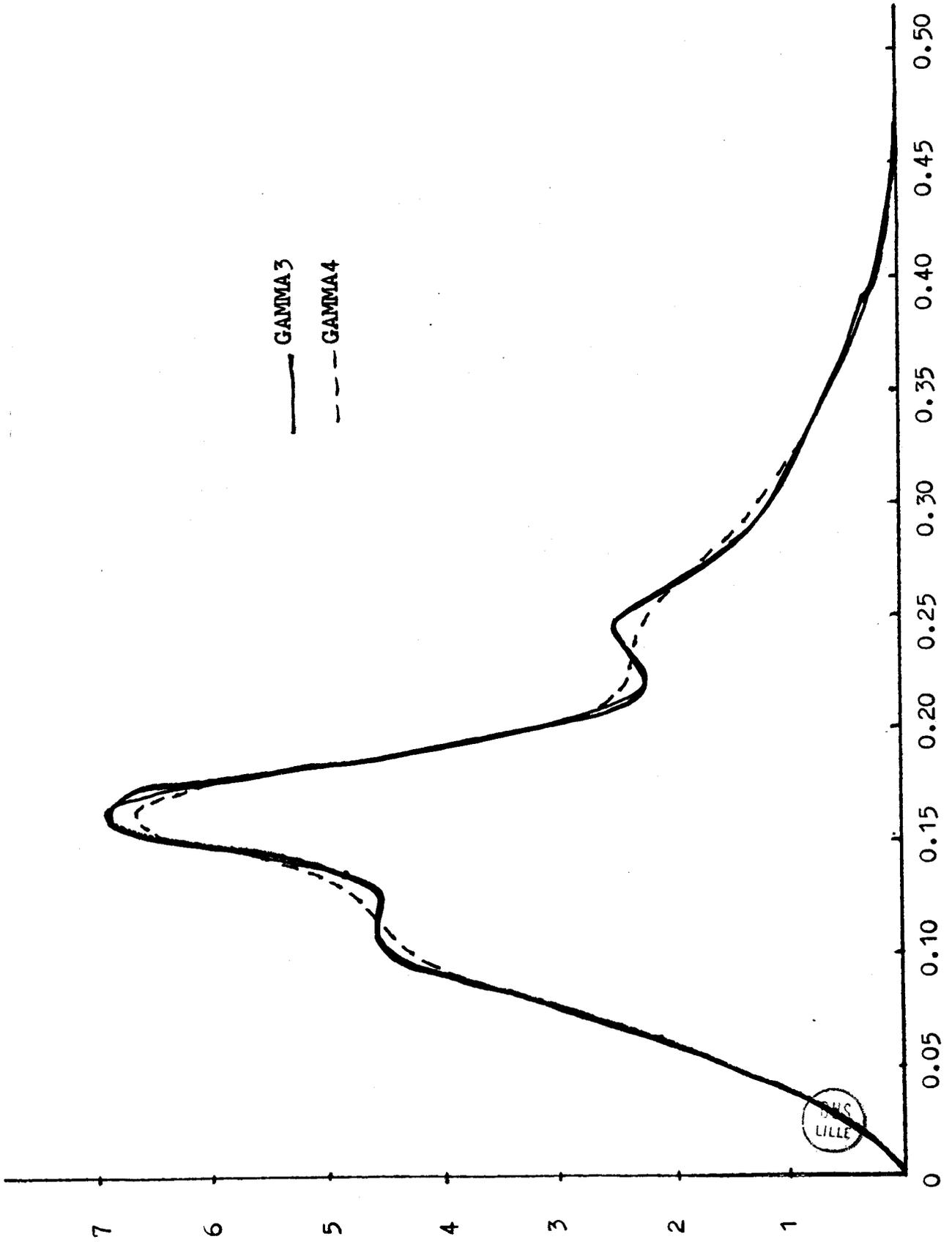
Courbe 9.

Estimation de
la hauteur des vagues.

— GAMMA1
- - - GAMMA2



Courbe 10.



B I B L I O G R A P H I E

=====

- [1] ANDERSON-FIGUEIREDO AS 80 - An adaptive orthogonal series
estimator for probability density.
- [2] AGGARWAL AMS 67 - Some minimax invariant procedures for estimating
a cumulative distribution.
- [3] BLUM ET ROSENBLATT AMS 67 - On partial a priori information in
statistical inference.
- [4] BERLINET A. 80 - Espaces autoreproduisants et mesure empirique
Thèse de 3ème cycle, Lille I.
- [5] BLEUEZ J. 76 - Conditions nécessaires et suffisantes de conver-
gence pour une classe d'estimateurs de la densité,
Thèse de 3ème cycle, Lille I.
- [6] BOSQ D. 70 - Contribution à la théorie de l'estimation
fonctionnelle, Thèse d'Etat, Publications de l'ISUP
70 .
- [7] BOSQ D. 74 - Cours de DEA rédigé par L. Nobécourt et Wibaux.
- [8] BOSQ D. 76 - Sur l'estimation d'un paramètre à valeurs
dans un espace de Hilbert.
Publication interne Lille n° 70.
- [9] BOSQ D. 77 - Sur l'estimation sans biais d'un paramètre
fonctionnel. Publication interne Lille n° 104.
- [10] BOSQ D. et BLEUEZ J. - Conditions nécessaires et suffisantes de
76 convergence pour une classe d'estimateurs de la
densité. Comptes Rendus 282 - Série A.
- [11] BOSQ D. et BLEUEZ J. - Etude d'une classe d'estimateurs non paramé-
78 triques de la densité.
Annales Institut H. Poincaré XIV, n° 4.
- [12] BOYD et STELE AS 78 - Lower bounds for non parametric density
estimation rates.
- [13] BRETAGNOLLE-HUBERT 78 - Estimation de la densité
E.R.A.
532 Risque minimax.
CNRS
- [14] BRUNK 76 - Univariate density estimation by orthogonal
series.
Technical report 51, Oregon State Univers.

- [15] BRUNK Biom 78 - Même titre.
- [16] BRUNK-PIERCE Biom 74 - Estimation of discrete multivariate densities.
- [17] CENCOV 62 - Evaluation of an unknown density observations.
- [18] DAVIS AS 75 - Mean square error properties of density estimates.
- [19] DEHEUVELS 77 - Estimation non paramétrique de la densité, Revue de Stat. Appliquées, Vol. XXV, n° 3.
- [20] EDWARDS T(1) et (2) - Fourier Series.
- [21] FARRELL AMS 67 - On the Lack of an uniformly consistent sequence of estimators of a density function in certain cases.
- [22] FARREL AMS 72 - On the best obtainables asympt rates of convergence in estimation of a density function at a point.
- [23] FELLNER Biom 74 - Heuristic estimation of probability densities.
- [24] FERGUSSON AS 73 - A bayesian analysis of some non parametric problems.
- [25] FERGUSSON AS 74 - Prior distr on spaces of probab. measures.
- [26] GHOSH AMS 64 - Uniform approximation of minimax point estimates.
- [27] GIRSHICK et SAVAGE - Bayes and minimax estimates for quadratic
AMS 51 loss functions.
- [28] GOOD-GASKINS Biom 71 - Non parametric roughness penalties for probability densities.
- [29] GUILBART C. 78 - Etude des produits scalaires sur l'espace des mesures. Thèse d'état, Lille I.
- [30] HODGES-LEHMANN AMS 50 - Some problems in minimax point estimation.
- [31] HODGES-LEHMAN AMS 52 - The use of previsions experience in reaching statistical decisions.
- [32] HOMINAL 79 - Estimation non paramétrique de la densité compte tenu d'informations sur le support. Thèse de 3ème cycle. Paris.
- [33] JACKSON & others Biom 70 - G_{α} Minimax estimators in the exponential family.
- [34] KATZ 60 - Admissible and minimax procedures estim. of parameter in truncated spaces.
- [35] KOREVAAR T(1) - Mathematical Methods.
- [36] KRONMAL-TARTER 68 - The estimation of probability densities and cumulatives by Fourier Series methods.

- [37] LEONARD JRSSB 78 - Density estimation. Stochastic processes and prior information.
- [38] LEBEDEV - Specials functions and their applications.
- [39] LORENTZ - Approximation of functions.
- [40] MIKUSINSKI-SIKORSKI - Théorie élémentaire des distributions.
- [41] MONTRICHER-TAPIA - Non parametric maximum like lihood estimation
THOMPSON A.S.75 estimation of probability densities by
penalty funct. methods.
- [42] NADARAYA, Theory of Proba. - On the Mise of some non parametric
estimates for the density function.
- [43] NELSON A.M.S.66 - Minimax solution of statistical decision
problems by iteration.
- [44] NOBECOURT L. - Séminaire du 7 mars 1979.
Sur l'estimation bayésienne.
- [45] OULEDCHEIKH M. 79 - Construction d'estimateurs de la densité par
la méthode des fonctions orthogonales.
Thèse 3ème cycle, Lille I.
- [46] PARZEN A.M.S.61 - On the estimation of a probability dens.
function.
- [47] PHADIA A.S. 73 - Minimax estimation of a cumulative distribu-
tion function.
- [48] PITCHER P.J. Math. - A more general property than domination for
65 sets of probability measures.
- [49] READ A.M.S.72 - The asympt. inadmissibility of the sample
distribution function.
- [50] ROBBINS A.M.S.63 - The empirical bayes approuch to statistical
decision problems.
- [51] ROSENBLATT 55 - Remarks on some non parametric estimates of
a density function.
- [52] SAMANIEGO - The American - On T Minimax estimation.
Stat. 75
- [53] SCHWARTZ A.M.S.67 - Estimation of probability density by orthogonals
series.
- [54] SINGH A.S. 76 - Empirical bayes estimation with convergences
rates in non continuous Lebesgue exponential
families.
- [55] SPECHT Techn. 71 - Series estimation of a probab. density function.
- [56] SZÉGO - Orthogonal polynomiales.

- [57] TARTER-RAMAN - A systematic approach to graphical methods in
Sixth Berkeley Symposium biometry.
- [58] WAHBA A.M.S. 71 - A polynomial algorithm for density estimation.
- [59] WAHBA A.M.S. 75 - Optimal convergence properties of variable
Knot, Kernel and orthogonal series methods
for density estimation.
- [60] WAHBA 78 - Data bases optimal smoothing of orthogonal
series density estimates.
Technical report 509 Wisconsin.
- [61] WAHBA 77 - Optimal smoothing of density estimates.
Ac. Press
- [62] WALD 45 - Statistical decision functions which minimize
A. of Math. the maximum risk.
- [63] WALTER A.S. 77 - Properties of hermite series estimation of
probability density.
- [64] WALTER A.S. 79 - Probability density estimation using delta
BLUM sequence.
- [65] WATSON A.M.S. 69 - Density estimation by orthogonal series.
- [66] WATSON A.M.S. 73 - On bayesian inference which incompletely
specified prior distribution.
- [67] WATSON A.M.S. 62 - On the estimation of the probability density.
LEADBETTER
- [68] WERTZ Studia - On the existence of density estimators.
Scientiarum Math.
Hungarica 74
- [69] WHITTLE R.J.S. 58 - On the smoothing of probability density functions.
- [70] WOLFOWITZ - Minimax estimates of the mean of a normal
A.M.S. 50 distribution with known variance.
- [71] ZACHS - The theory of statistical inference.

