

N° d'ordre : 1186

50376  
1984  
95

50376  
1984  
95

THÈSE

PRÉSENTÉE A

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE I

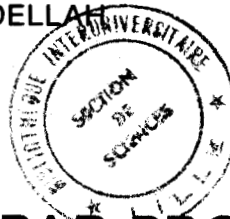
POUR OBTENIR

LE GRADE DE DOCTEUR DE 3ÈME CYCLE

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

PAR

SBAI ABDELLAH



**SUR L'ESTIMATION PAR PROJECTION  
DANS LES ENSEMBLES DE MESURES**

MEMBRES DU JURY : P. JACOB, Président

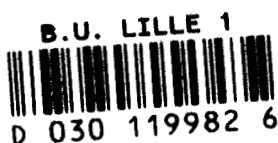
D. BOSQ, Rapporteur et Directeur de Recherche

C. LANGRAND

M. DELECROIX

} Examineurs

SOUTENUE LE 27 JUIN 1984



INTRODUCTION.-

<u>CHAPITRE</u>	I	<u>GENERALITES SUR LES ESPACES AUTOREPRODUISANTS ET LES ESPACES DE MESURES BORNEES A SIGNE.-</u>	1
	I	Espaces de Hilbert à noyau reproduisant.	1
	II	Etude des produits scalaires sur l'espace des mesures bornées à signe : préliminaires et notations.	2
	III	Relation entre produit scalaire et topologie faible.	5
	IV	Mesures aléatoires à valeurs dans $H_K$ .	7
	V	Théorème de "Glivenko-Cantelli" et estimation.	23
<u>CHAPITRE</u>	II	<u>ESTIMATION D'UN PARAMETRE FONCTIONNEL PAR PROJECTION SUR LES ENSEMBLES DE MESURES.-</u>	26
	I	Introduction.	26
	II	Estimation sans biais.	28
	III	Conditions suffisantes de convergence.	29
	IV	Loi limite de l'estimateur.	36
	V	Comparaison avec la méthode des fonctions orthogonales.	39
<u>CHAPITRE</u>	III	<u>ESTIMATION DU DEGRE DE CONTAMINATION.-</u>	44
	I	Généralités.	44
	II	Estimation du degré de contamination.	49
		1) Introduction.	49
		2) Méthode des fonctions orthogonales.	50
		3) Méthode par projection sur les ensembles de mesures.	64
	III	Autres estimateurs de $\epsilon$ .	74
<u>CHAPITRE</u>	IV	<u>SIMULATION.-</u>	80
	I	Introduction.	80
	II	Méthode par projection sur les ensembles de mesures : le calcul commun à toutes les lois qu'on va traiter par la suite.	82
	III	Loi polynômiale : $x + \frac{3}{2} x^2$ sur $[0,1]$ .	87
	IV	Loi exponentielle négative.	94
	V	Loi normale $N(0,1)$ .	98
<u>BIBLIOGRAPHIE.</u>			107

*Nous remercions très vivement Monsieur le Professeur Pierre JACOB pour l'honneur qu'il nous fait de présider le jury de cette thèse.*

*Nous remercions très vivement Monsieur le Professeur Denis BOSQ qui nous a proposé le sujet de cette thèse, qui nous a fait bénéficier de ses conseils précieux et qui nous a constamment aidé durant l'élaboration de ce travail. Notre reconnaissance envers lui est profonde.*

*Messieurs les Professeurs Claude LANGRAND et Michel DELECROIX nous ont fait l'honneur de juger ce travail ; nous les remercions très vivement.*

*Nous remercions très vivement Monsieur VAN INGELANDT pour l'aide remarquable et efficace apportée à la réalisation du travail informatique.*

*Que soient également remerciés tous les membres du Laboratoire de Statistique et Probabilités de l'U.E.R. de Mathématiques Pures et Appliquées de Lille I et en particulier Monsieur le Professeur Jean DELPORTE, qui nous ont toujours accordé aide et sympathie.*

*Nous tenons à rendre un hommage ému à Christian GUILBART qui nous a, hélas, quittés prématurément.*

*Nous remercions très vivement Madame Arlette Lengaigne pour la rapidité et le soin qu'elle a apportés à la présentation du texte, ainsi que les personnes qui ont contribué à la réalisation matérielle de cette thèse.*

## INTRODUCTION.

En statistique fonctionnelle l'estimation par projection dans des sous-espaces de dimension finie de l'espace des paramètres est utilisée fréquemment.

Dans [8] C. Guilbart a effectué d'abord la projection dans l'espace des mesures (muni d'un produit scalaire) et lui a associé ensuite son image dans l'espace des paramètres.

C'est cette nouvelle méthode qui va être l'objet de notre travail. Le chapitre I est essentiellement consacré à un rappel de certains résultats généraux sur les espaces à noyau reproduisant et les espaces de mesures. La caractérisation d'une mesure aléatoire gaussienne considérée comme une v.a. à valeurs dans un espace de Hilbert à noyau reproduisant est donnée. La convergence stochastique des mesures aléatoires est étudiée.

Dans le chapitre II, on s'intéresse à l'estimation d'un paramètre fonctionnel par un estimateur proposé par C. Guilbart dans [8]. Plusieurs modes de convergence de cet estimateur sont étudiés. La loi limite de l'estimateur est fournie. Une comparaison avec la méthode des fonctions orthogonales est faite. Les démonstrations de ces résultats sont basées sur les résultats d'un article de Denis Bosq et Jacques Bleuez [18].

Dans le chapitre III, on considère une densité de probabilité  $f = (1-\varepsilon) g + \varepsilon h$  appartenant à un modèle contaminé, le couple  $(g, h)$  est supposé connu et  $\varepsilon$  inconnu. On construit deux estimateurs sans biais de  $\varepsilon$  :

- l'un par la méthode des fonctions orthogonales ,
- l'autre par la méthode de la projection sur les ensembles de mesures.

La borne de Cramer-Rao, la loi limite et la borne supérieure pour  $0 < \epsilon < 1$  de ces deux estimateurs sont fournies. L'efficacité de ces estimateurs est étudiée. Une comparaison avec d'autres estimateurs de  $\epsilon$  est faite. On illustre cette étude par quelques exemples.

Enfin dans chapitre IV, on effectue des simulations pour comparer graphiquement 3 méthodes de l'estimation de la densité :

- méthode des fonctions orthogonales,
- méthode du noyau,
- méthode par projection sur les ensembles de mesures.

## CHAPITRE I

### GENERALITES SUR LES ESPACES AUTOREPRODUISANTS ET LES ESPACES DE MESURES BORNEES A SIGNE.

---

#### I - ESPACES DE HILBERT A NOYAU REPRODUISANT.-

La théorie des noyaux reproduisants a été introduite par N. Aronsjajn en 1943 dans [1].

Depuis elle a été l'objet de nombreux travaux, à titre d'exemples voir [2], [3], [4], [5], [6], ...

Définition : Soit  $E$  un ensemble. Une fonction  $K : E \times E \rightarrow \mathbb{R}$  est dite noyau reproduisant si  $K(x,y) = K(y,x)$  (symétrie) et si  $\forall (a_i, x_i) \in \mathbb{R} \times E \quad 1 \leq i \leq j \quad \sum a_i a_j K(x_i, x_j) \geq 0$  (semi-définie positive).

Un espace  $H$  d'applications réelles sur  $E$  est dit de Hilbert à noyau reproduisant  $K$ , s'il est de Hilbert pour un produit scalaire  $\langle \cdot, \cdot \rangle$  tel que :

- $\forall x \in E, K(\cdot, x) \in H$
- $\forall f \in H, \forall x \in E \quad \langle f, K(\cdot, x) \rangle = f(x)$  (propriété d'autoreproduction).

Comme rappel citons quelques résultats généraux :

Théorème 1.- (de caractérisation) (voir [1] page 138).

Soit  $H$  un espace de Hilbert de fonctions réelles définies sur  $E$ .  
 $H$  est à noyau reproduisant si et seulement si :

$$\forall x \in E, \quad \psi_x : H \rightarrow \mathbb{R}$$
$$f \rightarrow f(x)$$

est continue.

Remarque : Le représentant dans  $H$  de  $\delta_x$  est donc  $K(.,x)$  (propriété d'autoreproduction).

Théorème 2.- (voir [1] ou [8])

. Si  $H$  est  $K$ -reproduisant alors  $H_0$ , espace engendré par les combinaisons linéaires finies de  $K(.,y)$  est dense dans  $H$ .

. A tout noyau reproduisant  $K$  on peut associer un Hilbert  $K$ -reproduisant, fermeture pour  $\langle K(.,x), K(.,y) \rangle = K(x,y)$  de  $H_0$ .

Ce théorème de construction et dont la démonstration se trouve dans [1] ou [4] montre que l'espace  $H_K$  associé à un noyau reproduisant  $K$  n'est autre que l'ensemble des applications de  $E$  dans  $\mathbb{R}$  qui sont limites pour la topologie de la convergence simple sur  $E$  de suites d'éléments de  $H_0$  qui sont de Cauchy.

Théorème 3.- (voir [8])

$H$ , Hilbert  $K$ -reproduisant borné (c'est-à-dire  $K$  est borné), est séparable si et seulement si il existe une suite  $(f_i)_{i \in \mathbb{N}}$  d'applications de  $E \rightarrow \mathbb{R}$  telle que  $K(x,y) = \sum_{i \in \mathbb{N}} f_i(x) f_i(y)$ , la limite étant uniforme en  $x$  pour tout  $y$  fixé.

## II - ETUDE DES PRODUITS SCALAIRES SUR L'ESPACE DES MESURES BORNEES A SIGNE : PRELIMINAIRES ET NOTATIONS.-

Définition :

. Une mesure  $\mu$  à signe sur un espace mesurable  $(E, \mathcal{B})$  est une application de  $\mathcal{B}$  dans  $]-\infty, +\infty]$  telle que :

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i) \quad (\sigma\text{-additivité}) \quad \text{pour toute famille}$$

dénombrable  $\{A_i, i \in I\}$  de parties de  $E$  dans  $\mathcal{B}$  deux à deux disjoints.

. Si  $\sup \{|\mu(A)|, A \in \mathcal{B}\} < +\infty$ , alors  $\mu$  est dite bornée .

Remarque :  $-\infty$  est exclu pour la valeur de  $\mu$  pour que la somme  $\sum_{i \in I} \mu(A_i)$  définie ci-dessus ne soit jamais indéterminée.

On note :

- $M(E)$  : l'espace des mesures bornées à signe ,
- $M^+(E)$  : l'espace des mesures bornées positives ,
- $M_0(E)$  : l'espace des combinaisons linéaires finies de mesures de Dirac  $\delta_x, x \in E$  ,
- $P(E)$  : le sous-espace de  $M(E)$  , formé par l'ensemble des mesures de probabilités ,
- $P_0(E) = P(E) \cap M_0(E)$  .

C. Guilbart en se servant de la théorie des noyaux reproduisants a introduit un produit scalaire sur  $M(E)$  et sous certaines conditions il a montré que  $M(E)$  peut être considéré comme un sous-espace d'un espace de Hilbert  $H$  à noyaux reproduisants : c'est le

Théorème de plongement.- (voir [8])

Soient  $(E, \mathcal{B})$  un espace mesurable,  $K$  un noyau reproduisant borné et  $\mathcal{B} \otimes \mathcal{B}$  -mesurable.

Soit  $\langle \cdot, \cdot \rangle_K$  le pseudo-produit scalaire défini sur  $M(E)$  par  $\langle \mu, \nu \rangle_K = \langle \phi(\mu), \phi(\nu) \rangle_{H_K}$  où  $\phi$  désigne l'application linéaire :

$$\phi : M(E) \longrightarrow H_K$$

$$\mu \longmapsto \int K(\cdot, y) d\mu(y) .$$



Alors si  $\langle \cdot, \cdot \rangle_K$  est non dégénéré,  $\phi$  est une isométrie de  $(M(E), \langle \cdot, \cdot \rangle_K)$  dans  $(\phi(M(E)), \langle \cdot, \cdot \rangle_{H_K})$  et  $(M(E), \langle \cdot, \cdot \rangle_K)$  peut être considéré comme un sous-espace de  $H_K$ .

Si  $\langle \cdot, \cdot \rangle_K$  est non dégénéré deux remarques découlent de ce théorème :

- $M_0(E)$  est dense dans  $(M(E), \langle \cdot, \cdot \rangle_K)$ .
- $\forall f \in H_K, \forall \mu \in M(E) : \langle f, \phi(\mu) \rangle_{H_K} = \int f d\mu$ .

L'identification des produits scalaires, sur l'espace des mesures, qui se réintègrent s'annonce comme suit :

Théorème de réintégration. - (cf. [8]).

Soient  $(E, \mathcal{B})$  un espace mesurable et  $\langle \cdot, \cdot \rangle$  un produit scalaire sur  $M(E)$ , où  $\langle \delta_x, \delta_y \rangle$  est  $\mathcal{B} \otimes \mathcal{B}$ -mesurable et borné. Pour que  $\langle \cdot, \cdot \rangle$  se réintègre (i.e.  $\forall \mu, \nu \in M(E) : \langle \mu, \nu \rangle = \int \langle \delta_x, \delta_y \rangle d\mu(x) d\nu(y)$ ) il faut et il suffit que  $T_1 : x \mapsto \delta_x$  soit un estimateur faiblement sans biais d'ordre 1 de la loi. Autrement dit  $\forall \mu \in \mathcal{P}(E), T_1$  est  $\mu$ -intégrable au sens de Pettis et  $\int T_1 d\mu = \mu$ .

Remarques :

- Le moins de l'on puisse demander à un produit scalaire sur  $M(E)$  est que la mesure empirique soit un estimateur sans biais de la loi. Les seuls produits scalaires intéressants sont donc de la forme  $\int K d\mu \otimes \nu$ .

- Un produit scalaire sur  $M(E)$  ne se réintègre pas nécessairement.

Exemple : (voir [8]).

III - RELATION ENTRE PRODUIT SCALAIRE ET TOPOLOGIE FAIBLE.-

Soient  $(E, d)$  un espace métrique et  $(C_b(E), ||\cdot||_U)$  l'espace des fonctions réelles continues bornées, muni de la norme de la convergence uniforme  $||f||_U = \sup_E |f(x)|$ .

Définition : On appelle topologie faible de  $M(E)$  (que l'on note  $F_d$ ) la topologie engendrée par la famille d'écartes  $(d_f(\mu, \nu) = \left| \int f d(\mu - \nu) \right|)_{f \in C_b(E)}$ .

C. Guilbart a donné une caractérisation des produits scalaires dont la topologie sur l'espace des mesures de probabilités n'est autre (resp. : est moins fine) que la topologie faible, c'est le :

Théorème de faible comparaison (cf. [8]).-

Soient  $(E, d)$  un espace métrique séparable et  $\langle \cdot, \cdot \rangle$  un pseudo-produit scalaire sur  $M(E)$ . Si  $\langle \delta_x, \delta_y \rangle$  est bornée. Alors :

1) pour que  $\langle \cdot, \cdot \rangle$  soit moins fin que la topologie faible sur  $M^+(E)$  il faut et il suffit que  $\langle \cdot, \cdot \rangle$  se réintègre et que  $\langle \delta_x, \delta_y \rangle$  soit  $d \otimes d$  continu.

2) si  $\langle \cdot, \cdot \rangle$  vérifie 1), il existe une suite  $(f_i)$  uniformément bornée de  $C_b(E)$ , telle que  $\langle \delta_x, \delta_y \rangle = \sum_{i=1}^{\infty} f_i(x) f_i(y)$  (la convergence est telle que, pour tout  $x$  fixé de  $E$ ,  $\sum_{i=1}^n f_i(x) f_i(y)$  converge uniformément vers  $\langle \delta_x, \delta_y \rangle$ ).

3) si de plus  $\langle \cdot, \cdot \rangle$  est non dégénéré, pour tout compact  $F$  de  $E$   $(f_i|_F)_{i \in \mathbb{N}^*}$  est totale dans  $C_b(F)$  et  $\langle \cdot, \cdot \rangle$  a pour trace sur  $M^+(F)$  la topologie faible.

Remarques :

. Si  $F_d$  est plus fine que  $\langle \dots \rangle$  sur  $M^+(E)$  et si  $\langle \delta_x, \delta_y \rangle = \sum_{i=0}^{\infty} f_i(x) f_i(y)$  converge uniformément,  $F_d$  est alors plus fine que  $\langle \dots \rangle$  sur  $M(E)$  tout entier.

. La topologie de  $\langle \dots \rangle$  est en général strictement moins fine que  $F_d$  sur  $M(E)$  tout entier.

Exemple : [8]

On prend  $E = \mathbb{R}$  et pour tout  $i \in \mathbb{N}^*$  une suite  $(f_{i,j})_{j \in \mathbb{N}}$  de fonctions réelles, continues, bornées par 1, nulles sur  $]-\infty, -(i+1)[ \cup ]i+1, +\infty[$ , telle que la suite des restrictions  $(f_{i,j}|_{[-i,i]})_{j \in \mathbb{N}}$  soit totale dans  $C_b[-i,i]$ , on pose  $K = \sum_{i,j} \frac{1}{2^{i+j}} f_{i,j} \otimes f_{i,j}$ .

Alors  $\langle \dots \rangle_K$  est non dégénéré et vérifie l'assertion (1) du théorème de faible comparaison, et pourtant  $(\delta_n)$  est une suite tendant vers zéro pour  $\langle \dots \rangle_K$  alors qu'elle ne tend pas vers zéro pour la métrique de Prokhorov.

Théorème d'existence.- [8]

Si  $(E,d)$  est un espace métrique séparable, il existe un noyau reproduisant  $K$ , non dégénéré tel que la topologie trace de  $\langle \dots \rangle_K$  sur  $M^+(E)$  soit la topologie faible.

La démonstration de ce théorème met en évidence l'existence d'un noyau  $K$  de la forme  $K(x,y) = \sum_{i \in \mathbb{N}} f_i(x) f_i(y)$  où  $(f_i)_{i \in \mathbb{N}}$  est une suite de fonctions réelles, bornées, uniformément continues par rapport à une métrique  $d'$  équivalente à  $d$  et qui est dense dans l'ensemble des fonctions réelles, bornées, uniformément continues  $U_b(E,d')$  muni de la norme de la convergence uniforme.

Remarques :

Sous les conditions du théorème d'existence on a :

- l'ensemble des mesures de Dirac est fermé dans  $(M(E), \langle \dots \rangle_K)$ ,
- $M^+(E)$  et  $P(E)$  sont fermés dans  $(M(E), \langle \dots \rangle_K)$ .

Un théorème qui permet d'exhiber des noyaux reproduisants sur  $\mathbb{R}$  dont la métrique trace sur  $M^+(\mathbb{R})$  est équivalente à la métrique de Prokhorov et qui est dû à C. Guilbart s'énonce comme suit :

Théorème.- [8]

Si  $\psi$  est un homéomorphisme de  $\mathbb{R}$  sur  $\psi(\mathbb{R}) \subset ]-a, +a[$  (avec  $a < \infty$ ) si  $(\alpha_n)$  est une suite strictement positive et si  $K(x,y) = \sum \alpha_n (\psi(x))^n (\psi(y))^n$  (resp :  $K(x,y) = \sum \alpha_n e^{-n\psi(x)} e^{-n\psi(y)}$  est borné), alors le produit scalaire  $\langle \dots \rangle_K$  a pour topologie trace sur  $M^+(\mathbb{R})$  la topologie faible.

IV - MESURES ALEATOIRES A VALEURS DANS  $H_K$ .

Dans ce paragraphe on étudie les convergences stochastiques des mesures aléatoires considérées comme des variables aléatoires à valeurs dans un espace de Hilbert à noyau reproduisant. Soient  $(E,d)$  un espace métrique séparable et  $K$  un noyau reproduisant borné non dégénéré tel que  $H_K$  soit séparable. Notons  $d_K$  la métrique associée à  $\langle \dots \rangle_K$  et  $(M(E), \mathcal{B}_{d_K})$  l'espace des mesures bornées à signe muni de la tribu borélienne associée à  $d_K$ .

1 - Mesurabilité dans  $M(E)$ .

Lemme IV.1.1. (voir [9])

Il existe un sous-ensemble dénombrable  $D_0$  de  $H_0$  qui est dense dans  $H_K$ .

Corollaire IV.1.1.-

$(M(E), d_K)$  est séparable et il existe un sous-ensemble dénombrable  $N_0(E)$  de  $M_0(E)$  qui est dense dans  $M(E)$ .

Démonstration : Soit l'application

$$\phi : (M(E), \langle \cdot, \cdot \rangle_K) \longrightarrow (H_K, \langle \cdot, \cdot \rangle_{H_K}) \quad \text{d'après le théorème de plongement } M(E)$$

$$\mu \longmapsto \int K(\cdot, y) d(y)$$

est isométrique à  $\phi(M(E))$ ,  $H_K$  est séparable donc  $(M(E), d_K)$  est séparable.

$\phi(M_0(E)) = H_0$  donc d'après le lemme précédent il existe un sous-ensemble  $N_0(E)$  de  $M_0(E)$  et qui est dense dans  $M(E)$ .

Proposition IV.1.1.-

La tribu borélienne  $B_{d_K}$  de  $M(E)$  est engendrée par les applications  $(\psi_x)_{x \in E}$  qui sont définies par  $\psi_x(\mu) = \langle \mu, \delta_x \rangle_K$ .

Démonstration : La démonstration est analogue à celle du théorème 1 de [9]. Soit  $T$  la tribu engendrée par les  $\psi_x$ ,  $x \in E$ .

.  $\forall x \in E$   $\psi_x$  est une application continue donc  $T \subset B_{d_K}$ ,

. montrons que  $B_{d_K} \subset T$  :

$(M(E), d_K)$  étant séparable tout ouvert est réunion dénombrable de boules fermées et par conséquent il suffit de montrer que toute boule fermée  $\in T$ .

Soit  $r > 0$ ,  $\mu \in M(E)$  et  $B(\mu, r) = \{v \in M(E) : d_K(\mu, v) \leq r\}$

$$d_K(\mu, v) = \|\mu - v\|_K = \sup_{\substack{\gamma \in M(E) \\ \|\gamma\|_K \leq 1}} |\langle \mu - v, \gamma \rangle| = \sup_{\substack{\gamma \in N_0(E) \\ \|\gamma\|_K \leq 1}} |\langle \mu - v, \gamma \rangle|$$

donc  $B(\mu, r) = \bigcap_{\gamma \in N_0(E)} \{v : |\langle \mu - v, \gamma \rangle| \leq r\}$

$T$  est engendrée par les  $\psi_x$ ,  $x \in E$  donc pour tout  $x \in E$  :  $v \mapsto \langle v, \delta_x \rangle_K$  est  $T$ -mesurable et par conséquent  $\forall \gamma = \sum_{i=1}^m \alpha_i \delta_{x_i} \in N_0(E)$  l'application  $v \mapsto \langle v, \gamma \rangle_K$  est  $T$ -mesurable et pour  $\mu$  fixé :  $v \mapsto |\sum \alpha_i \langle \mu - v, \delta_{x_i} \rangle|$  est  $T$ -mesurable donc  $\{v : |\langle \mu - v, \gamma \rangle_K| \leq r, \gamma \in N_0(E)\} \in T$ .

La condition de mesurabilité dans  $(M(E), B_{d_K})$  s'énonce comme suit :

Proposition IV.1.2.-

Soit  $(\Omega, A)$  un espace mesurable et  $\mu^\bullet : (\Omega, A) \rightarrow (M(E), B_{d_K})$  une application alors  $\mu^\bullet$  est une mesure aléatoire si et seulement si  $\forall x \in E$   $\langle \mu^\bullet, \delta_x \rangle_K$  est une variable aléatoire réelle.

La démonstration est une conséquence directe de la proposition IV.1.1.

Corollaire IV.1.2.-

La donnée d'une mesure aléatoire  $\mu^\bullet : (\Omega, A) \mapsto (M(E), B_{d_K})$  est équivalente à la donnée d'un processus  $(X_t)_{t \in E}$  défini sur  $(\Omega, A)$  et dont la trajectoire appartient au sous-espace  $H_K^{M(E)} = \left\{ \int K(\cdot, y) d\mu(y), \mu \in M(E) \right\}$  de  $H_K$ .

2 - Mesures aléatoires gaussiennes dans  $(M(E), B_{d_K})$ .

Dans [10] Rajput et Cambanis ont montré l'équivalence entre processus gaussien et mesure gaussienne dans différents espaces fonctionnels (espace des fonctions, continues, absolument continues et espace  $L_2$ ) dans [9] Berlinet a étendu leurs résultats aux espaces de Hilbert à noyau reproduisant.

Nous allons étudier l'équivalence entre mesure aléatoire gaussienne et processus gaussien.

Définition IV.2.1.-

. Soient  $(F, || ||)$  un espace normé séparable,  $F^*$  son dual topologique et  $\mathcal{B}_c$  la tribu engendrée par les éléments de  $F^*$ .

Une mesure de probabilité  $\mu$  sur  $(F, \mathcal{B}_c)$  est dite gaussienne si et seulement si  $\forall f \in F^*$   $f$  est une variable aléatoire réelle gaussienne de  $(E, \mathcal{B}_c, \mu)$  dans  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ .

. Un processus  $(X_t)_{t \in T}$  est dit gaussien si et seulement si toute combinaison linéaire finie des  $(X_t)_{t \in T}$  est une variable aléatoire réelle gaussienne.

Remarque : Si  $(F, T)$  est un espace de Fréchet séparable et si on note par  $\mathcal{B}(T)$  la tribu borélienne de  $F$  alors  $\mathcal{B}_c = \mathcal{B}(T)$  (voir [11] page 100).

Définition IV.2.2.-

Une mesure aléatoire  $\mu^\bullet : (\Omega, \mathcal{A}, P) \rightarrow (M(E), \mathcal{B}_c)$  est gaussienne si et seulement si  $\forall f \in M^*(E)$   $f \circ \mu^\bullet$  est une v.a.r. gaussienne.

Avant de caractériser les mesures aléatoires gaussiennes énonçons un lemme caractérisant les formes linéaires continues sur  $(M(E), || ||_K)$ .

Lemme IV.2.1.-

. Les formes linéaires continues sur  $(M(E), || ||_K)$  sont de la forme  $\psi(\mu) = \int g d\mu$ ,  $g \in H_K$ .

.  $\mathcal{B}_{d_K} = \mathcal{B}_c$ .

Démonstration :

.)  $\forall g \in H_K$  l'application linéaire  $\psi : (M(E), || ||_K) \rightarrow \mathbb{R}$

$$\mu \longmapsto \int g d\mu$$

est continue, en effet :

$$\mu_n \xrightarrow{d_K} \mu \iff \int K(\cdot, y) d\mu_n(y) \xrightarrow{H_K} \int K(\cdot, y) d\mu(y) \implies$$

$\forall g \in H_K \quad \left\langle \int K(\cdot, y) d\mu_n(y), g \right\rangle \longrightarrow \left\langle \int K(\cdot, y) d\mu(y), g \right\rangle$  et comme pour tout  $\nu \in M(E)$  et pour tout  $f \in H_K$  on a  $\left\langle \int K(\cdot, y) d\nu(y), g \right\rangle = \int g d\nu$  (voir [8]) alors  $\int g d\mu_n \longrightarrow \int g d\mu$ .

.) Soit  $\psi$  une application linéaire continue de  $(M(E), d_K) \rightarrow \mathbb{R}$  et  $\phi$  l'isométrie de  $M(E)$  dans  $\phi(M(E))$  définie par  $\phi(\mu) = \int K(\cdot, y) d\mu(y)$ . Alors  $\Psi = \psi \circ \phi^{-1} : \phi(M(E)) \rightarrow \mathbb{R}$  est une application linéaire continue et comme  $\phi(M(E))$  est dense dans  $H_K$   $\Psi$  se prolonge d'une façon unique en une forme linéaire  $\tilde{\Psi}$  sur  $H_K$  avec  $||\tilde{\Psi}|| = ||\Psi||$ .  $H_K$  étant un espace de Hilbert, il existe donc  $g \in H_K : \tilde{\Psi} = \langle \cdot, g \rangle_{H_K}$

$$\tilde{\Psi} /_{H_K} M(E) = \Psi : \int K(\cdot, y) d\mu(y) \longrightarrow \left\langle \int K(\cdot, y) d\mu(y), g \right\rangle_{H_K} = \int g d\mu \text{ donc}$$

$$\forall \mu \in M(E) \quad \psi(\mu) = \int g d\mu.$$

Démonstration évidente d'après ce qu'on vient de montrer et la proposition IV.1.1.

Théorème IV.2.1.-

Une mesure aléatoire  $\mu^\cdot : (\Omega, A, P) \mapsto (M(E), \mathcal{B}_{d_K})$  est gaussienne si et seulement si pour tout  $x$  de  $E, \langle \mu^\cdot, \delta_x \rangle_K$  est une variable aléatoire réelle gaussienne.

Démonstration :

. Si  $\mu^\cdot$  est gaussienne alors  $\forall g \in H_K \int g d\mu^\cdot$  est une v.a.r. gaussienne en particulier  $\forall x \in E \int K(x, y) d\mu^\cdot(y) = \langle \mu^\cdot, \delta_x \rangle_K$  est gaussienne.



. Réciproquement supposons que  $\forall x \in E : \langle \mu^\cdot, \delta_x \rangle_K$  soit une v.a.r. gaussienne.

Soit  $g \in H_K : H_0$  étant dense dans  $H_K$ , il existe  $g_n \in H_0 : \|g_n - g\|_{H_K} \leq \frac{1}{n}$ . Soit  $\omega \in \Omega$ .

$$\begin{aligned} \left| \int g_n(y) d\mu^\omega(y) - \int g(y) d\mu^\omega(y) \right| &= \left| \langle g_n - g, K(\cdot, y) \rangle d\mu^\omega(y) \right| \quad (\text{propriété d'autoreproduction}) \\ &\leq \|g_n - g\|_{H_K} \int \|K(\cdot, y)\|_{H_K} d\mu^\omega(y) \quad (\text{inégalité de Schwarz}) \end{aligned}$$

$K$  étant borné donc  $\exists M > 0$  tel que  $\|K(\cdot, y)\|_{H_K} < M$  d'où :

$$\left| \int g_n(y) d\mu^\omega(y) - \int g(y) d\mu^\omega(y) \right| \leq M \|g_n - g\| \mu^\omega(E) \xrightarrow{n \rightarrow \infty} 0$$

$g_n \in H_0$  donc  $g_n = \sum_{i=1}^s \alpha_{i_n} K(\cdot, x_{i_n})$

$$\int g_n(x) d\mu^\omega(x) = \sum_{i=1}^s \alpha_{i_n} \int K(x, x_{i_n}) d\mu^\omega(x) = \sum_{i=1}^s \alpha_{i_n} \langle \mu^\omega, \delta_{x_{i_n}} \rangle_K$$

combinaison linéaire finie de gaussiennes donc  $\int g_n(x) d\mu^\omega(x)$  est gaussienne et  $\int g(x) d\mu^\omega(x)$  est gaussienne. (Comme limite p.s. de gaussiennes).

Exemple :

$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, N = N(0,1))$  et  $\mu_0$  une mesure fixe de  $M(\mathbb{R})$ .

On suppose que le noyau reproduisant  $K$  est tel que  $\int K(x,y) d\mu_0(y) > 0, x \in \mathbb{R}$ .

Alors  $\mu^\cdot : \mathbb{R} \rightarrow M(\mathbb{R})$  est une mesure aléatoire gaussienne. En effet :

$$\omega \rightarrow \omega \mu_0$$

.  $\forall x \in \mathbb{R} \quad \omega \rightarrow (\omega \mu_0, \delta_x) = \omega \langle \mu_0, \delta_x \rangle_K$  est une variable aléatoire réelle.

. Pour tout  $\alpha \in \mathbb{R}$  et pour tout  $x \in \mathbb{R}$  on a :

$$\begin{aligned} N(\omega : \langle \mu^\omega, \delta_x \rangle_K < \alpha) &= N(\omega : \omega \int K(x,y) d\mu_0(y) < \alpha) \\ &= N(\omega : \omega < \frac{\alpha}{\int K(x,y) d\mu_0(y)}) \end{aligned}$$

$$N(\omega : \langle \mu^\omega, \delta_x \rangle_K < \alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} K(x, y) d\mu_0(y) e^{-t^2/2} dt \quad \text{donc } \mu \text{ est gaussienne.}$$

3 - Notion de cylindre dans  $M(E)$ .

Notation et définition d'un sous-ensemble de  $M(E)$ .

Pour tout  $x_1, \dots, x_p \in E$  on note par  $\Pi_{(x_1, \dots, x_p)}$  l'application de  $M(E)$  dans  $\mathbb{R}^p$  définie par  $\Pi_{(x_1, \dots, x_p)}(\mu) = (\langle \mu, \delta_{x_1} \rangle_K, \dots, \langle \mu, \delta_{x_p} \rangle_K)$ .

Définition IV.3.-

$W \subset M(E)$  est dit cylindre de dimension finie si  $W$  est de la forme  $\Pi_{(x_1, \dots, x_p)}^{-1}(B)$  avec  $B$  un élément de la tribu borélienne  $\mathcal{B}_{\mathbb{R}^p}$  de  $\mathbb{R}^p$ .

Proposition IV.3.-

La classe  $\mathcal{U}$  des cylindres est une algèbre de boole engendrant la tribu borélienne de  $M(E)$ , donc c'est une classe déterminante (au sens de Billingsley [15]).

Démonstration :

\* )  $\mathcal{U}$  est une algèbre de boole :

- Stabilité par complémentation :

Soit  $W = \Pi_{(x_1, \dots, x_p)}^{-1}(B) \in \mathcal{U}$

$$W^c = \Pi_{(x_1, \dots, x_p)}^{-1}(B^c) \quad \text{et comme } B^c \text{ est un borélien de } \mathbb{R}^p$$

on a  $W^c \in \mathcal{U}$ .

-  $M(E) \in \mathcal{U}$  : pour  $x \in E$   $M(E) = \Pi_x^{-1}(\mathbb{R})$

- si  $W = \Pi_{(x_1, \dots, x_p)}^{-1}(B)$  et  $W' = \Pi_{(y_1', \dots, y_q')}^{-1}(B')$

$W \cap W' = \Pi^{-1}_{(x_1, \dots, x_p, y_1, \dots, y_q)}(B \times B')$  et  $B \times B'$  est un borélien de  $\mathbb{R}^{p+q}$   
 donc  $W \cap W'$  est un cylindre.

\*\* )  $U$  engendre la tribu  $\mathcal{B}_{d_K}$  :

Soit  $T$  la tribu engendrée par  $U$  :

-  $\forall p \in \mathbb{N}^*$ ,  $\forall (x_1, \dots, x_p) \in E^p$   $\Pi_{(x_1, \dots, x_p)}$  est continue de  $(M(E), d_K)$  dans  $\mathbb{R}^p$  donc  $T \subset \mathcal{B}_{d_K}$ .

- Il nous reste à montrer que  $\mathcal{B}_{d_K} \subset T$ , pour cela il suffit de montrer que toute boule fermée de  $M(E)$  est dans  $T$ . La démonstration proposée est analogue à la démonstration de la proposition IV.1.1.

Soit  $B(\mu, r) = \{v : \|\mu - v\|_K \leq r\}$  une boule fermée

$B(\mu, r) = \bigcap_{\substack{\gamma \in N_0(E) \\ \|\gamma\|_K \leq 1}} \{v : |\langle \mu - v, \gamma \rangle_K| \leq r\}$  ( $N_0(E)$  étant l'ensemble dénombrable

dense dans  $M(E)$  (corollaire IV.1.1.))  $\gamma \in N_0(E)$  donc  $\gamma = \sum_{i=1}^m \alpha_i \delta_{x_i}$

Considérons  $\psi : (\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m}) \longmapsto (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  est mesurable.

$$(\beta_1, \dots, \beta_m) \longmapsto \sum_{i=1}^m \alpha_i |\beta_i - \langle \mu, \delta_{x_i} \rangle_K|$$

Posons  $B = \psi^{-1}([0, r]) \in \mathcal{B}_{\mathbb{R}^m}$  et on a :  $\{v : |\langle \mu - v, \gamma \rangle_K| \leq r\} = \Pi^{-1}_{(x_1, \dots, x_m)}(B) \in T$

donc comme intersection dénombrable d'éléments de  $T$   $B(0, r) \in T$ .

4 - Convergence stochastique dans  $M(E)$ .

i) Définition et préliminaires.

Définition IV.4.i. (Geffroy).

Soit  $(E, \mathcal{B})$  un espace mesurable,  $M$  un ensemble non vide de mesures sur  $(E, \mathcal{B})$  et  $T$  une tribu de parties de  $M$ . On dit que  $(M, T)$  est un espace

mesurable de mesures adapté à l'espace mesurable  $(E, \mathcal{B})$  si :

$$\forall B \in \mathcal{B}, \text{ l'application } \varphi_B : (M, \mathcal{T}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \text{ est mesurable.}$$

$$\mu \longmapsto \mu(B)$$

Dans la suite on suppose que  $(E, d)$  est un espace métrique séparable que l'on munit de sa tribu borélienne  $\mathcal{B}_d$ .

$B(0, r)$  désignera une boule de centre 0 et de rayon r

Soit  $A \in \mathcal{B}_d$ , on désignera par

$$\left\{ \begin{array}{l} - \text{ pour } \delta > 0 \quad A^\delta = \{x : d(x, A) < \delta\} \\ \quad \text{le dilaté de } A \text{ à } \delta \text{ près} \\ - \delta A : \text{ la frontière de } A. \end{array} \right.$$

Dans [12] J. Geffroy a introduit sur l'ensemble  $M$  des mesures positives "θ-finies" (c'est-à-dire finies sur tout borélien borné de  $\mathcal{B}_d$ ) la distance :

$$\rho(\mu, \mu') = \sum_{n=1}^{\infty} 2^{-n} \frac{\rho_n(\mu, \mu')}{1 + \rho_n(\mu, \mu')} \quad \text{avec}$$

$$\rho_n(\mu, \mu') = \sup \{ \theta_n(\mu, \mu'), \theta_n(\mu', \mu) \} \quad \text{et}$$

$$\theta_n(\mu, \mu') = \inf \{ \delta > 0 / \forall A \in \mathcal{B}_d \text{ avec } A^\delta \subset B(0, r), \mu(A) \leq \mu(A^\delta) + \delta \}$$

pour étudier la convergence stochastique des mesures aléatoires à valeurs dans  $M$ . Citons quelques uns de ces résultats qui vont nous être utiles par la suite.

Théorème IV.4.i.1. (voir [12]).

Etant donné une mesure  $\mu \in M$  et une suite généralisée  $\{\mu_\theta, \theta \in \Theta\}$ , où  $\Theta$  est un ensemble filtrant supérieurement, on a l'équivalence :

$$\lim_{\Theta} \rho(\mu, \mu_\theta) = 0 \iff \forall A \in \mathcal{B}_b \text{ tel que } \mu(\delta A) = 0 \quad \lim_{\Theta} \mu_\theta(A) = \mu(A)$$

$\mathcal{B}_b$  : désigne l'ensemble des boréliens bornés de  $E$ .

Théorème IV.4.i.2.- (voir [12]).

La fonction  $\psi_B : (M, \mathcal{B}_\rho) \longrightarrow (\mathbb{R}, \mathcal{B}_\mathbb{R})$  est mesurable  $\forall B \in \mathcal{B}_b$ .

Remarques :

. La distance  $\rho$  induit une topologie équivalente à la topologie faible sur  $P(E)$  mais non sur un ensemble de mesures quelconques.

. Si on considère  $M$  : l'ensemble des mesures positives finies sur les boréliens bornés de  $E$  alors  $M^+(E) \subset M$ .

Proposition IV.4.i.1.-

La topologie induite par  $d_K$  sur  $M^+(E)$  est plus fine que la topologie induite par  $\rho$ .

Démonstration : Soit  $\mu_n \xrightarrow{d_K} \mu \iff \mu_n \xrightarrow{\rho} \mu \iff \forall A \in \mathcal{B}_d$  tel que  $\mu(\delta A) = 0$   $\mu_n(A) \longrightarrow \mu(A)$  (voir [13] page 366 théorème 2). En particulier  $\forall A \in \mathcal{B}_d$  borélien borné tel que  $\mu(\delta A) = 0$  on a  $\mu_n(A) \longrightarrow \mu(A)$  et le théorème IV.4.i.1. implique que  $\mu_n \xrightarrow{\rho} \mu$ .

Notre but est d'étudier la convergence stochastique des mesures aléatoires à valeurs dans  $(M^+(E), d_K)$ . Mais avant d'étudier cette convergence, montrons comme dans [12] que  $(M(E), \mathcal{B}_{d_K})$  est adapté à  $(E, \mathcal{B}_d)$ . Pour cela énonçons deux lemmes :

Lemme IV.4.i.1.-

Pour tout ouvert  $\theta$  de  $E$   $\psi_\theta : (M^+(E), d_K) \longrightarrow \mathbb{R}$  est

$$\mu \longmapsto \mu(\theta)$$

semi-continue inférieurement.

Démonstration :

La démonstration de ce lemme est analogue à celle du lemme 4.2.

de ([12], page 21).

On va montrer que si  $\mu_n \xrightarrow{d_K} \mu$  alors pour  $\varepsilon > 0 \exists N : n \geq N$  on a  $\psi_\theta(\mu_n) \geq \psi_\theta(\mu) - \varepsilon$ .

(E,d) étant séparable, soit  $(x_i)_{i \in \mathbb{N}}$  une suite dense dans  $\theta$ .

$\forall i \exists r_i > 0$  tel que  $B(x_i, r_i) \subset \theta$  et  $\theta = \bigcup_{i \in \mathbb{N}} B(x_i, r_i)$ . Soit  $s_i \in ]0, r_i]$  tel que :  $\mu(\partial B(x_i, s_i)) = 0$   $s_i$  existe sinon  $B(x_i, s_i)$  ne serait pas de mesure finie. On a  $\theta = \bigcup_{i \in \mathbb{N}} B(x_i, s_i)$ .

Posons :

$$\begin{aligned} C_1 &= B(x_1, s_1) \\ C_2 &= B(x_2, s_2) - B(x_1, s_1) \\ &\vdots \\ C_i &= B(x_i, s_i) - \bigcup_{j=1}^{i-1} B(x_j, s_j) \end{aligned}$$

Les  $C_i$  sont des boréliens deux à deux disjoints et comme

$$\delta(A \cap B) \subset \delta(A) \cup (\partial B) \implies \delta C_i \subset \bigcup_{j=1}^i \delta B(x_j, s_j) \text{ et donc}$$

$$\forall i \quad \mu(\delta C_i) = 0 \quad \mu_n \xrightarrow{d_K} \mu \implies \mu_n \not\equiv \mu \implies \mu_n(C_i) \xrightarrow{i} \mu(C_i) \quad (\text{voir [13],$$

$$\text{p. 12), or } \mu(\theta) = \mu\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mu(C_i), \quad \mu_n(\theta) = \sum_{i=1}^{\infty} \mu_n(C_i).$$

$$\text{Soit } \varepsilon > 0 \quad \mu_n(C_i) \rightarrow \mu(C_i) \quad \forall i \quad \exists N(i) : n > N(i)$$

$$|\mu_n(C_i) - \mu(C_i)| < \frac{\varepsilon}{2^{i+1}} \implies \mu(C_i) \leq \mu_n(C_i) + \frac{\varepsilon}{2^{i+1}}$$

$$\mu(\varepsilon) = \sum_{i=1}^{\infty} \mu(C_i) < \infty \implies \forall \varepsilon > 0 \quad \exists N_0 : \sum_{i=N_0+1}^{\infty} \mu(C_i) < \varepsilon/2,$$

$$\text{d'où } \mu(\theta) \leq \sum_{i=1}^{N_0} \mu(C_i) + \varepsilon/2.$$

$$\text{Posons } N_1 = \sup(N(i), i = 1, \dots, N_0).$$

Alors pour  $n > N_1$  on a  $\mu(\theta) \leq \sum_{i=1}^{N_0} \mu(C_i) + \epsilon/2$

$$\mu(\theta) \leq \sum_{i=1}^{N_0} \left( \mu_n(C_i) + \frac{\epsilon}{2^{i+1}} \right) + \epsilon/2 \leq \mu_n(\theta) + \epsilon, \text{ soit pour}$$

$n > N_1$ ,  $\psi_n(\mu_n) > \psi_n(\mu) - \epsilon$  et par conséquent  $\psi_\theta$  est semi-continue inférieurement et donc mesurable.

Lemme IV.4.i.2.-

$\forall B \in \mathcal{B}_d \quad \psi_B : (M^+(E), \mathcal{B}_{d_K}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  est mesurable.

Démonstration :

Soit  $W = \{B \in \mathcal{B}_d : \psi_B \text{ est mesurable}\}$

$F = \{\theta, \theta \text{ ouvert de } E\} \subset W$ ,  $F$  est un  $\Pi$ -système et la tribu engendrée par  $F$  est  $\mathcal{B}_d$ . Si  $W$  est un  $\lambda$ -système, alors  $W$  contient le  $\lambda$ -système engendré par  $F$  donc  $W$  serait égale à  $\mathcal{B}_d$ . Pour démontrer le lemme il suffit donc de montrer que  $W$  est un  $\lambda$ -système :

-  $B_1 \in W$ ,  $B_2 \in W$ ,  $B_1 \subset B_2$  :

$$B_2 = B_1 \cup (B_2 - B_1) \implies \mu(B_2) = \mu(B_1) + \mu(B_2 - B_1) \implies \psi_{B_2 - B_1} = \psi_{B_2} - \psi_{B_1}$$

différence de deux fonctions mesurables donc mesurable.

- Si  $(B_n)_{n \in \mathbb{N}} \subset W : B_n \subset B_{n+1}$

$$\mu(\cup B_n) = \lim_{n \rightarrow \infty} \mu(B_n) = \mu(\lim B_n)$$

donc  $\psi_{(\cup B_n)}(\mu) = \lim_{n \rightarrow \infty} \psi_{B_n}(\mu)$  est mesurable comme limite de fonctions mesurables.

Proposition IV.4.i.2.-

$(M(E), \mathcal{B}_{d_K})$  est un espace mesurable de mesures bornées à signe

adapté à  $(E, \mathcal{B}_d)$ .

Démonstration :

$\forall \mu \in M(E) : \mu = \mu^+ - \mu^-$  (décomposition de Hahn - Jordan)

donc  $\forall B \in \mathcal{B}_d \quad \varphi_B(\mu) = \varphi_B(\mu^+ - \mu^-) = \mu^+(B) - \mu^-(B) = \varphi_B(\mu^+) - \varphi_B(\mu^-)$  donc  $\varphi_B$  est la différence de deux fonctions mesurables donc mesurable.

Conséquence : Si  $\mu^\bullet : (\Omega, \mathcal{A}) \mapsto (M(E), \mathcal{B}_{d_K})$  est une mesure aléatoire alors  $\forall B \in \mathcal{B}_d : \mu^\bullet(B) : (\Omega, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  est une variable aléatoire réelle.

ii) Convergence presque sûre.

Proposition IV.4.ii.1.-

Soient  $\mu_n, \mu : (\Omega, \mathcal{A}, P) \mapsto (M^+(E), \mathcal{B}_{d_K})$  des mesures aléatoires,  $n \in \mathbb{N}$ .

Considérons les conditions suivantes :

- 1)  $\mu_n^\bullet \xrightarrow{\text{P.S.}} \mu^\bullet$ .
- 2)  $\int K(\cdot, y) d\mu_n^\bullet \xrightarrow{\text{P.S.}} \int K(\cdot, y) d\mu^\bullet(y)$  dans  $H_K$ .
- 3)  $\forall B \in \mathcal{B}_d$  tel que  $\mu^\bullet(\partial B) = 0$  p.s.  $\mu_n^\bullet(B) \rightarrow \mu^\bullet(B)$  p.s.
- 4)  $\int K(\cdot, y) d\mu_n^\bullet \rightarrow \int K(\cdot, y) d\mu^\bullet$  uniformément presque sûrement.
- 5) Pour toute famille  $\Pi$  de  $H_K$  bornée au sens de la norme de  $H_K$   
 $\sup_{f \in \Pi} \left| \int f d\mu_n^\bullet - \int f d\mu^\bullet \right| \xrightarrow{\text{P.S.}} 0$ .
- 6)  $\forall g \in H_K \quad \int g d\mu_n^\bullet \rightarrow \int g d\mu^\bullet$  p.s.

Alors on peut énoncer les résultats suivants :

1)  $\iff$  2)  $\iff$  3)  $\implies$  6) , et si on suppose en plus que le noyau  $K$  est séparément continu et que les mesures aléatoires  $\mu_n^\bullet, \mu^\bullet$  sont à valeurs dans  $P(E)$  alors :

1)  $\iff$  2)  $\iff$  3)  $\iff$  4)  $\iff$  5)  $\implies$  6) .



Démonstration :

1)  $\iff$  2)

$$\mu_n^\bullet \xrightarrow{\text{p.s.}} \mu^\bullet \iff \exists \Omega_0 \in A : P(\Omega_0) = 1 \quad \text{et} \quad \forall \omega \in \Omega_0 \quad \mu_n^\omega \xrightarrow{d_K} \mu^\omega$$

$$\iff \exists \Omega_0 \in A : P(\Omega_0) = 1 \quad \text{et} \quad \forall \omega \in \Omega_0 \quad \int K(\cdot, y) d\mu_n^\omega \rightarrow \int K(\cdot, y) d\mu_n^\omega$$

dans  $H_K$

$$\iff \int K(\cdot, y) d\mu_n^\bullet(y) \rightarrow \int K(\cdot, y) d\mu^\bullet(y) \quad \text{p.s. dans } H_K$$

1)  $\iff$  3 : Evident d'après [12].

1)  $\implies$  6) :

$$\mu_n^\bullet \rightarrow \mu^\bullet \text{ p.s. } \iff \exists \Omega_0 \in A : P(\Omega_0) = 1 \quad \text{et tel que}$$

$$\forall \omega \in \Omega_0 \quad \mu_n^\omega \xrightarrow{d_K} \mu^\omega$$

$$\iff \exists \Omega_0 \in A : P(\Omega_0) = 1 \quad \text{et tel que}$$

$$\forall \omega \in \Omega_0 \quad \int K(\cdot, y) d\mu_n^\omega \rightarrow \int K(\cdot, y) d\mu^\omega \quad \text{dans } H_K .$$

$$\implies \exists \Omega_0 \in A : P(\Omega_0) = 1 \quad \text{et} \quad \forall g \in H_K \quad \langle g, \int K(\cdot, y) d\mu_n^\omega \rangle_{H_K} \rightarrow \langle g, \int K(\cdot, y) d\mu^\omega \rangle_{H_K}$$

$$\implies \exists \Omega_0 \in A : P(\Omega_0) = 0 \quad \text{et} \quad \forall g \in H_K \quad \int g d\mu_n^\omega \rightarrow \int g d\mu^\omega .$$

On suppose maintenant que  $\mu_n, \mu : (\Omega, A, P) \rightarrow (P(E), d_K)$

2)  $\Longleftarrow$  4)

$$2) \Longleftarrow \exists \Omega_0 \in A : P(\Omega_0) = 1 \text{ et } \forall \omega \in \Omega_0 \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} \rightarrow 0$$

$$\text{Or } \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right| = \left| \langle K(x, \cdot), \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \rangle_{H_K} \right| \\ \leq \left\| K(x, \cdot) \right\|_{H_K} \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} \quad (\text{inégalité de Schwarz})$$

K étant borné donc  $\exists M > 0$  :

$$\left\| K(x, \cdot) \right\|_{H_K} \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\| \leq M \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} \text{ pour } \omega \in \Omega_0 \\ \leq M \left( \int \left| K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right|^2 d(\mu_n^\omega - \mu^\omega)(x) \right)^{1/2} \\ \leq M \sqrt{2} \left( \sup_{x \in E} \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right|^2 \right)^{1/2}$$

On a donc

$$\sup_{x \in E} \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right| \leq \left\| K(x, \cdot) \right\|_{H_K} \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} \\ \leq \sqrt{2} M \left( \sup_{x \in E} \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right|^2 \right)^{1/2} .$$

3)  $\implies$  5)

$$3) \Longleftarrow \exists \Omega_0 \in A : P(\Omega_0) = 1 \text{ et pour tout } \omega \in \Omega_0 \mu_n^\omega \implies \mu^\omega \text{ soit } \Pi$$

une famille bornée de  $H_K$ . K étant séparément continue  $\implies H_K \subset C_b(E)$  (théorème

II.5.1. page II.10 de [4]) et  $\Pi$  est équicontinue (voir [2] page 345) et par

conséquent pour tout  $\omega \in \Omega_0$   $\limsup_{f \in \Pi} \left| \int f d\mu_n^\omega - \int f d\mu^\omega \right| = 0$  (voir [16] théorème 6.8 page 51).

5)  $\implies$  4)

Soit  $\Pi = \{K(\cdot, x), x \in E\}$  c'est une famille bornée dans  $H_K$  donc

équicontinue donc  $\limsup_{x \in E} \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right| = 0$  p.s.

iii) Convergence en probabilité dans  $M^+(E)$ .

On suppose qu'il existe une constante  $C > 0$  telle que

$\forall n \in \mathbb{N}$ ,  $\mu_n^\cdot, \mu^\cdot : (\Omega, \mathcal{A}, P) \rightarrow (M^+(E), \mathcal{B}_{d_K})$  soient uniformément bornées par  $C$  presque sûrement

Proposition IV.4.iii.1.-

Les conditions suivantes sont équivalentes.

- 1)  $\mu_n^\cdot \rightarrow \mu^\cdot$  en probabilité,
- 2) pour toute famille  $\Pi$  de  $H_K$  bornée dans  $H_K$

$$\sup_{f \in \Pi} \left| \int f d\mu_n^\cdot - \int f d\mu^\cdot \right| \xrightarrow{P} 0 .$$

Démonstration :

Supposons que  $\mu_n^\cdot \rightarrow \mu^\cdot$  en probabilité. Soit  $\Pi$  une famille bornée de  $H_K$ . Pour  $\omega \in \Omega$  on a :

$$\begin{aligned} \sup_{g \in \Pi} \left| \int g(y) d(\mu_n^\omega - \mu^\omega)(y) \right| &= \sup_{g \in \Pi} \left| \langle g, \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \rangle_{H_K} \right| \\ &\leq \sup_{g \in \Pi} \|g\|_{H_K} \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} \end{aligned}$$

$$\Rightarrow \left\{ \omega : \sup_{g \in \Pi} \left| \int g(y) d(\mu_n^\omega - \mu^\omega)(y) \right| > \varepsilon \right\} \subset \left\{ \omega : \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} > \frac{\varepsilon}{\sup_{g \in \Pi} \|g\|_{H_K}} \right\}$$

$$\text{donc } P(\omega : \sup_{g \in \Pi} \left| \int g d(\mu_n^\omega - \mu^\omega) \right| > \varepsilon) \leq P(\omega : \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} > \frac{\varepsilon}{\sup_{g \in \Pi} \|g\|_{H_K}})$$

$$\mu_n^\cdot \xrightarrow{P} \mu^\cdot \iff \forall \varepsilon > 0 \quad P(\omega : \|\mu_n^\omega - \mu^\omega\|_K > \varepsilon) = P(\omega : \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} > \varepsilon)$$

converge vers 0 .

$$\Rightarrow P(\omega : \sup_{g \in \Pi} \left| \int g d(\mu_n^\omega - \mu^\omega) \right| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0 .$$

2)  $\implies$  1) :

$$\begin{aligned} \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K}^2 &= \int \left( \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right) d(\mu_n^\omega - \mu^\omega)(x) \\ &\leq 2C \sup_{x \in E} \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right| \end{aligned}$$

donc

$$\{\omega : \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} > \varepsilon\} \subset \{\omega : \sup_{x \in E} \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right| > \frac{\varepsilon}{2C}\}$$

$$\text{donc si } P(\omega : \sup_{x \in E} \left| \int K(x, y) d(\mu_n^\omega - \mu^\omega)(y) \right| > \frac{\varepsilon}{2C}) \xrightarrow[n \rightarrow \infty]{} 0$$

$$\implies P(\omega : \left\| \int K(\cdot, y) d(\mu_n^\omega - \mu^\omega)(y) \right\|_{H_K} > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0$$

Pour conclure il suffit de remarquer que  $\Pi = \{K(\cdot, y), y \in E\}$  est une famille de  $H_K$  bornée.

V - THEOREME DE "GLIVENKO-CANTELLI" ET ESTIMATION.-

En théorie d'estimation au lieu de considérer la projection sur les espaces de paramètres, C. Guilbart a considéré la projection sur les espaces de mesures auxquelles il a associé ensuite son image dans l'espace des paramètres. Comme base, il a énoncé un "théorème de Glivenko-Cantelli" et plus précisément.

Théorème ([8] page 26).-

Si  $K = \sum \alpha_i f_i \otimes f_i$  est normal (c'est-à-dire les  $f_i$  sont des v.a.r. bornées par 1 et  $(\alpha_i)$  une série convergente de termes strictement positifs). Alors :

$$a) \sup_{n \geq N_1} \mu^{\otimes \infty} \left( \bigcup_{n \geq N_1} \{\omega \in \Omega^\infty : \left\| \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} - \mu \right\|_K \geq a\} \right) \leq \frac{16 n_0(a) \sum_{\ell=1}^{\infty} \alpha_\ell}{a^2 \cdot N_1}$$

$$b) \sup_{\mu} \mu^{\otimes \infty} (\omega \in \Omega^{\infty} : \left\| \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} - \mu \right\|_K \geq a) \leq 2 n_0(a) \exp\left(-\frac{n a^2}{4 \sum_{\ell=1}^{\infty} \alpha_{\ell}}\right)$$

avec  $a > 0$  et  $n_0(a)$  est le plus petit entier tel que :

$$\left\{ \begin{array}{l} \cdot \sum_{i=n_0(a)+1}^{\infty} \alpha_i < \frac{a^2}{2} \quad \text{si les } f_i \text{ sont positives} \\ \cdot \sum_{i=n_0(a)+1}^{\infty} \alpha_i < \frac{a^2}{8} \quad \text{sinon .} \end{array} \right.$$

La démonstration de ce théorème repose sur certaines inégalités de Hoeffding (voir [17]).

Soient  $\mathcal{P}_1$  un sous-espace de  $P(E)$  et  $(\mu_n)$  une suite dense dans  $(\mathcal{P}_1, d_K)$ .  $\langle \mu_1, \dots, \mu_n \rangle$  et  $e(\mu_1, \dots, \mu_n)$  désignent respectivement l'enveloppe convexe et vectorielle de  $(\mu_1, \dots, \mu_n)$ .

Proposition ([8] page 53).-

a)  $\text{prj}_{\langle \mu_1, \dots, \mu_n \rangle}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right)$  est un estimateur convergeant presque sûrement vers la loi sur  $\mathcal{P}_1$ .

b)  $\text{prj}_{e(\mu_1, \dots, \mu_n)}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right)$  est un estimateur asymptotiquement sans biais (c'est-à-dire :  $\forall \mu \in \mathcal{P}_1, \int_{\Omega^n} \text{prj}_{e(\mu_1, \dots, \mu_n)}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right) d\mu^{\otimes n}(\omega_1, \dots, \omega_n) \xrightarrow{n \rightarrow \infty} \mu$ ) converge presque sûrement vers la loi.

c)  $\text{prj}_{\langle \mu_1, \dots, \mu_n \rangle}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right)$  (resp.  $\text{prj}_{e(\mu_1, \dots, \mu_n)}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right)$ ) est un estimateur de la loi séquentiellement plus rapide que la loi empirique (c'est-à-dire :  $\forall \mu \in \mathcal{P}_1, \exists N(\mu) : n \geq N(\mu)$

$$\implies \text{prj}_{\langle \mu_1, \dots, \mu_n \rangle}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right) - \mu \Big\|_K \leq \left\| \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} - \mu \right\|_K .$$

d) pour toute application linéaire  $\phi : e((\mu_n : n \in \mathbb{N})) \rightarrow (F, || ||)$   
 (F espace de Banach séparable),  $\phi \circ \text{prj}_{e(\mu_1, \dots, \mu_n)}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right)$  est un  
 estimateur séquentiellement sans biais de  $\phi$ , (c'est-à-dire :

$$\forall \mu \in P_1, \exists N(\mu) : n \geq N(\mu) \implies \int_{\Omega^n} \phi \circ \text{prj}_{e(\mu_1, \dots, \mu_n)}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right) d\mu^{\otimes n}(\omega_1, \dots, \omega_n) = \phi(\mu) .$$

Pour finir ce paragraphe énonçons une proposition relative à l'estimation d'un paramètre linéaire.

Proposition [8].-

Soient  $K = \sum \alpha_i f_i \otimes f_i$  un noyau reproduisant normal, non dégénéré  
 et  $(F, || ||)$  un espace vectoriel normé.

1) Si  $\phi$  est une application linéaire de  $e(\mu_1, \dots, \mu_k)$  dans  
 F alors :

a)  $\phi \circ \text{prj}_{e(\mu_1, \dots, \mu_n)}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right)$  est un estimateur fortement sans biais  
 d'ordre n de la restriction de  $\phi$  à  $P(E) \cap e(\mu_1, \dots, \mu_k)$

b)  $\inf_{\mu \in P(E) \cap e(\mu_1, \dots, \mu_k)} \mu^{\otimes n} \left[ \omega \in \Omega^n : \left| \phi \circ \text{prj}_{e(\mu_1, \dots, \mu_k)}^K \left( \frac{1}{n} \sum_{i=1}^n \delta_{\omega_i} \right) - \phi(\mu) \right| \right]$

$$< a \varepsilon, \forall n \geq N_1 ] \geq 1 - 16 n_0(a \varepsilon) \frac{\sum_{i=1}^{\infty} \alpha_i}{a^2 \varepsilon^2 N_1} ,$$

où a est la norme de  $\phi$  et  $\varepsilon > 0$ .

2) Si  $\phi$  est une application linéaire définie sur  $e(\mu_n : n \in \mathbb{N})$   
 alors on peut exhiber (à l'aide des projections orthogonales sur des sous-espaces  
 vectoriels de dimension finie) un estimateur  $(S_n)$  séquentiellement fortement  
 sans biais et convergeant presque sûrement de la restriction de  $\phi$  à  
 $P(E) \cap e(\mu_n ; n \in \mathbb{N})$ .

La démonstration de ces deux propositions est essentiellement basée  
 sur le théorème de "Glivenko-Cantelli" cité précédemment.

## CHAPITRE II

### ESTIMATION D'UN PARAMETRE FONCTIONNEL PAR PROJECTION SUR LES ENSEMBLES DE MESURES.

---

Dans ce chapitre on garde les notations du chapitre I.

#### I - INTRODUCTION.

Soient  $(\Omega, \mathcal{A}, P)$  un espace probabilisé,  $(E, d)$  un espace métrique séparable,  $P_1$  un sous-ensemble de  $P(E)$ ,  $\phi$  une application de  $e(P_1)$  dans  $F$  sous-espace de  $\mathbb{R}^E$  prolongeable en une application linéaire à  $e(P_1)$  et  $X$  une variable aléatoire définie sur  $(\Omega, \mathcal{A})$  à valeurs dans  $(E, \mathcal{B}_d)$ , dont la loi  $P_X \in P_1$ . On se propose d'étudier la convergence d'un estimateur de  $\phi(P_X)$  (proposé par C. Guilbart dans [8]) suivant des modes que nous préciserons plus loin, lorsque la taille  $n$  de l'échantillon augmente indéfiniment. Soit  $K : E \times E \rightarrow \mathbb{R}$ , un noyau reproduisant induisant un produit scalaire  $\langle \cdot, \cdot \rangle_K$  sur  $M(E)$  supposé de dimension infinie. A partir de  $n$  observations une estimation directe d'un paramètre de dimension infinie n'est pas possible. Il faut nous ramener donc à l'estimation d'un paramètre de dimension finie pour espérer une information raisonnable pour cela on se donne :

Une suite d'entiers naturels  $(q(n))$  telle que  $q(n) \rightarrow \infty$ .

Pour tout  $n \in \mathbb{N}^*$  on pose  $H_n = e(\mu_1, \dots, \mu_{q(n)})$  l'espace engendré par  $\mu_1, \dots, \mu_{q(n)}$  muni de la topologie induite par  $d_K$ . On considère  $H = \bigcup_{n \in \mathbb{N}} H_n$  limite inductive des  $H_n$ .  $H$  est complet (une limite inductive d'espaces complets est complète).

On peut donc considérer la projection  $\text{prj}_H^K$  de  $M(E)$  sur  $H$ .

On prend :

$$P_1 = P(E) \cap H$$

Comme estimateur de  $\phi(P_X)$  on prend  $\hat{\phi}_n = \phi \circ \text{prj}_H^K \left( \frac{1}{n} \sum_{j=1}^n \delta_{X_j} \right)$ .

Un calcul simple montre qu'il existe un entier  $N_0$  tel que pour  $n \geq N_0$  :

$$(1) \quad \hat{\phi}_n = \frac{1}{n} \sum_{j=1}^n \left[ \sum_{i=1}^{q(n)} \sum_{i'=1}^{q(n)} (-1)^{k+i} \frac{\det M_{q(n),ki}}{\det M_{q(n)}} \left( \int K(X_j, y) d_{\mu_k}(y) \right) \phi(\mu_i) \right]$$

où  $M_{q(n)}$  est la matrice  $(\langle \mu_i, \mu_{i'} \rangle_K)_{\substack{1 \leq i \leq q(n) \\ 1 \leq i' \leq q(n)}}$  et  $M_{q(n),ki}$  est la matrice déduite de  $M_{q(n)}$  obtenue en supprimant la  $k^{\text{ième}}$  ligne et la  $i^{\text{ème}}$  colonne.

Remarques :

Si on pose :

$$R_{q(n)}(x, t) = \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} (-1)^{k+1} \frac{\det M_{q(n),ki}}{\det M_{q(n)}} \left( \int K(x, y) d_{\mu_k}(y) \right) \phi(\mu_i) \right](t)$$

Alors :

1) Pour  $n$  assez grand, pour tout  $x \in E$  on a  $\hat{\phi}_n(x) = \frac{1}{n} \sum_{j=1}^n R_{q(n)}(X_j, x)$  avec  $R_{q(n)}(\cdot, t) \in H_K$ .

2) Si  $P_1$  est dominé par une mesure  $\mu$   $\sigma$ -finie telle que  $\forall v \in P_1 \frac{dv}{d\mu} \in L^2(\mu)$  et si on prend  $\phi(v) = \frac{dv}{d\mu}$ , alors pour  $n$  assez grand  $\hat{\phi}_n(x)$  est un estimateur de la densité qui a la forme d'un estimateur bien connu et qui a été étudié dans [18] par Denis Bosq et Jacques Bleuez.

3) Si on suppose que les  $\mu_i$ ,  $i \in \mathbb{N}^*$  sont orthogonales pour  $\langle \cdot, \cdot \rangle_K$  l'estimateur  $\hat{\phi}_n$  se réduit pour  $n$  assez grand à

$$\hat{\phi}_n = \frac{1}{n} \sum_{j=1}^n \left[ \sum_{i=1}^{q(n)} \frac{\int K(X_j, y) d\mu_i(y)}{\|\mu_i\|_K^2} \phi(\mu_i) \right]$$



II - ESTIMATION SANS BIAIS.-

Dans ce paragraphe on suppose que  $\phi(P_1) \in L^2(\mu)$  où  $\mu$  est une mesure  $\sigma$ -finie définie sur  $(E, \mathcal{B}_d)$  telle que  $L^2(\mu)$  soit séparable. On munit  $L^2(\mu)$  de sa tribu borélienne  $\mathcal{C}$ .

Définition (D. Bosq [19]).

. T est un estimateur (d'ordre n) de  $\phi$  si T est une v.a.  $\mathcal{B}_d^n - \mathcal{C}$  mesurable.

. Soit T un estimateur d'ordre n de  $\phi$ , on dira que  $T^*$  est une version régulière de T si

- a)  $T^*$  est une fonction numérique  $\mathcal{B}_d^{n+1} - \mathcal{C}$  mesurable .
- b)  $T^*(x_1, \dots, x_n)$  est une version de  $T(x_1, \dots, x_n)$  pour tout  $(x_1, \dots, x_n) \in E^n$  .

Proposition II.1.-

a)  $\hat{\phi}_1(x) = \phi \circ \text{prj}_{\mathcal{H}_n}^K(\delta_{X_1})$  est un estimateur sans biais d'ordre 1 de  $\phi/P_1 \cap \mathcal{H}_n$  .

b) Si la tribu des événements symétriques est complète pour  $P_1 \cap \mathcal{H}_n$  alors  $\hat{\phi}_n(x) = \frac{1}{n} \sum_{j=1}^n \left[ \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} (-1)^{k+i} \frac{\det M_{q(n),ki}}{\det M_{q(n)}} \left( \int K(X_j, y) d\mu_k(y) \right) \right] \phi(\mu_i)(x) \right]$

définit une version régulière de l'unique estimateur sans biais d'opérateur de covariance minimum pour  $\phi/P_1 \cap \mathcal{H}_n$  .

c)  $\hat{\phi}_n(x) = \phi \circ \text{prj}_{\mathcal{H}_n}^K\left(\frac{1}{n} \sum_{j=1}^n \delta_{X_j}\right)$  est un estimateur asymptotiquement sans biais de  $\phi/P_1 \cap \mathcal{H}$  .

Pour la démonstration de cette proposition il suffit d'adopter celle de la proposition 1 de ([19], page 3) pour a) et b) et de remarquer pour c) que pour tout  $v \in P_1 \cap \mathcal{H} \quad ] n_0 : n \geq n_0 \quad v \in P_1 \cap \mathcal{H}_n$  .

Pour une étude très détaillée de l'estimation sans biais d'un paramètre fonctionnel voir [19], dans cet article D. Bosq a donné un résultat très précis : dans  $L^2(\mu)$ , en supposant les densités bornées et  $\mathcal{P}_1$  convexe, une condition nécessaire et suffisante pour que la densité soit estimable sans biais, est que l'ensemble des densités à  $\mathcal{P}_1$  soit inclus dans un espace de dimension finie.

Dans [8] C. Guilbart a donné une condition nécessaire et suffisante pour qu'un paramètre soit estimable sans biais cette condition porte sur la continuité au sens de la topologie faible du paramètre à estimer.

### III - CONDITIONS SUFFISANTES DE CONVERGENCE.

#### 1) Notation et hypothèses.

On suppose que la suite  $(\mu_n)_{n \in \mathbb{N}^*}$  de  $M(E)$  est telle que  $\forall i \in \mathbb{N}^* \mu_i = \mu_i^+ - \mu_i^-$  avec  $\mu_i^+$  et  $\mu_i^-$  deux mesures positives dominées par une mesure  $\mu$   $\sigma$ -finie non atomique définie sur  $E$  et telle que :

-  $L^2(\mu)$  soit séparable ,

-  $\forall i \in \mathbb{N}^* (\frac{d\mu_i^+}{d\mu}, \frac{d\mu_i^-}{d\mu}) \in L^2(\mu) \times L^2(\mu)$  .

On pose  $f_i = \frac{d\mu_i}{d\mu} = \frac{d\mu_i^+}{d\mu} - \frac{d\mu_i^-}{d\mu}$  ,  $i \in \mathbb{N}^*$

comme paramètre  $\phi$  on choisit la densité :  $\phi : \mathcal{P}_1 \longmapsto L^2(\mu)$  .  
 $v \longmapsto \frac{dv}{d\mu}$

On pose :

.  $D = \frac{d\mathcal{P}_1}{d\mu}$  et  $D_F = \{f \in D : \limsup_{n \rightarrow \infty} \sup_{t \in F} \left| \int R_{q(n)}(x,t) d\mu(x) - f(t) \right| = 0\}$  ,  $F$

étant un sous-espace de  $E$ .

.  $D_n(t) = \left| \hat{f}_n(t) - f(t) \right|$  ,  $t \in E$  ,  $n$  assez grand .

.  $f_n(t) = E_f \hat{f}_n(t)$

.  $S_n(t) = \hat{f}_n(t) - f_n(t)$  ,  $t \in E$  ,  $n$  assez grand

$$g_{i_n}(t) = \sum_{k=1}^{q(n)} (-1)^{k+i} \frac{\det M_{q(n),ki}}{\det M_{q(n)}} \left( \int K(t,y) du_k(y) \right), i \in \mathbb{N}^*.$$

On suppose que les densités de  $D$  sont bornées.

On considère les hypothèses suivantes :

$$H_1) \text{ (resp. } H'_1) \exists h : E \rightarrow \mathbb{R}^+ \text{ (resp. définies sur } F \text{ et bornée) tel que}$$

$$\sup |f_i(x)| \leq h(x), x \in E.$$

$$H_2) \exists \beta > 0, \exists \xi_0 \text{ tel que } \left\| \sum_{i=1}^{q(n)} |f_i| \right\|_{L^2(\mu)}^2 \leq \xi_0 q(n)^\beta$$

$$H_3) \exists C_0 > 0 \text{ tel que } \left\| \sum_{i=1}^{q(n)} |g_{i_n}| \right\|_{L^2(\mu)}^2 \leq C_0 q(n)^\beta$$

$$H_4) \text{ (resp. } H'_4) \exists M > 0 : \sup_{i \in \mathbb{N}^*} \sup_{x \in E} |g_i(x)| \leq M \text{ (resp. } \forall i = 1, \dots, q(n)$$

$$\sup_{x \in E} |g_i(x)| \leq M q(n)^\beta).$$

$$H_5) \exists A > 0 \exists x_0 \in E \text{ tel que } \sum_{i=1}^{q(n)} g_i(x_0) f_i(x_0) \geq A q(n)^\beta$$

$$H_6) \exists B(t) > 0 \text{ tel que } E \left( \sum_{i=1}^{q(n)} g_{i_n}(X_1) f_i(t) \right)^2 \geq B(t)$$

2) Convergence simple en moyenne d'ordre  $p$  ( $p = 1, 2$ ).

Proposition III.2.2.-

Considérons les propriétés suivantes :

$$a) \frac{q(n)^\beta}{n} \rightarrow 0,$$

$$b) E_f(D_n(t)) \rightarrow 0, t \in E, f \in \frac{dP_1}{d}$$

$$c) E_f(D_n^2(t)) \rightarrow 0, t \in E, f \in \frac{dP_1}{d}$$

Alors sous les hypothèses  $H_1)$  et  $H_3)$  on a les implications  $a) \Rightarrow c) \Rightarrow b)$ .

Démonstration : On va montrer que les hypothèses  $H_1)$  et  $H_3)$  entraînent l'hypothèse II de [18] ce qui nous permettra de conclure.

$$\begin{aligned} \int R_{q(n)}^2(x,t) d\mu(x) &= \int_E \left( \sum_{i=1}^{q(n)} g_{i_n}(x) f_i(t) \right)^2 d\mu(x) \\ &\leq \int \left( \sum_{i=1}^{q(n)} |g_{i_n}(x)| |f_i(t)| \right)^2 d\mu(x) \\ &\leq \int \left( \sum_{i=1}^{q(n)} |g_{i_n}(x)| h(t) \right)^2 d\mu(x) \quad (H_1) \\ &\leq h^2(t) \int \left( \sum_{i=1}^{q(n)} |g_{i_n}(x)| \right)^2 d\mu(x) \end{aligned}$$

$$\int R_{q(n)}^2(x,t) d\mu(x) \leq C_0 h^2(t) q(n)^\beta \quad (H_3) .$$

Il suffit de poser  $B(t) = C_0 h^2(t)$ ,  $r(n) = q(n)$  et  $\alpha = \beta$  et d'utiliser la proposition 4 de D. Bosq et J. Bleuez ([18], p. 488) pour affirmer que a)  $\Rightarrow$  c)  $\Rightarrow$  b).

### 3) Convergence en moyenne quadratique intégrée.

Proposition III.3.1.-

Sous  $H_2)$  et  $H_4)$ , si  $\frac{q(n)^\beta}{n} \rightarrow 0$  on a alors

$$E_f \int (\hat{f}_n(x) - f(x))^2 d\mu(x) \rightarrow 0, \quad f \in \frac{dP_1}{d\mu}$$

Démonstration :

$$\begin{aligned} E \int (\hat{f}_n(x) - f(x))^2 d\mu(x) &= E \int (\hat{f}_n(x) - f_n(x))^2 d\mu(x) + \int (f_n(x) - f(x))^2 d\mu(x) \\ \hat{f}_n(x) - f_n(x) &= \frac{1}{n} \sum_{j=1}^n \left[ \sum_{i=1}^{q(n)} g_{i_n}(X_j) f_i(x) \right] - \sum_{i=1}^{q(n)} \left( \int g_{i_n}(y) f(y) d\mu(y) \right) f_i(x) \end{aligned}$$

$$\begin{aligned}
 E \int (\hat{f}_n(x) - f_n(x))^2 d\mu(x) &\leq \frac{1}{n} \int E \left( \sum_{i=1}^{q(n)} g_{i_n}(X_1) f_i(x) \right)^2 d\mu(x) \\
 &\leq \frac{1}{n} \int \left( \int \left( \sum_{i=1}^{q(n)} g_{i_n}(t) f_i(x) \right)^2 f(t) d\mu(t) \right) d\mu(x) \\
 &\leq \frac{1}{n} \int \left( \int \left( \sum_{i=1}^{q(n)} |g_{i_n}(t)| |f_i(x)| \right)^2 f(t) d\mu(t) \right) d\mu(x) \\
 &\leq \frac{1}{n} M^2 \int \left( \sum_{i=1}^{q(n)} |f_i(x)|^2 \right) f(t) d\mu(t) d\mu(x) \quad (H_4)
 \end{aligned}$$

$$E \int (\hat{f}_n(x) - f_n(x))^2 d\mu(x) \leq M^2 \xi_0 \frac{q(n)^\beta}{n}$$

et comme pour tout  $f \in \frac{dP_1}{d\mu}$ ,  $\int (f_n(x) - f(x))^2 d\mu(x) \rightarrow 0$

on a le résultat cherché.

4) Convergence simple presque complète et en moyenne d'ordre  $p$  ( $p = 1, 2$ ).

Enonçons d'abord un lemme qui donne la vitesse de convergence de  $\hat{f}_n(x)$  vers  $f(x)$ .

Lemme III.4.1.-

Sous les hypothèses  $H_1)$ ,  $H_3)$  et  $H_4)$  (resp.  $H_1)$ ,  $H_3)$  et  $H_4'$ ), pour tout  $t \in E$ , pour tout  $\varepsilon > 0$  on a :

$$P(D_n(t) \geq \varepsilon) \leq 2 \exp\left(-\rho(t) \frac{n \varepsilon}{q(n)^\beta}\right), \quad n \text{ assez grand,}$$

$$\rho(t) = \frac{1}{4M h(t)} \left[ \left(1 + \frac{\sup f(x)}{M h(t)} C_0 h^2(t)\right) \cdot \text{Log}\left(1 + \frac{M h(t)}{(\sup f(x)) C_0 h^2(t)}\right) - 1 \right]$$

(resp.  $P(D_n(t) \geq \varepsilon) \leq 2 \exp\left(-\rho(t) \frac{n \varepsilon}{q(n)^{2\beta}}\right)$ ,  $n$  assez grand,

$$\rho(t) = \frac{1}{4M h(t)} \left[ \left(1 + \frac{\sup f(x) C_0 h^2(t)}{M h(t)}\right) \cdot \text{Log}\left(1 + \frac{M h(t)}{(\sup f(x)) C_0 h^2(t)}\right) - 1 \right]$$

Démonstration : On a montré dans la proposition III.2.2. que  $H_1)$  et  $H_3)$  entraîne II) de [18]. Montrons que  $H_1)$  et  $H_4)$  (resp.  $H_1)$  et  $H'_4)$ )  $\implies$  I) de [18].

$$\sup_{x \in E} |R_{q(n)}(x, t)| = \sup_{x \in E} \left| \sum_{i=1}^{q(n)} g_{i_n}(x) f_i(t) \right| \leq \sup_{x \in E} \sum_{i=1}^{q(n)} |g_{i_n}(x)| |f_i(t)|$$

$$\begin{aligned} \sup_{x \in E} |R_{q(n)}(x, t)| &\leq M \sum_{i=1}^{q(n)} |f_i(t)| \text{ sous } H_4 \text{ (resp. } \leq M q(n)^\beta \sum_{i=1}^{q(n)} |f_i(t)| \text{ sous } H'_4) \\ &\leq M q(n)^\beta h(t) \text{ sous } H_1) \text{ (resp. } \leq M q(n)^{2\beta} h(t) \text{ sous } H_1) \end{aligned}$$

il suffit de poser  $A_1(t) = M h(t)$  et  $\alpha = \beta$  (resp.  $\alpha = 2\beta$ ) pour pouvoir appliquer le lemme 1 de ([18] p. 490).

Remarque : La démonstration du lemme 1 de ([18] p. 490) est basée sur une inégalité de type exponentiel dû à Hoeffding (cf. [21]).

Avant de donner une condition suffisante de convergence presque complète et en moyenne d'ordre 1 et 2 rappelons la définition de la convergence presque complète (p. co.).

$(Y_n)_{n \in \mathbb{N}}$  une suite de v.a.r. converge p.co. vers  $Y$  si  $\forall \epsilon > 0 \sum P(|Y_n - Y| \geq \epsilon) < +\infty$ .

Remarque : La convergence presque complète entraîne la convergence presque sûre (lemme de Borel - Cantelli).

Proposition III.4.4.-

Soient les propriétés suivantes :

a)  $\forall n > 0 \sum \exp(-n \frac{n}{q(n)^\beta}) < +\infty$

b)  $D_n(t) \xrightarrow[p.co.]{} 0$  et  $E(D_n(t)) \rightarrow 0$  ;  $t \in E$ ,  $f \in \frac{dP_1}{d\mu}$

c)  $D_n(t) \xrightarrow[p.co.]{} 0$  et  $E(D_n^2(t)) \rightarrow 0$  ;  $t \in E$ ,  $f \in \frac{dP_1}{d\mu}$

Alors sous les hypothèses  $H_1)$ ,  $H_3)$  et  $H_4)$  on a les implications a)  $\implies$  c)  $\implies$  b).

Démonstration : a) entraîne c) d'après le lemme II.1. et la proposition III.2.2., il est évident que c)  $\implies$  c).

5) Convergence uniforme presque sûre.

Énonçons d'abord un lemme dû à Hoeffding.

Lemme III.5.2. ([17] page 16).

Soient  $X_1, \dots, X_n$  n v.a.r. indépendantes, centrées, majorées par une constante positive  $b$ , définies sur  $(\Omega, \mathcal{A}, P)$  et ayant des moments d'ordre 1 et 2. Si on pose  $S_n = X_1 + \dots + X_n$ ,  $\sigma^2 = \frac{1}{n} \sum_{k=1}^n E(X_k^2)$  alors pour  $0 < \epsilon < b$  on a :

$$P(S_n \geq n \epsilon) \leq \exp\left(-\frac{n \epsilon}{b} \left[\left(1 + \frac{\sigma^2}{b \epsilon}\right) \log\left(1 + \frac{b \epsilon}{\sigma^2}\right) - 1\right]\right).$$

Proposition III.5.5.-

Sous les hypothèses  $H_1), H'_4)$  et si en plus  $h$  est bornée dans  $E$ . Alors une condition suffisante pour que  $\sup_{x \in E} |\hat{f}_n(x) - f(x)| \xrightarrow{P.S.} 0$ ,  $f \in \frac{dP}{d\mu}$  est que  $\sum_{n=1}^{\infty} q(n) \exp\left(-\lambda \frac{n}{q(n) 2\beta}\right) < +\infty \quad \forall \lambda > 0$ .

Démonstration :

$$\sup_{x \in E} |\hat{f}_n(x) - f(x)| \leq \sup_{x \in E} |\hat{f}_n(x) - f_n(x)| + \sup_{x \in E} |f_n(x) - f(x)|$$

$\hat{f}_n$  est asymptotiquement sans biais donc il suffit de montrer que

$$\sup_{x \in E} |\hat{f}_n(x) - f_n(x)| \xrightarrow{P.S.} 0$$

$$\begin{aligned} \sup_{x \in E} |\hat{f}_n(x) - f_n(x)| &= \sup_{x \in E} \left| \frac{1}{n} \sum_{i=1}^{q(n)} \left( \sum_{j=1}^n g_{i_n}(X_j) \right) f_i(x) - \sum_{i=1}^{q(n)} \left( \int g_{i_n}(y) f(y) d\mu(y) \right) f_i(x) \right| \\ &\leq \sup_{x \in E} \left( \sum_{i=1}^{q(n)} \left| \frac{1}{n} \sum_{j=1}^n g_{i_n}(X_j) - \int g_{i_n}(y) f(y) d\mu(y) \right| |f_i(x)| \right) \\ &\leq \sup_{x \in E} h(x) \left( \sum_{i=1}^{q(n)} \left| \frac{1}{n} \sum_{j=1}^n g_{i_n}(X_j) - \int g_{i_n}(y) f(y) d\mu(y) \right| \right) \quad (H_1) \end{aligned}$$

h étant supposé bornée il suffit donc de montrer que

$$\sum_{i=1}^{q(n)} \left| \frac{1}{n} \sum_{j=1}^n g_{i_n}(X_j) - \int g_{i_n}(y) f(y) d\mu(y) \right| \xrightarrow{\text{P.S.}} 0 \text{ pour cela on va utiliser}$$

le lemme précédent.

Posons 
$$Y_{i_n, j} = g_{i_n}(X_j) - \int g_{i_n}(y) f(y) d\mu(y)$$

Les  $Y_{i_n, j}$  sont des variables aléatoires réelles indépendantes et centrées.

$$|Y_{i_n, j}| \leq 2M q(n)^\beta \quad (H'_4)$$

$$\begin{aligned} E((Y_{i_n, j})^2) &= E(g_{i_n}(X_j))^2 - \left( \int g_{i_n}(y) f(y) d\mu(y) \right)^2 \\ &\leq E(g_{i_n}(X_j)^2) \leq M^2 q(n)^{2\beta} \end{aligned} \quad (H'_4)$$

donc si on pose  $b = 2M q(n)^{2\beta}$  alors  $\frac{\sigma^2}{b\varepsilon} \leq \frac{M^2 q(n)^{2\beta}}{2M q(n)^{2\beta}} = \frac{M}{2}$ .

La fonction  $h(u) = (1+u) \text{Log}(1 + \frac{1}{u}) - 1$  est strictement positive et décroissante sur  $]0, +\infty[$  on a donc

$$\left(1 + \frac{\sigma^2}{b\varepsilon}\right) \text{Log}\left(1 + \frac{b\varepsilon}{\sigma^2}\right) - 1 \geq \left(1 + \frac{M}{2}\right) \text{Log}\left(1 + \frac{2}{M}\right) - 1$$

en appliquant le lemme III.5.2. à  $S_{i, n} = \sum_{j=1}^n Y_{i_n, j}$  on obtient

$$P\left(\frac{1}{n} \sum_{j=1}^n g_{i_n}(X_j) - \int g_{i_n}(y) f(y) d\mu(y) > \varepsilon\right) = P(S_{i, n} > \varepsilon) \leq \exp\left(-\frac{n\varepsilon}{2M q(n)^2} \rho\right)$$

avec  $\rho = \left(1 + \frac{M}{2}\right) \text{Log}\left(1 + \frac{2}{M}\right) - 1$

on a donc :

$$\begin{aligned} P\left(\sum_{i=1}^{q(n)} \left| \frac{1}{n} \sum_{j=1}^n g_{i_n}(X_j) - \int g_{i_n}(y) f(y) d\mu(y) \right| > \varepsilon\right) \\ \leq \sum_{i=1}^{q(n)} P\left(\left| \frac{1}{n} \sum_{j=1}^n g_{i_n}(X_j) - \int g_{i_n}(y) f(y) d\mu(y) \right| > \frac{\varepsilon}{q(n)}\right) \\ \leq 2q(n) \exp\left(-n \frac{\varepsilon}{2M q(n)^{2\beta+1}}\right) \end{aligned}$$

d'où le résultat.



Remarque : dans ([25] p. 102) on trouve une condition analogue avec  $\beta = 1$ .

6) Convergence uniforme en moyenne quadratique.

Proposition III.6.6.-

Sous les hypothèses  $H_1)$  et  $H'_4)$  et si en plus dans  $H_1)$  la fonction  $h$  est bornée. Alors si  $\frac{q(n)^{(\beta+1)^2}}{n} \xrightarrow{n \rightarrow \infty} 0$  on a

$$\sup_{x \in E} E_f (f_n(x) - f(x))^2 \xrightarrow{n \rightarrow \infty} 0 .$$

Démonstration : Il suffit de montrer que  $\sup_{x \in E} E_f (\hat{f}_n(x) - f_n(x))^2 \rightarrow 0$ .

$$\begin{aligned} E(\hat{f}_n(x) - f_n(x))^2 &\leq \frac{1}{n} E \left( \sum_{i=1}^{q(n)} g_{i_n}(X_1) f_i(x) \right)^2 \\ &\leq \frac{1}{n} E \left( \sum_{i=1}^{q(n)} |g_{i_n}(X_1)| |f_i(x)| \right)^2 \\ &\leq \frac{1}{n} h^2(x) E_v \left( \sum_{i=1}^{q(n)} |g_{i_n}(X_1)| \right)^2 \end{aligned} \quad (H_1)$$

$$E(\hat{f}_n(x) - f_n(x))^2 \leq h^2(x) M^2 \frac{q(n)^{(\beta+1)^2}}{n}$$

$$\sup_{x \in E} E(\hat{f}_n(x) - f(x))^2 \leq \left( \sup_{x \in E} h^2(x) \right) M^2 \frac{q(n)^{(\beta+1)^2}}{n}$$

d'où le résultat.

IV - LOI LIMITE DE L'ESTIMATEUR.-

Soit  $t$  un point de  $E$ , pour  $n$  assez grand on veut étudier la loi limite de l'estimateur  $\hat{f}_n(t) = \frac{1}{n} \sum_{j=1}^n \left[ \sum_{i=1}^{q(n)} g_{i_n}(X_j) f_i(t) \right]$  et plus précisément

la loi limite de

$$\frac{\hat{f}_n(t) - f_n(t)}{\sigma(\hat{f}_n(t))} \quad (1)$$

Proposition IV.1.-

Sous  $H_1)$ ,  $H_4)$  et  $H_6)$  et si  $\frac{q(n)^4}{n} \rightarrow 0$  alors

$$\frac{\hat{f}_n(t) - f_n(t)}{\sigma(\hat{f}_n(t))} \xrightarrow{L} N(0,1) .$$

Démonstration : La démonstration est analogue à celle qui se trouve dans [21]. Posons

$$Y_{jq(n)}(t) = \frac{1}{n} \sum_{i=1}^{q(n)} g_{i_n}(X_j) f_i(t) , Z_{jq(n)}(t) = \frac{Y_{jq(n)}(t) - E(Y_{jq(n)}(t))}{\sqrt{n} \sigma(Y_{jq(n)}(t))} .$$

Les  $Z_{jq(n)}(t)$  sont des v.a.r. de même loi, indépendantes, centrées et d'écart type  $\frac{1}{n}$  et on a pour  $n$  assez grand :

$$\frac{\hat{f}_n(t) - f_n(t)}{\sigma(\hat{f}_n(t))} = \sum_{j=1}^n Z_{jq(n)}(t) .$$

Pour trouver la loi de (1) on va appliquer le théorème limite centrale ([22] page 328) aux variables  $Z_{jq(n)}(t)$ . Il suffit alors de montrer :

a)  $\forall \epsilon > 0 \quad \sum_{j=1}^n P(|Z_{jq(n)}(t)| > \epsilon) \rightarrow 0 .$

b)  $\exists \alpha > 0$  tel que  $\sum_{j=1}^n \sigma^2(Z_{jq(n)}(t)(\alpha)) \rightarrow 1$  et  $\sum_{j=1}^n E(Z_{jq(n)}(t)(\alpha)) \rightarrow 0$

avec  $Z_{jq(n)}(t)(\alpha)$  la variable aléatoire tronquée à  $\alpha$ .

Démonstration de a) :

Soit  $\epsilon > 0 \quad \sum_{j=1}^n P(|Z_{jq(n)}(t)| > \epsilon) = n P(|Z_{1q(n)}(t)| > \epsilon)$

$$\leq \frac{n E |(Y_{1q(n)}(t) - E(Y_{1q(n)}(t)))^4|}{\epsilon^4 n^2 \sigma^4(Y_{1q(n)}(t))}$$

$$\begin{aligned} E[(Y_{1q(n)}(t) - E(Y_{1q(n)}(t)))^4] &\leq E[ (|Y_{1q(n)}(t)| + |E(Y_{1q(n)}(t))|)^4 ] \\ &\leq E\left[\left(\frac{1}{n} Mq(n) h(t) + \frac{1}{n} Mq(n) h(t)\right)^4\right] \quad (H'_1 \text{ et } H_4) \end{aligned}$$

$$E[(Y_{1q(n)}(t) - E(Y_{1q(n)}(t)))^4] \leq \frac{16}{n^4} M^4 q(n)^4 h^4(t) .$$

$$\sigma^2(Y_{1q(n)}(t)) = E(Y_{1q(n)}^2(t)) - (E(Y_{1q(n)}(t)))^2$$

$$E(Y_{1q(n)}^2(t)) \geq B(t)/n^2 \quad (H_9)$$

$$E(Y_{1q(n)}(t)) \leq M h(t) \frac{q(n)}{n} \rightarrow 0 \quad (H_1 \text{ et } H_4)$$

donc  $\exists N_0 : n \geq N_0 \quad E(Y_{1q(n)}(t))^2 \leq \frac{B(t)}{2n^2}$  et par conséquent pour

$$n \geq N_0 \quad \sigma^2(Y_{1q(n)}(t)) \geq \frac{B(t)}{2n^2} ,$$

d'où pour  $n$  assez grand  $\sum_{j=1}^n P(|Z_{jq(n)}(t)| > \varepsilon) \leq \frac{n}{4} \frac{16 M^4 q(n)^4}{n^2 n^4} h^4(t) \times \frac{4 n^4}{B^2(t)}$

$$\sum_{j=1}^n P(|Z_{jq(n)}(t)| > \varepsilon) \leq \frac{74 M^4 h^4(t) q^4(n)}{\varepsilon^4 B^2(t) n} \rightarrow 0$$

si  $\frac{q^4(n)}{n} \rightarrow 0$  .

Démonstration de b) :

$$\begin{aligned} |Z_{jq(n)}(t)| &= \left| \frac{Y_{jq(n)}(t) - E(Y_{jq(n)}(t))}{\sqrt{n} \sigma(Y_{jq(n)}(t))} \right| \frac{|Y_{jq(n)}(t)| + |E(Y_{jq(n)}(t))|}{\sqrt{n} \sigma[Y_{jq(n)}(t)]} \\ &\leq \frac{M_{q(n)} h(t)/n + M_{q(n)} h(t)/n}{\sqrt{n} \sigma(Y_{jq(n)}(t))} \quad (H_1 \text{ et } H_4) \end{aligned}$$

$$|Z_{jq(n)}(t)| \leq \frac{2M h(t) q(n)}{\sqrt{B(t)}} \frac{q(n)}{\sqrt{n}} \rightarrow 0 \quad \text{si} \quad \frac{q(n)^4}{n} \rightarrow 0$$

donc  $\forall \varepsilon > 0 \quad Z_{jq(n)}(t)(\alpha) = Z_{jq(n)}(t)$  pour  $n$  assez grand et comme les  $Z_{jq(n)}(t)$  sont centrées et d'écart type  $\frac{1}{n}$  on a le résultat souhaité.

V - COMPARAISON AVEC LA METHODE DES FONCTIONS ORTHOGONALES.

La méthode des fonctions orthogonales a été proposée par Tiago de Oliveira en 1961 (23) et indépendamment par Cencôv en 1962 (24).

On suppose que la densité à estimer  $f \in L^2(\mu)$ . On pose comme estimateur

$$\hat{g}_n(x) = \sum_{i=1}^{q(n)} \hat{a}_{ki,n} e_{ki}(x) \quad \text{avec} \quad \hat{a}_{ki,n} = \frac{1}{n} \sum_{j=1}^n e_{ki}(x_j), \quad (e_k)_{k \in I} \quad \text{étant une}$$

base de  $L^2(\mu)$ ,  $q(n)$  une suite d'entiers positifs tendant vers l'infini.

Si  $L^2(\mu)$  est séparable on prend  $k_i = i \quad \forall i = 1, q(n)$ .

Dans ce paragraphe on suppose que  $L^2(\mu)$  est séparable. On va montrer que dans certains cas la méthode par projection sur les ensembles de mesures coïncide avec la méthode des fonctions orthogonales.

En effet soit  $(f_i)_{i \in \mathbb{N}^*}$  une base orthonormale de  $L^2(\mu)$  telle que  $\forall i \in \mathbb{N}^* \quad \|f_i\|_U = \sup_{x \in E} |f_i(x)| < \infty$ .

Posons 
$$h_i = \frac{f_i}{\|f_i\|_U}, \quad i \in \mathbb{N}^* \quad \text{et} \quad K(x,y) = \sum_{i=1}^{\infty} \alpha_i h_i(x) h_i(y)$$

avec  $(\alpha_i)_{i \in \mathbb{N}^*}$  une suite à termes strictement positifs et telle que  $\sum \alpha_i < \infty$ .

Posons 
$$\mu_i = \mu_i^+ - \mu_i^- \quad \text{avec} \quad \frac{d\mu_i^+}{d\mu} = f_i^+ \quad \text{et} \quad \frac{d\mu_i^-}{d\mu} = f_i^-$$
. On peut alors

énoncer la proposition :

Proposition V.1.-

Si la forme bilinéaire symétrique  $\langle \cdot, \cdot \rangle_K$  définie sur  $M(E)$  est un produit scalaire. Alors :

a) les  $(\mu_i)_{i \in \mathbb{N}}$  sont orthogonaux par rapport à  $\langle \cdot, \cdot \rangle_K$ .

b)  $\forall v \in H = \bigcup_{n \in \mathbb{N}} (H_n \cap P(E))$ . L'estimateur de  $\frac{dv}{d\mu}$  par projection sur les ensembles de mesures  $\hat{f}_n$  coïncide avec l'estimateur de la méthode des fonctions orthogonales  $\hat{g}_n$ .

Démonstration :

$$\begin{aligned} \text{a) } \langle \mu_\ell, \mu_{\ell'} \rangle_K &= \sum_{i=1}^{\infty} \alpha_i \left( \int h_i(x) d\mu_\ell(x) \right) \left( \int h_i(y) d\mu_{\ell'}(y) \right) \\ &= \sum_{i=0}^{\infty} \alpha_i \frac{1}{\|f_i\|_U^2} \left( \int f_i(x) f_\ell(x) d\mu(x) \right) \left( \int f_i(y) f_{\ell'}(y) d\mu(y) \right) \end{aligned}$$

or  $(f_i)_{i \in \mathbb{N}^*}$  est une base orthonormale de  $L^2(\mu)$  donc

$$- \langle \mu_\ell, \mu_\ell \rangle_K = \frac{\alpha_\ell}{\|f_\ell\|_U^2} \|f_\ell\|_{L^2(\mu)}^4$$

$$- \langle \mu_\ell, \mu_{\ell'} \rangle_K = 0 \quad \text{pour } \ell \neq \ell'$$

b) Les  $\mu_i$ ,  $i \in \mathbb{N}^*$  sont orthogonales pour  $\langle \cdot, \cdot \rangle_K$  et par conséquent

$$g_{i_n}(x) = \sum_{k=1}^{q(n)} (-1)^{k+i} \frac{\det M_{q(n),ki}}{\det M_{q(n)}} \left( \int K(X_j, y) d\mu_k(y) \right) \text{ se réduit à}$$

$$g_{i_n}(x) = \frac{\int K(X_j, y) d\mu_i(y)}{\|\mu_i\|_K^2} = \frac{\frac{\alpha_i}{\|f_i\|_U} f_i(X_j) \cdot \|f_i\|_{L^2(\mu)}^2}{\frac{\alpha_i}{\|f_i\|_U^2} \|f_i\|_{L^2(\mu)}^4} = f_i(X_j)$$

$$\text{donc } \hat{f}_n(t) = \frac{1}{n} \sum_{i=1}^{q(n)} \left[ \sum_{j=1}^n f_i(X_j) \right] f_i(t) = \hat{g}_n(t) \quad \forall t \in E.$$

## VI - EXEMPLES. -

Exemple 1 :

On considère  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ .

Soit  $\Delta$  la partition de  $\mathbb{R}$  définie par

$$\begin{aligned} \Delta_1 &= ]-1, +1[ \\ \Delta_2 &= ]-2, +2[ - \Delta_1 \\ &\vdots \\ \Delta_n &= ]-n, +n[ - \bigcup_{i=1}^{n-1} \Delta_i \end{aligned}$$

Le noyau reproduisant  $K = \sum_{i=1}^{\infty} \frac{1}{2^i} 1_{\Delta_i} \otimes 1_{\Delta_i}$  définie un produit scalaire  $\langle \cdot, \cdot \rangle_K$  sur  $M(\mathbb{R})$ .

Considérons la suite des mesures de probabilités  $(\mu_n)_{n \in \mathbb{N}^*}$  ;  
 $\mu_n = U|_{[n-1, n]}$  loi uniforme sur  $[n-1, n]$ .

On vérifie que les  $\mu_n$ ,  $n \in \mathbb{N}^*$  sont orthogonales pour  $\langle \cdot, \cdot \rangle_K$  et que  $\langle \mu_i, \mu_i \rangle = \frac{1}{2^i}$ .

On vérifie que les hypothèses  $H_1, H_2, H_3, H_4$  (resp.  $H'_4$ ),  $H_5$  et  $H_6$  sont vérifiées avec  $h = 1$ ,  $\xi_0 = 1$ ,  $\beta = 1$ ,  $C_0 = 1$ ,  $M = 1$  et  $A = 1$  pour  $x_0 \in \bigcup_{i=1}^{q(n)} [i-1, i]$ .

Exemple 2 : Polynômes de Legendre sur  $E = [-1, +1]$ .

Les polynômes de Legendre sont définis de la façon suivante :

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2-1)^n, \quad n \in \mathbb{N}^*, \quad x \in [-1, +1].$$

Ils vérifient la relation de récurrence :

$$(n+1) P_{n+1}(x) - (2n+1)x P_n(x) + n P_{n-1}(x) = 0, \quad n \in \mathbb{N}^*.$$

Les premiers polynômes de Legendre sont :

$$P_0(x) = 1 ; P_1(x) = x ; P_2(x) = \frac{3}{2} x^2 - \frac{1}{2} ; P_3(x) = \frac{5}{2} x^3 - \frac{3}{2} x.$$

Les polynômes de Legendre sont bornés, on a  $\sup_{x \in [-1, 1]} |P_n(x)| \leq 1, \forall n \in \mathbb{N}$ .

Le système  $\{((2n+1)/2)^{1/2} P_n\}$  est orthonormal dans  $L^2[-1, +1]$ .

Soient  $(\alpha_i)_{i \in \mathbb{N}}$  une série à termes strictement positifs et  $K$  le noyau reproduisant sur  $[-1, +1]$  définie par  $K(x, y) = \sum_{i \in \mathbb{N}} \alpha_i P_i(x) P_i(y)$ .

Considérons le système de mesures  $\{\mu_i\}_{i \in \mathbb{N}}$  suivant :

$$\mu_i = \mu_i^+ - \mu_i^- \quad \text{avec} \quad \frac{d\mu_i^+}{d\mu} = P_i^+ \quad \text{et} \quad \frac{d\mu_i^-}{d\mu} = P_i^-, \quad i \in \mathbb{N}$$

Posons  $H = \bigcup_{n \in \mathbb{N}} E(\mu_0, \dots, \mu_{q(n)})$  et  $P_1 = \{v \in \mathcal{P}([-1, +1]) \cap H : \frac{dv}{d\mu} \text{ soit borné}\}$ .

Proposition VI.1.-

La forme bilinéaire symétrique  $\langle \cdot, \cdot \rangle_K$  sur  $M(F)$  est un produit scalaire induisant la topologie faible sur  $M^+(E)$  et pour tout

$f \in \frac{d(P(E) \cap H)}{d\mu}$  l'estimateur  $\hat{f}_n$  de  $f$  par la méthode de projection sur les ensembles de mesures coïncide avec l'estimateur  $\hat{g}_n$  de  $f$  par la méthode des fonctions orthogonales.

Démonstration :

. Soit  $m \in M(E)$  :

$$\begin{aligned} \|m\|_K^2 &= \sum_{i \in \mathbb{N}} \alpha_i \left( \int_E P_i(x) dm(x) \right)^2 = 0 \iff \forall i \in \mathbb{N} \int_E P_i(x) dm(x) = 0 \\ \implies \forall Q = \sum \gamma_i P_i \int_E Q(x) dm(x) &= 0 \implies \forall i \in \mathbb{N} \int_E x^i dm(x) = 0 . \\ \text{Or } (x^i)_{i \in \mathbb{N}} \text{ est dense dans } C[-1, +1] &\implies \forall f \in C[-1, +1] \int f(x) dm(x) = 0 \\ \iff m = 0 \text{ donc } \langle \cdot, \cdot \rangle_K &\text{ est un produit scalaire.} \end{aligned}$$

. Pour montrer que  $\langle \cdot, \cdot \rangle_K$  induit la topologie faible sur  $M^+(E)$  il suffit d'appliquer le théorème de faible comparaison du chapitre I page 5.

. Les  $\mu_i$ ,  $i \in \mathbb{N}$  sont orthogonales pour  $\langle \cdot, \cdot \rangle_K$ , on trouve :

$$\langle \mu_\ell, \mu_{\ell'} \rangle_K = \alpha_\ell \frac{4}{(2\ell+1)^2} \text{ et } \langle \mu_\ell, \mu_{\ell'} \rangle_K = 0 \text{ pour } \ell \neq \ell'$$

Une vérification simple montre que  $\forall x \in E \hat{f}_n(x) = \hat{g}_n(x)$ .

Pour cet exemple on vérifie que les hypothèses  $H_1$  (resp.  $H'_1$ ),  $H_2$ ,  $H_3$ ,  $H'_4$ ,  $H_5$ ,  $H_6$  sont vérifiées.

Exemple 3 : Base de Haar.

Ici on reprend l'exemple qui a été traité dans [20] pour montrer que pour un système de mesures  $\{\mu_i\}_{i \in \mathbb{N}}$  bien défini la méthode par projection sur les ensembles de mesures coïncide avec la méthode des fonctions orthogonales. Soit  $(E, \mathcal{B}, \mu)$  un espace probabilisé, où  $\mathcal{B}$  est engendrée par une suite croissante  $\mathcal{B}_n$  de sous-tribus de  $\mathcal{B}$  construites de la façon suivante :

•  $B_0 = \{\phi, E\}$

•  $B_1$  est engendrée par la partition de  $E$   $P_1 = \{B_0^{(1)}, B_1^{(1)}\}$  avec  $\mu(B_0^{(1)}) = \mu(B_1^{(1)}) = \frac{1}{2}$  d'une manière générale si  $n = 2^k + \ell$   $\ell = 0, 1, \dots, 2^k - 1$ .

•  $B_n$  est engendrée par la partition de  $E$  :  $P_n = \{B_0^{(n)}, B_1^{(n)}, \dots, B_n^{(n)}\}$  avec  $\mu(B_0^{(n)}) = \mu(B_1^{(n)}) = \dots = \mu(B_{2^k-1}^{(n)}) = \frac{1}{2^{k+1}}$

$$B_{2\ell}^{(n)} \cup B_{2\ell+1}^{(n)} \subset B_{2\ell}^{(n-1)}$$

$$\mu(B_{2\ell+2}^{(n)}) = \dots = \mu(B_n^{(n)}) = \frac{1}{2^k}.$$

La suite de partition de  $E$ ,  $(P_n)_{n \in \mathbb{N}}$  constitue un système de Haar (voir (26) page 51).

A ce système, on associe une base de Haar définie de la façon suivante :

$$U_0(t) = 1$$

$$U_n(t) = \begin{cases} + 2^{k/2} & \text{si } t \in B_{2\ell}^{(n)} \\ - 2^{k/2} & \text{si } t \in B_{2\ell+1}^{(n)} \\ 0 & \text{sinon} \end{cases} \quad \text{pour } n = 2^k + \ell \quad (k \in \mathbb{N}, \ell = 0, \dots, 2^k - 1)$$

comme noyau reproduisant sur  $E$  on prend :

$$K(x, y) = \sum_{i=0}^{\infty} \alpha_i \frac{U_i(x)}{\|U_i\|_U} \cdot \frac{U_i(y)}{\|U_i\|_U} \quad \text{avec } \forall i \in \mathbb{N} \quad \alpha_i > 0 \quad \text{et} \quad \sum_{i \in \mathbb{N}} \alpha_i < \infty$$

comme système de mesures  $\{\mu_i\}_{i \in \mathbb{N}}$  on prend :

$$\mu_i = \mu_i^+ - \mu_i^- \quad , \quad \text{avec } \mu_i^+ \quad \text{tel que} \quad \frac{d\mu_i^+}{d\mu} = U_i^+ \quad \text{et} \quad \frac{d\mu_i^-}{d\mu} = U_i^-.$$

Si on considère  $P_1 = \{v \in \mathcal{P}(E) \cap H : \frac{dv}{d\mu} \text{ soit bornée sur } E\}$ . Alors on vérifie de  $\langle \cdot, \cdot \rangle_K$  est un produit scalaire sur  $M(E)$  et pour tout  $v \in P_1$  l'estimateur de  $\frac{dv}{d\mu}$  par projection sur les ensembles de mesures  $\hat{f}_n$  coïncide avec l'estimateur  $\hat{g}_n$  de la méthode des fonctions orthogonales.



### CHAPITRE III

#### ESTIMATION DU DEGRE DE CONTAMINATION.

##### I - GENERALITES.-

Dans un problème statistique, l'analyse avant l'expérience du phénomène aléatoire étudié, conduit le plus souvent au choix d'un modèle idéal  $(E, \mathcal{B}, P)$ . Mais à cause des erreurs de plusieurs types, il y a des cas où il y a contradiction entre ce modèle préexpérimental  $P$  et le résultat des expériences. Pour tenir compte de ces erreurs, une notion de déviation est introduite entre le modèle  $P$  et le modèle étendu qui sert à exploiter le résultat  $P_1 : P \subset P_1$ .

Le choix de  $P$  se fait sur deux types de connaissance a priori :

- 1) celle du modèle "théorique"  $P$ .
- 2) celles des conditions expérimentales plausibles qui conduisent à étendre le modèle  $P$  à  $P_1$ .

Cependant les procédures classiques qui sont très sensibles à de légères déviations deviennent instables dès qu'on s'éloigne de  $P$ . Ceci a conduit alors à définir sur les modèles quelque chose ressemblant à une topologie, de manière à rechercher des procédures stables aux voisinages des modèles préexpérimentaux, le voisinage étant le modèle de l'exploitation de l'expérience. Une méthode statistique qui reste stable au voisinage du modèle préexpérimental est dite robuste.

Si  $(P_\theta)_{\theta \in \Theta}^{\otimes n}$  est le modèle initial, on peut citer comme exemples de voisinages :

a) Le modèle contaminé :

$H_c(\epsilon) = \{(1-\epsilon) P_\theta + \epsilon h\}_{\theta \in \Theta, h \in H}$  avec  $\epsilon \in ]0, \frac{1}{2}[$  et  $H$  un ensemble assez vaste.

Ce modèle a été étudié dans [27] ; [28] ; [29].

b) Le modèle à variation totale :

Pour  $0 < \epsilon < 1$  : le modèle à variation totale est :

$H_K(\epsilon) = \bigcup_{\theta \in \Theta} B(P_\theta, \epsilon)$  avec  $B(P_\theta, \epsilon)$  la boule associée à la distance  $K(P, Q) = \sup_{A \in \mathcal{B}} |P(A) - Q(A)|$ , de centre  $P_\theta$  et de rayon  $\epsilon$ . Ce modèle peut être considéré comme une extension du modèle contaminé.

c) Le modèle de Prokhorov :

Le modèle de Prokhorov est défini pour  $\epsilon \in \mathbb{R}^+$  par :

$H_\Pi(\epsilon) = \bigcup_{\theta \in \Theta} B_\Pi(P_\theta, \epsilon)$  avec  $B_\Pi(P_\theta, \epsilon)$  la boule associée à la distance de Prokhorov  $\Pi(P, Q) = \inf \{ \epsilon \mid \forall A \in \mathcal{B} \ P(A) \leq Q(A^\epsilon) + \epsilon \}$  de centre  $P_\theta$  et de rayon  $\epsilon$ .

Dans la suite nous allons nous intéresser au modèle contaminé.

Nous reprenons ici les exemples de C. Huber (voir [30]) qui illustre le mauvais comportement des méthodes usuelles en statistique paramétrique dès qu'on s'éloigne du modèle.

Soit  $P = \{N(\theta, \sigma^2), \sigma^2 \text{ connu}\}_{\theta \in \mathbb{R}}$ .

1) Pour estimer la moyenne  $\theta$  de  $N(\theta, \sigma^2)$  on peut, pour tenir compte des mesures aberrantes, choisir le modèle normal contaminé

$$N_c(\epsilon) = \{(1-\epsilon) N(\theta, \sigma^2) + \epsilon N(\theta, A \sigma^2), \epsilon > 0, A > 0, \sigma^2 > 0, \theta \in \mathbb{R}\}.$$

L'exemple suivant montre qu'il est difficile de déceler une telle contamination :

Soit un échantillon de taille 1 000 de  $X$  telle que

$$P_X = (1-\varepsilon) N(\theta, \sigma^2) + \varepsilon N(\theta, A \sigma^2) \quad \text{avec } \varepsilon = 0,01.$$

En moyenne il y aura 10 observations provenant de la loi contaminante  $N(\theta, A \sigma^2)$ , mais parmi ces observations, seules donneront l'éveil celles qui, se situent en dehors de la zone  $[- 2,56 ; 2,56]$  où les fonctions de répartition des lois contaminées et non contaminées coïncident pratiquement.

Si  $A = 9$  et si  $P_X = N(0,96^2)$  on a  $P(|X| > 2,56) \approx 0,4$ , c'est-à-dire que l'on peut s'attendre à voir 2 observations à droite et 2 observations à gauche de  $[- 2,56 ; 2,56]$  provenant de la loi contaminante  $N(0,96^2)$ .

Si  $P_X = N(0, \sigma^2)$  :  $P(|X| > 2,56) = 12 \cdot 10^{-3}$  donc il y aura six observations à gauche et six observations à droite de  $[- 2,56 ; + 2,56]$  provenant de  $N(0, \sigma^2)$ .

Il ne sera guère possible de détecter que le modèle gaussien ne convient pas.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{est un estimateur sans biais de } \theta.$$

Soit  $I_n(\varepsilon, \theta)$  la quantité d'information de Fisher ,

$$V_n(\varepsilon) = \text{Var}(\bar{X}_n) = \frac{1+8\varepsilon}{n} \quad \text{l'efficacité asymptotique de } \bar{X}_n \quad \text{est } e(\varepsilon) = \lim_{n \rightarrow \infty} \frac{1}{I(n, \varepsilon) V_n(\varepsilon)}$$

un calcul numérique donne le résultat suivant :

$\varepsilon$	0,00	0,02	0,05	0,10
$e(\varepsilon)$	1	0,90	0,80	0,70

(voir [30])

L'efficacité de la moyenne empirique est sensiblement modifiée même avec  $\epsilon = 0,01$ .

Un autre estimateur très simple de la moyenne est la moyenne  $\alpha$ -tronquée : si  $\alpha \in ]0, \frac{1}{2}[$ , on appelle moyenne  $\alpha$ -tronquée de l'échantillon  $(X_1, \dots, X_n)$  de statistiques d'ordre  $(X_{(1)}, \dots, X_{(n)})$  la statistique  $\bar{X}_\alpha = \frac{1}{n-2[n\alpha]} [X_{[n\alpha]+1} + \dots + X_{n-[n\alpha]}]$  où  $[n\alpha]$  désigne la partie entière de  $n\alpha$ .

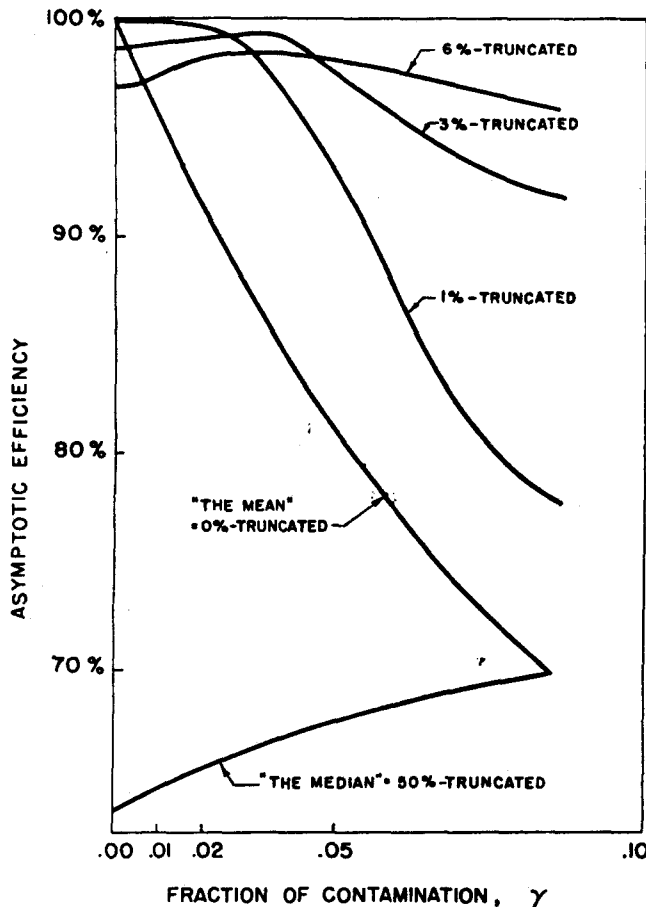
Pour  $\alpha = 0,03$  on a le tableau suivant :

$\epsilon$	0	0,02	0,05	0,10
$e_{\bar{X}_\alpha}$	0,98	0,99	0,97	0,91

(voir [31])

On constate que  $\bar{X}_\alpha$  est plus robuste que  $\bar{X}$ .

Le graphique suivant illustre le comportement de l'efficacité asymptotique des moyennes tronquées en fonctions du degré  $\epsilon$  de contamination.



(voir Tukey [29])

2) Estimation d'un paramètre d'échelle.

Si on suppose maintenant que le modèle est

$$\{(1-\epsilon) N(0, \sigma^2) + \epsilon N(0, 96^2), \sigma^2 > 0\} \text{ avec } \epsilon \text{ connu vérifiant } 0 \leq \epsilon \leq 1.$$

Pour estimer  $\sigma$ , on considère les estimateurs :

$$T_n = \frac{\sigma d_n}{E(d_n)} \quad \text{et} \quad T'_n = \frac{\sigma S_n}{E(S_n)} \quad \text{avec} \quad d_n = \frac{1}{n} \sum_{i=1}^n |X_i| \quad \text{et} \quad S_n = \sqrt{\frac{\sum (X_i)^2}{n}}$$

L'efficacité asymptotique relative de  $T_n$  par rapport à  $T'_n$  en fonction de  $\epsilon$  est : (voir [28], page 3).

$$ARE(\epsilon) = \lim_{n \rightarrow \infty} \frac{\text{Var } T'_n}{\text{Var } T_n} = \frac{\left[ \frac{3(1+80)}{(1+8)^2} - 1 \right]}{4 \left[ \frac{\pi(1+8)}{2(1+2)^2} - 1 \right]}$$

$\epsilon$	0	0.001	0.002	0.005	0.01	0.02	0.05	0.10	0.15	0.25	0.5	1.0
ARE( $\epsilon$ )	0.876	0.948	1.016	1.198	1.439	1.752	2.035	1.903	1.689	1.371	1.017	0.876

Ce tableau montre que pour  $\epsilon = 0$  ce qui correspond au modèle gaussien la performance asymptotique de l'estimateur  $T'_n$  (optimal dans la classe des estimateurs sans biais de  $\sigma$ ) est meilleure que celle de  $T_n$ . Cependant il suffit d'une contamination très faible ( $\epsilon = 0,002$ ) pour que  $T_n$  se comporte aussi bien que  $T'_n$ .

Ceci montre le caractère fragile de l'optimalité face à des modèles perturbés même très légèrement.

## II - ESTIMATION DU DEGRÉ DE CONTAMINATION.

### 1) Introduction.

Soient  $(E, d)$  un espace métrique séparable,  $P$  et  $Q$  deux sous-ensembles de  $P(E)$ , dominés par une mesure  $\mu$   $\sigma$ -finie et telle que :

$$\forall (g, h) \in (G = \frac{dP}{d\mu}, H = \frac{dQ}{d\mu}), (g, h) \in L^2(\mu) \times L^2(\mu). \text{ Soit } g \in G.$$

Pour une densité  $f$  appartenant au modèle contaminé :

$$(1) \quad F_c(g) = \{(1-\varepsilon)g + \varepsilon h, h \in H, 0 < \varepsilon < 1\}$$

connaître le degré de contamination est une situation concrète très particulière.

Aussi nous allons chercher à l'estimer. Soit  $(X_1, \dots, X_n)$  un échantillon qui suit une loi de densité  $f = (1-\varepsilon)g + \varepsilon h \in F_c(g)$ .

Avant de citer l'exemple de Bickel-Lehmann qui montre que si on suppose  $h$  inconnu il n'existe pas en général d'estimateur sans biais de  $\varepsilon$  énonçons leur célèbre théorème :

#### Théorème (Bickel-Lehmann, [32] page 1525)

Soient  $P_0$  un sous-ensemble convexe de  $P(E)$ ,  $F$  un espace de Banach séparable et  $\psi$  une application de  $P_0 \mapsto F$ . Alors les conditions suivantes équivalentes :

1) degré de  $\psi = n$  (c'est-à-dire le plus petit entier  $n$  tel que  $\psi$  soit estimable sans biais d'ordre  $n$ ).

2) -  $\forall \nu$  et  $\gamma \in P_0, \forall \alpha \in [0, 1], \psi(\alpha\nu + (1-\alpha)\gamma)$  est un polynôme en  $\alpha$  de degré inférieur ou égal à  $n$ .

-  $\forall \nu \in P_0$ , il existe  $\gamma$  tel que  $\psi(\alpha\nu + (1-\alpha)\gamma)$  soit exactement de degré  $n$ .

Exemple : (voir [32] page 1527).

Soit  $F_c(g) = \{(1-\varepsilon)g + \varepsilon h, h \in H, 0 \leq \varepsilon \leq 1\}$  où  $g$  est la densité de la loi normale réduite,  $H$  un ensemble convexe, complet, équivalent à  $g$  et tel que si  $(1-\varepsilon_1)g + \varepsilon_1 h_1 = (1-\varepsilon_2)g + \varepsilon_2 h_2$  alors  $\varepsilon_1 = \varepsilon_2$  et  $h_1 = h_2$

Soit

$$\begin{aligned} \psi : F_c(g) &\longmapsto [0,1] \\ (1-\varepsilon)g + \varepsilon h &\longmapsto \varepsilon \end{aligned}$$

Supposons que  $\psi$  est estimable sans biais.

$F_c(g)$  et  $\psi$  sont convexes donc d'après le théorème de Bickel-Lehmann  $\psi$  est de degré 1. Soit  $T$  un estimateur sans biais d'ordre 1 de  $\psi$ .

$\forall f = (1-\varepsilon)g + \varepsilon h \in F_c(g) : \int T(x) f(x) dx = \psi(f) = \varepsilon$  cette condition est équivalente à

$$\begin{cases} \int T(x) g(x) dx = 0 \\ \int T(x) h(x) dx = 1 \end{cases}$$

contradiction avec  $H$  complet et équivalent à  $g$ .

## 2) Méthode des fonctions orthogonales.

### a) Efficacité de l'estimateur et borne de Cramer-Rao.

Soit  $P_0$  un sous-ensemble de  $P(E)$  dominé par une mesure de probabilité  $\mu_0$  tel que  $\forall h \in \frac{dP_0}{d\mu_0}, h \in L^2(\mu_0)$ . Considérons le modèle contaminé :

$$F_c(h) = \{(1-\varepsilon)1 + \varepsilon h, h \in H, h \neq 1 \text{ p.s.}, 0 < \varepsilon < 1\},$$

$h$  supposé connu. Soit  $(X_1, \dots, X_n)$  un échantillon de loi  $\nu$  absolument continue par rapport à  $\mu_0$  tel que  $f = \frac{d\nu}{d\mu_0} = (1-\varepsilon)1 + \varepsilon h \in F_c(h)$ .

Notons par  $\| \cdot \|$  la norme de  $L^2(\mu_0)$ .

Posons  $\alpha = (\|h\|^2 - 1)^{1/2}$  ;  $e_1 \equiv 1$  ,  $e_2 = \frac{h-1}{\alpha}$ .

Supposons en plus que  $h^3 \in L^1(\mu_0)$ .

Proposition II.2.1.-

1)  $(e_1, e_2)$  est un système orthonormal de  $L^2(\mu_0)$ .

2)  $\hat{\varepsilon}_n = \frac{1}{n \alpha^2} \sum_{i=1}^n (h(X_i) - 1)$  est un estimateur sans biais de  $\varepsilon$ .

3) La variance de  $\hat{\varepsilon}_n$  s'écrit  $\text{var } \hat{\varepsilon}_n = \frac{1}{n \alpha^4} [a \varepsilon^2 + b \varepsilon + c]$  avec

$$a = -\alpha^4, \quad b = \int h^3 d\mu_0 - 3 \int h^2 d\mu_0 + 2, \quad c = \alpha^2$$

4) Si on pose  $\psi(\varepsilon) = a \varepsilon^2 + b \varepsilon + c$  alors

$$n \alpha^4 \sup_{0 < \varepsilon < 1} \text{var } \hat{\varepsilon}_n = \begin{cases} \cdot \psi(-\frac{b}{2a}) & \text{si } 0 < -\frac{b}{2a} < 1 \\ \cdot \psi(0) & \text{si } -\frac{b}{2a} \leq 0 \\ \cdot \psi(1) & \text{si } -\frac{b}{2a} \geq 1 \end{cases}$$

Démonstration :

1)  $\int e_1 \cdot e_2 d\mu_0 = \int 1 \cdot (\frac{h-1}{\alpha}) d\mu_0 = \frac{1}{\alpha} (\int d\mu_0 - 1) = 0$

$\int e_2^2 d\mu_0 = \int (\frac{h-1}{\alpha})^2 d\mu_0 = \frac{1}{\alpha^2} \int (h^2 - 2h + 1) d\mu_0 = \frac{\alpha^2}{\alpha^2} = 1$

2)  $f \in E(e_1, e_2) \cap L^2(\mu_0)$  : Les coefficients de Fourier de  $f$  sont

$a_1 = \int f \cdot 1 d\mu_0 = 1$

$a_2 = \int f \cdot e_2 d\mu_0 = \int f \cdot (\frac{h-1}{\alpha}) d\mu_0 = \frac{1}{\alpha} \left[ \int ((1-\varepsilon) \cdot 1 + \varepsilon h) d\mu_0 - \int ((1-\varepsilon) \cdot 1 + \varepsilon h) d\mu_0 \right]$

$a_2 = \frac{1}{\alpha} \left[ \int (1-\varepsilon) h d\mu_0 + \varepsilon \int h^2 d\mu_0 - 1 \right] = \frac{\varepsilon}{\alpha} (\|h\|^2 - 1)$



Soit

$$a_2 = \varepsilon \alpha$$

un estimateur sans biais de  $a_2$  est  $\frac{1}{n} \sum_{i=1}^n e_2(X_i)$  d'où un estimateur sans biais de  $\varepsilon$  :

$$\hat{\varepsilon}_n = \frac{1}{n \alpha^2} \sum_{i=1}^n [h(X_i) - 1]$$

$$3) \text{ var } \hat{\varepsilon}_n = \frac{1}{n \alpha^4} \text{ var } h(X_1)$$

$$\begin{aligned} \text{var } h(X_1) &= \int h^2 f \, d\mu_0 - \left( \int h f \, d\mu_0 \right)^2 \\ &= \int h^2 ((1-\varepsilon) \cdot 1 + \varepsilon h) \, d\mu_0 - \left( \int h [(1-\varepsilon) \cdot 1 + \varepsilon h] \, d\mu_0 \right)^2 \\ &= \varepsilon^2 \left[ 2 \int h^2 \, d\mu_0 - \left( \int h^2 \, d\mu_0 \right)^2 - 1 \right] + \varepsilon \left[ - \int h^2 \, d\mu_0 + \int h^3 \, d\mu_0 + 2 - 2 \int h^2 \, d\mu_0 \right] + \int h^2 \, d\mu_0 - 1 \end{aligned}$$

$$\text{var } h(X_1) = - \left( \int h^2 \, d\mu_0 - 1 \right)^2 \varepsilon^2 + \left( \int h^3 \, d\mu_0 - 3 \int h^2 \, d\mu_0 + 2 \right) \varepsilon + \left( \int h^2 \, d\mu_0 - 1 \right)$$

c'est un polynôme en  $\varepsilon$  de degré deux.

$$4) \frac{\partial \psi}{\partial \varepsilon} = 2a\varepsilon + b = 0 \iff \varepsilon = -\frac{b}{2a}$$

$$\frac{\partial^2 \psi}{\partial \varepsilon^2} = 2a < 0 \quad (a < 0) \quad \text{donc } \psi \text{ est concave et}$$

$$\sup_{0 < \varepsilon < 1} \psi(\varepsilon) = \begin{cases} \cdot \psi\left(-\frac{b}{2a}\right) & \text{si } 0 < -\frac{b}{2a} < 1 \\ \cdot \psi(0) & \text{si } -\frac{b}{2a} \leq 0 \\ \cdot \psi(1) & \text{si } -\frac{b}{2a} \geq 1 \end{cases}$$

Considérons le modèle  $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \prod_{i=1}^n [(1-\varepsilon) \cdot 1 + \varepsilon h(x_i)])$ ,  $\varepsilon \in ]0, 1[$  ouvert de  $\mathbb{R}$ .

$$\text{On a : } E_{\varepsilon}(\hat{\varepsilon}_n) = \int \hat{\varepsilon}_n(x_1, \dots, x_n) \prod_{i=1}^n [(1-\varepsilon) \cdot 1 + \varepsilon h(x_i)] \, d\mu_0(x_1) \dots d\mu_0(x_n) = \varepsilon$$

$\varepsilon \mapsto f_{\varepsilon}(x) = (1-\varepsilon) \cdot 1 + \varepsilon h(x)$  est dérivable pour tout  $x \in \mathbb{R}$  et

$$\frac{\partial f_{\varepsilon}(x)}{\partial \varepsilon} = h - 1 \leq h \in L^1(\mu_0) \quad \text{donc } x \mapsto \frac{\partial f_{\varepsilon}(x)}{\partial \varepsilon} \text{ est intégrable et on a}$$

$\frac{\partial}{\partial \varepsilon} \int f_{\varepsilon}(x) d\mu_0(x) = \int \frac{\partial f_{\varepsilon}}{\partial \varepsilon}(x) d\mu_0(x) = 0$  , on est donc dans les conditions d'application de l'inégalité de Cramer-Rao et on peut énoncer :

Proposition 11.2.2.-

Sous les hypothèses suivantes :

$H_1) h \in L^4(\mu_0)$

$H_2) \text{ pour } \varepsilon \in ]0, 1[, \int \frac{(h-1)^2}{1+\varepsilon(h-1)} d\mu_0 \text{ est finie}$

$H_3) \text{ pour } \varepsilon \in ]0, 1[, \int \frac{(h-1)^4}{1+\varepsilon(h-1)} d\mu_0 \text{ est majorée .}$

On a les résultats suivants :

1) pour  $\varepsilon \in ]0, 1[, \text{var } \hat{\varepsilon}_n \geq \frac{1}{n \int \frac{(h-1)^2}{(1-\varepsilon(h-1))} d\mu_0}$

2) l'efficacité de  $\hat{\varepsilon}_n$  "au voisinage de  $\varepsilon = 0$ " est de la forme

$$e_{\varepsilon}(\hat{\varepsilon}_n) = 1 - \varepsilon \frac{\int (h-1)^3 d\mu_0}{\int (h-1)^2 d\mu_0} + o(\varepsilon^2) .$$

Démonstration :

1)  $\frac{\partial}{\partial \varepsilon} \text{Log } f_{\varepsilon} = \frac{\partial}{\partial \varepsilon} \text{Log } (1+\varepsilon(h-1)) = \frac{h-1}{1+\varepsilon(h-1)}$

$1 + \varepsilon(h-1) = 0 \iff \varepsilon h = \varepsilon - 1 \iff h = \frac{\varepsilon-1}{\varepsilon} < 0$  impossible donc  $1 + \varepsilon(h-1) \neq 0$   
pour tout  $\varepsilon \in ]0, 1[$ .

La quantité d'information de Fisher est définie par

$$I(\varepsilon) = E_{\varepsilon} \left( \frac{\partial \text{Log } f_{\varepsilon}}{\partial \varepsilon} \right)^2 = \int \left( \frac{h-1}{1+\varepsilon(h-1)} \right)^2 f_{\varepsilon} d\mu_0 = \int \frac{(h-1)^2}{1+\varepsilon(h-1)} d\mu_0 < \infty$$

d'après l'hypothèse  $H_2$ ) et comme  $h \neq 1$  p.s.  $\implies I(\varepsilon) \neq 0$  d'où

$$\text{Var } \hat{\varepsilon}_n \geq \frac{1}{n \int \frac{(h-1)^2}{(1+\varepsilon(h-1))} d\mu_0}$$

d'après Cramer-Rao.

2) Calculons l'efficacité de  $\hat{\varepsilon}_n$  "au voisinage de  $\varepsilon = 0$ ".

"au voisinage de  $\varepsilon = 0$ "  $\text{var } \hat{\varepsilon}_n = \frac{1}{n \left[ \int h^2 d\mu_0 - 1 \right]}$  donc l'efficacité de  $\hat{\varepsilon}_n$  "au voisinage de  $\varepsilon = 0$ " est

$$e(\hat{\varepsilon}_n) = \frac{\text{Var } \hat{\varepsilon}_n}{\text{borne de C.R.}} = \frac{\int \frac{(h-1)^2}{(1+\varepsilon(h-1))} d\mu_0}{\int h^2 d\mu_0 - 1}$$

$$\text{Or on a : } \frac{1}{1+\varepsilon(h-1)} = 1 - \varepsilon(h-1) + \frac{\varepsilon^2(h-1)^2}{1+\varepsilon(h-1)}$$

$$\implies \int \frac{(h-1)^2}{1+\varepsilon(h-1)} d\mu_0 = \int (h-1)^2 d\mu_0 - \varepsilon \int (h-1)^3 d\mu_0 + \varepsilon^2 \int \frac{(h-1)^4}{1+\varepsilon(h-1)} d\mu_0$$

et comme  $\int \frac{(h-1)^4}{1+\varepsilon(h-1)} d\mu_0$  est supposée majorer on a :

$$e(\hat{\varepsilon}_n) = 1 - \varepsilon \frac{\int (h-1)^3 d\mu_0}{\int (h-1)^2 d\mu_0} + O(\varepsilon^2)$$

Remarques :

.  $H_2$ ) et  $H_3$ ) sont vérifiés si pour tout  $x \in E$   $h(x) - 1 \geq 0$ .

. Le minimum de  $\text{var } \hat{\varepsilon}_n$  n'est pas atteint pour  $\varepsilon \in ]0, 1[$  car le modèle n'est pas exponentiel.

b) Loi limite de l'estimateur.

Théorème.-

$$\frac{\sum_{i=1}^n [h(X_i) - (1+n\epsilon)]}{\sqrt{n} \sigma(h(X_1))} \xrightarrow{L} N(0,1) .$$

Démonstration : Posons  $\psi(X_i) = \frac{h(X_i) - 1}{\alpha}$ ,  $\psi$  est de carré intégrable

$$\begin{aligned} \frac{\hat{\epsilon}_n - \epsilon}{\sigma \hat{\epsilon}_n} &= \frac{\frac{1}{n \alpha^2} \sum_{i=1}^n (h(X_i) - 1) - \epsilon}{\alpha^2} = \frac{\frac{1}{n} \sum_{i=1}^n (\psi(X_i) - n\epsilon)}{\frac{1}{\sqrt{n}} \sigma(\psi(X_j))} \\ &= \frac{\sum_{i=1}^n (\psi(X_i) - n\epsilon)}{\sqrt{n} \sigma(\psi(X_j))} \rightarrow N(0,1) \end{aligned}$$

d'après le théorème central limite.

c) Exemples.

Exemple 1 : modèle uniforme contaminé :

Soit  $\mu_0 : \frac{d\mu_0}{dx} = \frac{1}{\theta_0} 1_{[0, \theta_0]}$  . Considérons le modèle  $\{(1-\epsilon) \cdot 1 + \epsilon h, 0 < \epsilon < 1\}$

avec  $h = \frac{\theta_0}{\theta_1} 1_{[0, \theta_0]}$  et  $0 < \theta_1 < \theta_0 < 2 \theta_1$  .

$$\int h^i(x) d\mu_0(x) = \left(\frac{\theta_0}{\theta_1}\right)^{i-1}, \quad i \geq 1$$

$$b = \int h^3 d\mu_0 - 3 \int h^2 d\mu_0 + 2 = \left(\frac{\theta_0}{\theta_1}\right)^2 - 3\left(\frac{\theta_0}{\theta_1}\right) + 2$$

On trouve que

$$\text{var } \hat{\epsilon}_n = \frac{1}{n} \left[ -\epsilon^2 + \frac{\theta_0^2 - 3\theta_0\theta_1 + 2\theta_1^2}{(\theta_0 - \theta_1)^2} \epsilon + \frac{\theta_1}{(\theta_0 - \theta_1)} \right]$$

$$\cdot - \frac{b}{2a} = \frac{\left(\frac{\theta_0}{\theta_1}\right)^2 - 3\left(\frac{\theta_0}{\theta_1}\right) + 2}{2\left(\frac{\theta_0}{\theta_1} - 1\right)^2} = \frac{1}{2} \left(1 - \frac{\theta_1}{\theta_0 - \theta_1}\right) \leq 0$$

donc  $\sup_{0 < \varepsilon < 1} \psi(\varepsilon) = \psi(0) = \frac{\theta_0 - \theta_1}{\theta_1}$

$$\sup_{0 < \varepsilon < 1} \text{var } \hat{\varepsilon}_n = \frac{1}{n} \frac{\theta_1}{\theta_0 - \theta_1}$$

. La borne de Cramer-Rao est :

$$\int \frac{(h(x)-1)^2}{1+\varepsilon(h(x)-1)} d\mu_0(x) = \int_0^{\theta_1} \frac{\left(\frac{\theta_0}{\theta_1} - 1\right)^2}{1 + \varepsilon \left(\frac{\theta_0}{\theta_1} - 1\right)} \cdot \frac{1}{\theta_0} dx$$

$$\int \frac{(h(x)-1)^2}{1+\varepsilon(h(x)-1)} d\mu_0(x) = \frac{(\theta_0 - \theta_1)^2}{\theta_0(\theta_1 + \varepsilon(\theta_0 - \theta_1))}$$

L'efficacité asymptotique  $e_\varepsilon(\hat{\varepsilon}_n)$  "au voisinage de  $\varepsilon = 0$ " :

$$\frac{\int (h-1)^3 d\mu_0}{\int (h-1)^2 d\mu_0} = \frac{\int h^3 d\mu_0 - 3 \int h^2 d\mu_0 + 2}{\int h^2 d\mu_0 - 1} = \frac{\theta_0}{\theta_1} - 2$$

donc

$$e_\varepsilon(\hat{\varepsilon}_n) = 1 - \left(\frac{\theta_0 - 2\theta_1}{\theta_1}\right) \varepsilon + O(\varepsilon^2)$$

Exemple 2 : modèle normal contaminé.

On considère  $\mu_0 = N(m_0, 1)$  et le modèle  $\{(1-\varepsilon).1 + \varepsilon h, 0 < \varepsilon < 1\}$

avec  $h(x) = e^{-1/2(x-m_1)^2} \cdot e^{1/2(x-m_0)^2}$  ;  $m_0 < m_1$  .

Après tout calcul on trouve :

$$h(x) = e^{(m_0 - m_1)^2 / 2} \cdot e^{-(m_0 - m_1)x}$$

$$\int h^i d\mu_0 = e^{i(i-1)(m_0 - m_1)^2 / 2} ; i = 2, 3$$

$$\text{var } \hat{\epsilon}_n = \frac{1}{n} \left[ -\epsilon^2 + \frac{e^{3(m_0 - m_1)^2} - 3e^{(m_0 - m_1)^2} + 2}{(e^{(m_0 - m_1)^2} - 1)^2} \epsilon + \frac{1}{(e^{(m_0 - m_1)^2} - 1)} \right]$$

$$\text{var } \hat{\epsilon}_n = \frac{1}{n} \left[ -\epsilon^2 + (e^{(m_0 - m_1)^2} + 2) \epsilon + \frac{1}{(e^{(m_0 - m_1)^2} - 1)} \right]$$

$$-\frac{b}{2a} = 1 + \frac{e^{(m_0 - m_1)^2}}{2} > 1 \text{ d'où } \sup_{0 < \epsilon < 1} \psi(\epsilon) = \psi(1)$$

$$\text{donc } \sup_{0 < \epsilon < 1} \text{var } \hat{\epsilon}_n = \frac{1}{n} \left[ -1 + e^{(m_0 - m_1)^2} + 2 + \frac{1}{(e^{(m_0 - m_1)^2} - 1)} \right]$$

$$\sup_{0 < \epsilon < 1} \text{var } \hat{\epsilon}_n = \frac{1}{n} \left[ \frac{e^{2(m_0 - m_1)^2}}{e^{(m_0 - m_1)^2} - 1} \right]$$

Pour la borne de Cramer-Rao et l'efficacité de  $\hat{\epsilon}_n$  "au voisinage de 0" on suppose que  $\epsilon \in ]0, \alpha[$ ,  $\alpha < 1$  et que  $m_0, m_1 \in [a, b]$ .

Dans ces conditions, on va montrer que  $H_2)$  et  $H_3)$  de la proposition II.2.2. sont vérifiées.

En effet  $h(x) = e^{(m_0 - m_1)^2 / 2} \cdot e^{-(m_0 - m_1)x}$  croît de 0 à  $+\infty$

donc  $1 - \alpha \leq 1 - \epsilon \leq 1 + \epsilon(h-1)$  d'où :

$$\int \frac{(h-1)^2}{1 + \varepsilon(h-1)} d\mu_0 \leq \frac{1}{1-\alpha} \int (h-1)^2 d\mu_0 = \frac{1}{1-\alpha} \left( \int h^2 d\mu_0 - 1 \right) < +\infty$$

$$\int \frac{(h-1)^4}{1 + \varepsilon(h-1)} d\mu_0 \leq \frac{1}{1-\alpha} \int (h-1)^4 d\mu_0 \quad \text{et on montre que} \quad \int \frac{(h-1)^4}{1 + \varepsilon(h-1)} d\mu_0$$

peut être majorée par :

$$(e^{(b-a)^2} + 1) [e^{4(b-a)^2} + 2e^{3(b-a)^2} + 3e^{2(b-a)^2} - 3]$$

Pour l'efficacité  $e_\varepsilon(\hat{\varepsilon}_n)$  "au voisinage de  $\varepsilon = 0$ " on trouve

$$e_\varepsilon(\hat{\varepsilon}_n) = 1 - (e^{(m_1 - m_2)^2} - 1) (e^{(m_1 - m_2)^2} + 2) \varepsilon + \mathcal{O}(\varepsilon^2)$$

Si on prend par exemple  $m_0 = 0$  et  $m_1 = 2$ , on a :

$$\text{var } \hat{\varepsilon}_n = \frac{1}{n} \left[ -\varepsilon^2 + (e^4 + 2) \varepsilon + \frac{1}{e^4 - 1} \right]$$

$$\text{var } \hat{\varepsilon}_n \approx \frac{1}{n} [-\varepsilon^2 + 57\varepsilon + 0,017]$$

$$n \sup \hat{v}_\varepsilon = 55,6$$

$$e(\hat{\varepsilon}_n) \approx 1 - 3078 \varepsilon + \mathcal{O}(\varepsilon^2)$$

Exemple 3 : Dans cet exemple on considère  $\mu_0 = N(0, \sigma_1^2)$  et le modèle

$$(1) \{(1-\varepsilon).1 + \varepsilon h, 0 < \varepsilon < 1\} \quad \text{avec} \quad h(x) = \frac{\sigma_1}{\sigma_2} \exp\left(-\frac{x^2}{2} \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right), \quad 0 < \sigma_1 < \sigma_2$$

mais pour que  $\int h^i d\mu_0$  soit fini,  $i = 1, 4$  il faut que  $\sqrt{\frac{3}{4}} < \frac{\sigma_1}{\sigma_2} < 1$ .

Cette condition assez contraignante nous a conduit alors à considérer au lieu de (1) le modèle  $F$  suivant :

Soient  $P, Q$  deux ensembles de probabilités dominés par une mesure  $\mu$   $\sigma$ -finie telle que  $\frac{dP}{d\mu} \times \frac{dQ}{d\mu} \in L^2(\mu) \times L^2(\mu)$ . On pose  $G = \frac{dP}{d\mu}$ ,  $H = \frac{dQ}{d\mu}$  et

$$F = \{f = (1-\epsilon)g + \epsilon h, 0 < \epsilon < 1, (g, h) \in G \times H\}.$$

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  tiré suivant  $\nu$  telle que  $\frac{d\nu}{d\mu} = f = (1-\epsilon)g + \epsilon h \in F$ , le couple  $(g, h)$  est supposé connu.

Posons  $\alpha = \frac{\langle g|h \rangle}{\|g\|^2}$ ,  $e_1 = \frac{g}{\|g\|}$ ,  $e_2 = \frac{h - \alpha g}{\|h - \alpha g\|}$  où  $\| \cdot \|$  désigne

la norme associée au produit scalaire usuelle de  $L^2(\mu)$ .

Proposition II.2.3.-

1)  $(e_1, e_2)$  est un système orthonormal de  $L^2(\mu)$ .

2)  $\hat{\epsilon}_n = \frac{\|g\|}{\sqrt{\|g\|^2 \cdot \|h\|^2 - \langle g|h \rangle^2}} \frac{1}{n} \sum_{i=1}^n e_2(X_i)$  est un estimateur sans

biais de  $\epsilon$ .

3) La variance de  $\hat{\epsilon}_n$  :

$$\text{var } \hat{\epsilon}_n = \frac{1}{n(\|g\|^2 \|h\|^2 - \langle g|h \rangle^2)^2} (u\epsilon^2 + v\epsilon + w)$$

avec  $u = -(\|g\|^2 \|h\|^2 - \langle g|h \rangle^2)^2$

$$v = \|g\|^4 \cdot \int h^2(h-g) d\mu - 2\|g\|^2 \langle g|h \rangle \int gh(h-g) d\mu + \langle g|h \rangle^2 \int g^2(h-g) d\mu$$

$$w = \|g\|^4 \cdot \int h^2 g d\mu - 2\|g\|^2 \langle g|h \rangle \int g^2 h d\mu + \langle g|h \rangle^2 \int g^3 d\mu$$

4) Si on pose  $P(w) = u\epsilon^2 + v\epsilon + w$  alors

$$n(\|g\|^2 \|h\|^2 - \langle g|h \rangle^2)^2 \sup_{0 < \epsilon < 1} \text{var } \hat{\epsilon}_n = \begin{cases} \cdot P(-\frac{v}{2u}) & \text{si } 0 < -\frac{v}{2u} < 1 \\ \cdot P(0) & \text{si } -\frac{v}{2u} \leq 0 \\ \cdot P(1) & \text{si } -\frac{v}{2u} \geq 0 \end{cases}$$



5) Si  $\int \frac{(h-g)^2}{g+\epsilon(h-g)} d\mu(x) < \infty$  alors pour tout  $\epsilon \in ]0,1[$  on a

$$\text{var } \hat{\epsilon}_n \geq \frac{1}{n \int ((h-g)^2 / (g+\epsilon(h-g))) d\mu}$$

Démonstration :

$$1) \int e_1 e_2 d\mu = \int \frac{g}{||g||} \cdot \frac{(h-\alpha g)}{||h-\alpha g||} d\mu = \frac{1}{||g|| ||h-\alpha g||} \left( \int gh d\mu - \alpha \int g^2 d\mu \right) = 0$$

2)  $f \in L^2(\mu) \cap E(e_1, e_2)$ . Les coefficients de Fourier par rapport à  $(e_1, e_2)$  sont :

$$a_1 = \int f e_1 d\mu$$

$$a_2 = \int f e_2 d\mu = \int ((1-\epsilon)g + \epsilon h) \frac{h-\alpha g}{||h-\alpha g||} d\mu$$

$$= \frac{1}{||h-\alpha g||} \left[ (1-\epsilon) \int g(h-\alpha g) d\mu + \epsilon \int h(h-\alpha g) d\mu \right]$$

$$a_2 = \frac{||h^2|| \langle -\alpha g | h \rangle}{||h-\alpha g||} \epsilon \quad \text{et comme :}$$

$$||h-\alpha g|| = \frac{\sqrt{||g||^2 ||h||^2 - \langle g | h \rangle^2}}{||g||}$$

alors  $a_2 = ||h-\alpha g|| \epsilon$  d'où  $\epsilon = \frac{a_2}{||h-\alpha g||}$

un estimateur sans biais de  $a_2$  est  $\frac{1}{n} \sum_{i=1}^n e_2(X_i)$  d'où un estimateur sans biais de  $\epsilon$  :

$$\hat{\epsilon}_n = \frac{1}{||h-\alpha g||} \frac{1}{n} \sum_{i=1}^n e_2(X_i)$$

$$\begin{aligned}
 3) \quad \text{var } \hat{\epsilon}_n &= \frac{1}{\| |h-\alpha g| \|^4} \frac{1}{n} \text{var}(h(X_1) - \alpha g(X_1)) \\
 &= \frac{1}{n \| |h-\alpha g| \|^4} [\text{var } h(X_1) - \alpha(Eh(X_1)g(X_1) - Eh(X_1) \cdot Eg(X_1)) + \\
 &\hspace{15em} \alpha^2 \text{var } g(X_1)]
 \end{aligned}$$

Var  $h(X_1) = Eh^2(X_1) - (Eh(X_1))^2$  on trouve :

$$\text{Var } h(X_1) = -\epsilon^2 \left( \int h^2 d\mu - \left( \int hg \right)^2 \right) + \epsilon \left[ \int h^3 - \int h^2 g + 2 \left( \int hg \right)^2 - 2 \left( \int hg \right) \int h^2 \right] + \int h^2 g - \left( \int hg \right)^2$$

$$\begin{aligned}
 \cdot \text{Vg}(X_1) &= -\epsilon^2 \left( \int gh d\mu - \int g^2 d\mu \right)^2 + \epsilon \left[ \int g^2 h d\mu - \int g^3 d\mu + 2 \left( \int g^2 d\mu \right)^2 - 2 \int g^2 d\mu \cdot \int gh d\mu \right] + \\
 &\hspace{15em} \int g^3 d\mu - \left( \int g^2 d\mu \right)^2
 \end{aligned}$$

$$\cdot \text{Eg}(X_1)h(X_1) - \text{Eg}(X_1) \cdot Eh(X_1) =$$

$$\begin{aligned}
 &-\epsilon^2 \left[ \int g^2 d\mu \cdot \int gh d\mu - \int g^2 d\mu \cdot \int h^2 d\mu - \left( \int gh d\mu \right)^2 + \int gh d\mu \cdot \int h^2 d\mu \right] + \\
 &\epsilon \left[ \int gh^2 d\mu - \int g^2 h d\mu + 2 \int g^2 d\mu \cdot \int gh d\mu - \int g^2 d\mu \cdot \int h^2 d\mu - \left( \int gh \right)^2 \right] + \\
 &\int g^2 h d\mu - \int g^2 d\mu \cdot \int gh d\mu .
 \end{aligned}$$

un simple calcul donne alors le résultat 3).

$$4) \text{ Si on pose } P(\epsilon) = u\epsilon^2 + v\epsilon + w$$

$$\text{alors } \sup_{0 < \epsilon < 1} \text{var } \hat{\epsilon}_n = \frac{1}{n \left( \| |g| \|^2 \| |h| \|^2 - \langle g|h \rangle^2 \right)^2} \sup_{0 < \epsilon < 1} P(\epsilon)$$

$$\cdot \frac{\partial P(\epsilon)}{\partial \epsilon} = 2u\epsilon + v = 0 \iff \epsilon = -\frac{v}{2u}$$

$$\cdot \frac{\partial^2 P(\epsilon)}{\partial \epsilon^2} = 2u < 0 \quad \text{donc } P(\epsilon) \text{ est concave et par conséquent}$$

$$\sup_{0 < \epsilon < 1} P(\epsilon) = \begin{cases} \cdot P\left(-\frac{v}{2u}\right) & \text{si } 0 < -\frac{v}{2u} < 1 \\ \cdot P(0) & \text{si } -\frac{v}{2u} \leq 0 \\ \cdot P(1) & \text{si } -\frac{v}{2u} \geq 1 \end{cases}$$

5)  $\varepsilon \in ]0, 1[$  ouvert de  $\mathbb{R}$ .

$\varepsilon \mapsto f_\varepsilon(x) = (1-\varepsilon)g(x) + \varepsilon h(x)$  est dérivable pour tout  $x \in \mathbb{R}$

$\frac{\partial f_\varepsilon(x)}{\partial \varepsilon} = h-g \leq h \in L^1(\mu)$  donc  $x \mapsto \frac{\partial f_\varepsilon}{\partial \varepsilon}$  est intégrable et on a

$$\frac{\partial}{\partial \varepsilon} \int f_\varepsilon(x) dx = \int \frac{\partial f_\varepsilon(x)}{\partial \varepsilon} dx = 0, \text{ on est donc dans les conditions d'application}$$

de l'inégalité de Cramer-Rao. La quantité d'information de Fisher est définie

par 
$$I(\varepsilon) = E_\varepsilon \left( \frac{\partial \text{Log } f_\varepsilon}{\partial \varepsilon} \right)^2 = \int \frac{(h-g)^2}{g+\varepsilon(h-g)} d\mu, \text{ d'où le résultat.}$$

Etude de l'exemple

$$\{ f_\varepsilon = (1-\varepsilon)g + \varepsilon h, 0 < \varepsilon < 1, g(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(1/2)/(x/\sigma_1)^2},$$

$$h(x) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(1/2)(x/\sigma_2)^2} \quad 0 < \sigma_1 < \sigma_2 \}$$

$$\int gh dx = \frac{1}{2\pi\sigma_1\sigma_2} \int e^{-(x^2/2)(1/\sigma_1^2 + 1/\sigma_2^2)} dx = \frac{1}{2\pi\sigma_1\sigma_2} \sqrt{\frac{\pi}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}}$$

$$\int gh d\mu = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}$$

$$\begin{aligned} \bullet \quad ||g||^2 &= \frac{1}{2\sigma_1\sqrt{\pi}} \\ \bullet \quad ||h||^2 &= \frac{1}{2\sigma_2\sqrt{\pi}} \end{aligned}$$

donc

$$u = - \left( \frac{1}{4\pi\sigma_1\sigma_2} - \frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)} \right)^2$$

(1)

$$\cdot \int h^2 g dx = \frac{1}{2\pi\sigma_2^2} \frac{1}{\sqrt{2\pi}\sigma_1} \int e^{-\frac{(x/\sigma_2)^2}{2}} e^{-\frac{(x/\sigma_1)^2}{2}} dx$$

$$\boxed{\int h^2 g dx = \frac{1}{2\pi\sigma_2 \sqrt{2\sigma_1^2 + \sigma_2^2}}}$$

$$\cdot \int h^3 d\mu = \frac{1}{2\sqrt{3}\pi\sigma_2^2}$$

$$\cdot \int g^3 d\mu = \frac{1}{2\sqrt{3}\pi\sigma_1^2}$$

$$\cdot \int g^2 h d\mu = \frac{1}{2\pi\sigma_1 \sqrt{2\sigma_2^2 + \sigma_1^2}}$$

on a donc :

$$v = \frac{1}{4\pi\sigma_1^2} \left[ \frac{1}{2\sqrt{3}\sigma_2^2\pi} - \frac{1}{2\pi\sigma_2 \sqrt{2\sigma_1^2 + \sigma_2^2}} \right] - \frac{1}{\sigma_1 \sqrt{\pi} \sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \left[ \frac{1}{2\pi\sigma_2 \sqrt{2\sigma_1^2 + \sigma_2^2}} - \frac{1}{2\pi\sigma_1 \sqrt{2\sigma_2^2 + \sigma_1^2}} \right]$$

$$+ \frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)} \left[ \frac{1}{2\pi\sigma_1 \sqrt{2\sigma_2^2 + \sigma_1^2}} - \frac{1}{2\sqrt{3}\pi\sigma_1^2} \right]$$

$$v = \frac{1}{8\pi^2\sigma_1^2\sigma_2^2} \left[ \frac{1}{\sqrt{3}\sigma_2} - \frac{1}{\sqrt{2\sigma_1^2 + \sigma_2^2}} \right] - \frac{1}{2\sigma_1\pi^2 \sqrt{2(\sigma_1^2 + \sigma_2^2)}} \left[ \frac{1}{\sigma_2 \sqrt{2\sigma_1^2 + \sigma_2^2}} - \frac{1}{\sigma_1 \sqrt{2\sigma_2^2 + \sigma_1^2}} \right]$$

$$+ \frac{1}{4\pi^2(\sigma_1^2 + \sigma_2^2)\sigma_1} \left[ \frac{1}{\sqrt{2\sigma_2^2 + \sigma_1^2}} - \frac{1}{\sqrt{3}\sigma_1} \right]$$

(2)

$$w = \frac{1}{2\sigma_1 \sqrt{\pi}} \left[ \frac{1}{4\pi \sqrt{\pi} \sigma_1 \sigma_2 \sqrt{2\sigma_1^2 + \sigma_2^2}} - \frac{2}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \cdot \frac{1}{2\pi \sigma_1 \sqrt{2\sigma_2^2 + \sigma_1^2}} \right] + \frac{1}{4\sqrt{3}\pi^2 \sigma_1^2 (\sigma_1^2 + \sigma_2^2)}$$

$$w = \frac{1}{4\pi^2 \sigma_1^2} \left[ \frac{1}{2\sigma_2 \sqrt{2\sigma_1^2 + \sigma_2^2}} - \frac{2}{\sqrt{2(\sigma_1^2 + \sigma_2^2)}(2\sigma_2^2 + \sigma_1^2)} + \frac{1}{\sqrt{3}(\sigma_1^2 + \sigma_2^2)} \right] \quad (3)$$

situer  $-\frac{v}{2u}$  par rapport à  $]0,1[$  est un problème très délicat, aussi pour avoir une idée plus précise traitons le cas

$$\sigma_1 = 1, \quad \sigma_2 = 3$$

D'après les formules (1), (2) et (3) on a :

$$u = -\frac{1}{4\pi^2} \left[ \frac{1}{\sigma} - \frac{1}{10} \right]^2 = -\frac{1}{4\pi^2 \cdot 225}$$

$$v = \frac{1}{2\pi^2} \left[ \frac{1}{12} \left( \frac{1}{3\sqrt{3}} - \frac{1}{\sqrt{11}} \right) + \frac{1}{\sqrt{20}} \left( \frac{1}{\sqrt{19}} - \frac{1}{3\sqrt{11}} \right) + \frac{1}{20} \left( \frac{1}{\sqrt{19}} - \frac{1}{\sqrt{3}} \right) \right]$$

on trouve  $v \approx \frac{0,0023405}{2\pi^2}$

$$w = \frac{1}{4\pi^2} \left( \frac{1}{6\sqrt{11}} - \frac{1}{\sqrt{95}} + \frac{1}{10\sqrt{3}} \right) \approx \frac{0,0053891}{4^2}$$

$-\frac{v}{2u} \approx 0,5266125 > 0$  donc on a :  $\sup_{0 < \epsilon < 1} \hat{\epsilon}_n = \frac{1}{n} \left[ -1 - \frac{v}{u} - \frac{w}{u} \right]$

$$\sup_{0 < \epsilon < 1} \hat{\epsilon}_n \approx \frac{1}{n} \times 1,2657725 \quad (4)$$

3) Méthode par projection sur les ensembles mesures.

a) Efficacité de l'estimateur et borne de Cramer-Rao.

Considérons le modèle contaminé  $F = (1-\epsilon) \mu_0 + \epsilon \mu_1, 0 < \epsilon < 1$

où  $\mu_0, \mu_1$  sont deux mesures de probabilités supposées connues et absolument continues par rapport à une mesure  $\mu$   $\sigma$ -finie définie sur  $(E, \mathcal{B}_d)$ .

Soit  $K$  un noyau reproduisant sur  $E$ , borné,  $B_d \otimes B_d$  mesurables et induisant un produit scalaire  $\langle \cdot, \cdot \rangle_K$  sur  $M(E)$ .

Soit  $(X_1, \dots, X_n)$  un échantillon de loi  $\nu = (1-\epsilon) \mu_0 + \epsilon \mu_1$ ,  $0 < \epsilon < 1$ . On veut estimer  $\epsilon$  par la méthode de projection sur les ensembles de mesures.

Posons

$$\left\{ \begin{array}{l} \cdot f_0 = \frac{d\mu_1}{d\mu}, \quad f_1 = \frac{d\mu_1}{d\mu} \\ \cdot \gamma = \mu_0 \\ \cdot \beta = \mu_1 - \lambda \mu_0 \quad \text{avec} \quad \lambda = \frac{\langle \mu_0, \mu_1 \rangle_K}{\|\mu_0\|_K^2} \end{array} \right.$$

Proposition 11.3.1.-

1)  $\gamma$  et  $\beta$  sont orthogonales pour  $\langle \cdot, \cdot \rangle_K$ .

2)  $\hat{\epsilon}_n = \frac{1}{n} \sum_{i=1}^n \frac{K(X_i, y) d\beta(y)}{\|\beta\|_K^2}$  est un estimateur sans biais de  $\epsilon$ .

3)  $\text{var } \hat{\epsilon}_n = \frac{1}{n \|\beta\|_K^4} [a \epsilon^2 + b\epsilon + c]$  avec

$$a = - \|\beta\|_K^4 = - \left( \|\mu_1\|_K^2 - \frac{\langle \mu_0, \mu_1 \rangle_K^2}{\|\mu_0\|_K^2} \right)$$

$$b = \int \left( \int K(x, y) d\mu_1(y) \right)^2 d(\mu_1 - \mu_0)(x) + \frac{\langle \mu_0, \mu_1 \rangle_K^2}{\|\mu_0\|_K^4} \int \left( \int K(x, y) d\mu_0(y) \right)^2 d(\mu_1 - \mu_0)(x) - 2 \frac{\langle \mu_0, \mu_1 \rangle_K}{\|\mu_0\|_K^2} \int \left( \int K(x, y) d\mu_1(y) \cdot \int K(x, y) d\mu_0(y) \right) d(\mu_1 - \mu_0)(x)$$

$$c = \int \left( \int K(x, y) d\mu_1(y) \right)^2 d\mu_0(x) - 2 \frac{\langle \mu_0, \mu_1 \rangle_K}{\|\mu_0\|_K^2} \int \left( \int K(x, y) d\mu_1(y) \cdot \int K(x, y) d\mu_0(y) \right) d\mu_0(x) + \frac{\langle \mu_0, \mu_1 \rangle_K^2}{\|\mu_0\|_K^4} \int \left( \int K(x, y) d\mu_0(y) \right)^2 d\mu_0(x)$$

4°) si on pose  $\psi(x) = a \epsilon^2 + b\epsilon + c$

$$\text{alors } \sup_{0 < \epsilon < 1} \psi(\epsilon) = \begin{cases} \cdot \psi(-\frac{b}{2a}) & \text{si } 0 < -\frac{b}{2a} < 1 \\ \cdot \psi(1) & \text{si } -\frac{b}{2a} \geq 1 \\ \cdot \psi(0) & \text{si } -\frac{b}{2a} \leq 0 \end{cases}$$

5°) Si pour tout  $\epsilon \in ]0, 1[$   $\int \frac{(f_1 - f_0)^2}{f_0 + \epsilon(f_1 - f_0)} d\mu < \infty$

$$\text{alors } \text{var } \hat{\epsilon}_n \geq \frac{1}{n \int ((f_1 - f_0)^2 / (f_0 + (f_1 - f_0))) d\mu}$$

Démonstration :

$$1) \langle \gamma, \beta \rangle_K = \langle \mu_0, \mu_1 - \frac{\langle \mu_0, \mu_1 \rangle_K}{\|\mu_0\|_K^2} \mu_0 \rangle_K = \langle \mu_0, \mu_1 \rangle_K - \langle \mu_0, \mu_1 \rangle_K = 0$$

$$2) \nu \in E(\gamma, \beta) \iff \exists! \lambda_1, \lambda_2 \text{ tel que } \nu = (1-\epsilon) \mu_0 + \epsilon \mu_1 = \lambda_1 \gamma + \lambda_2 \beta$$

considérons l'application  $\psi : E(\gamma, \beta) \longrightarrow \mathbb{R}$

$$\lambda_1 \gamma + \lambda_2 \beta \longmapsto \lambda_2$$

On sait que  $\hat{\lambda}_{2,n} = \psi \circ \text{prj}_{E(\gamma, \beta)} \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \right) = \frac{1}{n} \sum_{i=1}^n \frac{\int K(X_i, y) d\beta(y)}{\|\beta\|_K^2}$  est un

estimateur sans biais de  $\lambda_2$ .

$$\begin{aligned} \langle \nu, \beta \rangle_K &= \langle \lambda_1 \gamma + \lambda_2 \beta, \beta \rangle_K = \lambda_2 \|\beta\|_K^2 = \lambda_2 (\|\mu_1\|_K^2 - \frac{\langle \mu_0, \mu_1 \rangle_K^2}{\|\mu_0\|_K^2}) = \lambda_2 \langle \mu_1, \beta \rangle_K \\ &= \langle (1-\epsilon) \mu_0 + \epsilon \mu_1, \beta \rangle_K = \epsilon \langle \mu_1, \beta \rangle_K \end{aligned}$$

donc  $\epsilon = \lambda_2$  d'où un estimateur sans biais de  $\epsilon$  :  $\hat{\epsilon}_n = \frac{1}{n} \sum_{i=1}^n \frac{\int K(X_i, y) d\beta(y)}{\|\beta\|_K^2}$

$$3) \text{ var } \hat{\epsilon}_n = \frac{1}{n \|\beta\|_K^4} \text{ var } \int K(X_1, y) d\beta(y)$$

$$\begin{aligned} \text{var} \int K(X_1, y) d\beta(y) &= \int \left( \int K(x, y) d\beta(y) \right)^2 ((1-\epsilon)f_0(x) + \epsilon f_1(x)) d\mu(x) - \\ &\quad ((1-\epsilon) \int \left( \int K(x, y) d\beta(y) \right) f_0(x) d\mu(x) + \epsilon \int \left( \int K(x, y) d\beta(y) \right) f_1(x) d\mu(x))^2 \\ &= (1-\epsilon) \int \left( \int K(x, y) d\beta(y) \right)^2 f_0(x) d\mu(x) + \epsilon \int \left( \int K(x, y) d\beta(y) \right)^2 f_1(x) d\mu(x) \\ &\quad - \epsilon^2 \left( \int \left( \int K(x, y) d\beta(y) \right) f_1(x) d\mu(x) \right)^2 \end{aligned}$$

$$\begin{aligned} \text{var} \int K(X_1, y) d\beta(y) &= -\epsilon^2 \left( \int \int K(x, y) d\beta(y) d\mu_1(x) \right)^2 + \epsilon \left[ \int \left( \int K(x, y) d\beta(y) \right)^2 d\mu_1(x) - \right. \\ &\quad \left. \int \left( \int K(x, y) d\beta(y) \right)^2 d\mu_0(x) \right] + \int \left( \int K(x, y) d\beta(y) \right)^2 d\mu_0(x) \end{aligned}$$

$$\cdot \int \int K(x, y) d\beta(y) d\mu_1(x) = \langle \mu_1, \beta \rangle_K = \|\mu_1\|_K^2 - \frac{\langle \mu_0, \mu_1 \rangle_K^2}{\|\mu_0\|_K^2}$$

$$\begin{aligned} \cdot \int \left( \int K(x, y) d\beta(y) \right)^2 d\mu_1(x) &= \int \left( \int K(x, y) d(\mu_1 - \lambda \mu_0)(y) \right)^2 d\mu_1(x) \\ &= \int \left( \int K(x, y) d\mu_1(y) \right)^2 d\mu_1(x) + \frac{\langle \mu_0, \mu_1 \rangle_K^2}{\|\mu_0\|_K^4} \int \left( \int K(x, y) d\mu_0(y) \right)^2 d\mu_1(x) \\ &\quad - 2 \frac{\langle \mu_0, \mu_1 \rangle_K}{\|\mu_0\|_K^2} \int \left( \int K(x, y) d\mu_1(y) \cdot \int K(x, y) d\mu_0(y) \right) d\mu_1(x) \end{aligned}$$

pour 4) et 5) même démonstration que dans le cas des fonctions orthogonales.

b) Loi limite de l'estimateur.

Proposition II.3.2.-

$$\frac{\hat{\epsilon}_n - \epsilon}{\sigma \hat{\epsilon}_n} \xrightarrow{L} N(0, 1)$$



Démonstration : Même démonstration que dans le cas des fonctions

orthogonales. Il suffit de poser  $\psi(X_1) = \frac{\int K(X_1, y) d\beta(y)}{\|\beta\|_K^2}$ .

c) Exemples :

Exemple 1 : Considérons le modèle  $\{(1-\epsilon) N(0, \sigma_0^2) + \epsilon N(0, \sigma_1^2), 0 < \epsilon < 1\}$   $\sigma_0 < \sigma_1$   
on prend  $K(x, y) = e^{-(x-y)^2/2}$ .

$K$  borné,  $B_{\mathbb{R}} \otimes B_{\mathbb{R}}$  mesurable et induit un produit scalaire sur  $M(\mathbb{R})$ .

$$\begin{aligned} \langle \mu_0, \mu_1 \rangle_K &= \int K(x, y) d\mu_0(x) d\mu_1(y) = \frac{1}{2\pi\sigma_0\sigma_1} \iint e^{-(x-y)^2/2} e^{-x^2/2\sigma_0^2} e^{-y^2/2\sigma_1^2} dx dy \\ &= \frac{1}{2\pi\sigma_0\sigma_1} \int e^{-y^2/2\sigma_1^2} \left( \int e^{-(x-y)^2/2} e^{-x^2/2\sigma_0^2} dx \right) dy \quad (\text{Théorème de Fubini}) \end{aligned}$$

un simple calcul montre que pour  $\sigma > 0$  :

$$\begin{aligned} I_{\sigma}(y) &= \int e^{-(x-y)^2/2} \cdot e^{-x^2/2\sigma^2} dx = \frac{\sqrt{2\pi}\sigma}{\sqrt{\sigma^2+1}} e^{-y^2/2(\sigma^2+1)} \quad \text{donc} \\ \langle \mu_0, \mu_1 \rangle_K &= \frac{1}{2\pi\sigma_0\sigma_1} \cdot \frac{\sqrt{2\pi}\sigma_0}{\sqrt{\sigma_0^2+1}} \int e^{-(y^2/2)} \left[ 1/\sigma_1^2 + 1/(\sigma_0^2+1) \right] dy \end{aligned}$$

on trouve à la fin

$$\langle \mu_0, \mu_1 \rangle_K = \frac{1}{\sqrt{\sigma_0^2 + \sigma_1^2 + 1}}$$

On en déduit :

$$\begin{aligned} \|\mu_0\|_K^2 &= \frac{1}{\sqrt{2\sigma_0^2+1}} \\ \|\mu_1\|_K^2 &= \frac{1}{\sqrt{2\sigma_1^2+1}} \end{aligned}$$

Le coefficient a de la proposition II.3.1. est donc :

$$a = \left( \frac{1}{2\sigma_1^2+1} - \frac{2\sigma_0^2+1}{\sigma_0^2+\sigma_1^2+1} \right)^2$$

(5)

$$\cdot K(x,y) d\mu_1(y) = \frac{1}{\sqrt{2\pi\sigma_1}} I_{\sigma_1}(y) = \frac{1}{\sqrt{\sigma_1^2+1}} e^{-x^2/2(\sigma_1^2+1)}$$

$$\cdot \int \left( \int K(x,y) d\mu_1(y) \cdot \int K(x,y) d\mu_0(y) \right) d\mu_0(x) = \int \frac{1}{\sqrt{\sigma_1^2+1}} e^{-x^2/2(\sigma_1^2+1)} \cdot \frac{1}{\sqrt{\sigma_0^2+1}} e^{-x^2/2(\sigma_0^2+1)} d\mu_0(x)$$

$$\int \left( \int K(x,y) d\mu_1(y) \cdot \int K(x,y) d\mu_0(y) \right) d\mu_0(y) = \frac{1}{\sqrt{\sigma_0^4+2\sigma_0^2\sigma_1^2+3\sigma_0^2+\sigma_1^2+1}}$$

$$\cdot \int \left( \int K(x,y) d\mu_1(y) \right)^2 d\mu_0(x) = \int \frac{1}{(\sigma_1^2+1)} e^{-x^2/(\sigma_1^2+1)} d\mu_0(x)$$

$$\int \left( \int K(x,y) d\mu_1(y) \right)^2 d\mu_0(x) = \frac{1}{\sqrt{\sigma_1^4+2(\sigma_0^2\sigma_1^2+\sigma_0^2+\sigma_1^2)+1}}$$

$$\cdot \int \left( \int K(x,y) d\mu_0(y) \right)^2 d\mu_0(x) = \int \frac{1}{(\sigma_0^2+1)} e^{-x^2/(\sigma_0^2+1)} d\mu_0(x)$$

$$\int \left( \int K(x,y) d\mu_0(y) \right)^2 d\mu_0(x) = \frac{1}{\sqrt{3\sigma_0^4+4\sigma_0^2+1}}$$

$$\cdot \int \left( \int K(x,y) d\mu_1(y) \right)^2 d\mu_1(x) = \frac{1}{\sqrt{3\sigma_1^4+4\sigma_1^2+1}}$$

$$\cdot \int \left( \int K(x,y) d\mu_0(y) \right)^2 d\mu_1(x) = \frac{1}{\sqrt{\sigma_0^4+2(\sigma_0^2\sigma_1^2+\sigma_0^2+\sigma_1^2)+1}}$$

L'expression de b est alors :

$$b = \frac{1}{\sqrt{3\sigma_1^4 + 4\sigma_1^2 + 1}} - \frac{1}{\sqrt{\sigma_1^4 + 2(\sigma_0^2\sigma_1^2 + \sigma_0^2 + \sigma_1^2) + 1}} - \frac{2\sqrt{2\sigma_0^2 + 1}}{\sqrt{(\sigma_0^2 + \sigma_1^2 + 1) \cdot (3\sigma_0^4 + 4\sigma_0^2 + 1)}} + \frac{2\sigma_0 + 1}{\sigma_0^2 + \sigma_1^2 + 1}$$

$$\left( \frac{1}{\sqrt{\sigma_0^4 + 2(\sigma_0^2\sigma_1^2 + \sigma_0^2 + \sigma_1^2) + 1}} - \frac{1}{\sqrt{3\sigma_0^4 + 4\sigma_0^2 + 1}} \right)$$

(6)

L'expression de c est :

$$c = \frac{1}{\sqrt{\sigma_1^4 + 2(\sigma_0^2\sigma_1^2 + \sigma_0^2 + \sigma_1^2) + 1}} - 2\sqrt{\frac{2\sigma_0^2 + 1}{(\sigma_0^2 + \sigma_1^2 + 1)(\sigma_0^4 + 2\sigma_0^2\sigma_1^2 + 3\sigma_0^2 + \sigma_1^2 + 1)}} + \frac{2\sigma_0^2 + 1}{\sigma_0^2 + \sigma_1^2 + 1} \cdot \frac{1}{\sqrt{3\sigma_0^4 + 4\sigma_0^2 + 1}}$$

(7)

Il sort de ces calculs que trouver le  $\sup_{0 < \varepsilon < 1} \text{var } \hat{\varepsilon}_n$  n'est pas chose facile.

Comme dans le cas des fonctions orthogonales et à titre de comparaison traitons le cas

$$\sigma_0 = 1, \quad \sigma_1 = 3$$

d'après les formules (5), (6) et (7) on a :

$$a = \left( \frac{1}{19} - \frac{3}{11} \right)^2 = \frac{2116}{43681}$$

$$b = \frac{1}{\sqrt{280}} - \frac{1}{\sqrt{120}} - \sqrt{\frac{3}{22}} + \frac{3}{11} \left( \frac{1}{\sqrt{40}} - \frac{1}{\sqrt{8}} \right)$$

$$b \approx -0,7244295$$

$$c = \frac{1}{\sqrt{120}} - \sqrt{\frac{3}{88}} + \frac{3}{11\sqrt{8}}$$

$$c \approx -0,1320903$$

$$-\frac{b}{2a} \approx 7,4772696 \text{ donc } n \sup_{0 < \epsilon < 1} \text{ var } \hat{\epsilon}_n = -1 - \frac{b}{a} - \frac{c}{a}$$

$$n \sup_{0 < \epsilon < 1} \text{ var } \hat{\epsilon}_n \approx 16,681305$$

(8)

Remarque : D'après les formules (4) et (8) on peut dire que pour le modèle  $\{(1-\epsilon) N(0,1) + \epsilon N(0,9), 0 < \epsilon < 1\}$  la méthode des fonctions orthogonales est meilleure que la méthode par projection sur les ensembles de mesure.

Exemple 2 : modèle normal contaminé  $\{(1-\epsilon) N(m_0,1) + \epsilon N(m_1,1), 0 < \epsilon < 1, m_0 < m_1\}$ .

On prend  $K(x,y) = e^{-1/2(x-y)^2}$ ,  $\mu_0 = N(m_0,1)$  et  $\mu_1 = N(m_1,1)$

$$\begin{aligned} \langle \mu_0, \mu_1 \rangle_K &= \iint e^{-(x-y)^2/2} \frac{1}{2\pi} e^{-(x-m_0)^2/2} e^{-(x-m_1)^2/2} dx dy \\ &= \frac{1}{2\pi} \int \left( \int e^{-(x-y)^2/2} e^{-(m_0-x)^2/2} dx \right) e^{-(y-m_1)^2/2} dy \quad (\text{théorème de Fubini}). \end{aligned}$$

un simple calcul montre que

$$I(m,y) = \int e^{-(x-y)^2/2} e^{-(m_0-x)^2/2} dx = \sqrt{\pi} e^{-(m-y)^2/2}$$

(9)

donc  $\langle \mu_0, \mu_1 \rangle_K = \frac{\sqrt{\pi}}{2\pi} \int e^{-(m_0-y)^2/4} e^{-(y-m_1)^2/2} dy$

$$\begin{aligned} \langle \mu_0, \mu_1 \rangle_K &= \frac{1}{\sqrt{3}} e^{-(m_0-m_1)^2/6} \\ \|\mu_0\|_K^2 &= \|\mu_1\|_K^2 = \frac{1}{\sqrt{3}} \end{aligned}$$

(10)

$$a = - \left( \|\mu_1\|_K^2 - \frac{\langle \mu_0, \mu_1 \rangle_K^2}{\|\mu_0\|_K^2} \right)^2 = - \left( \frac{1}{\sqrt{3}} - \frac{\sqrt{3}}{3} e^{-(m_0-m_1)^2/3} \right)^2$$

$$a = -\frac{1}{3} (1 - e^{-(m_0 - m_1)^2/3})^2 \quad (11)$$

$$\begin{aligned} \int (K(x,y)d\mu_1(y) \cdot K(x,y)d\mu_0(y))d\mu_0(x) &= \frac{1}{\sqrt{2\pi}} \frac{1}{2\pi} \int (I(m_1,x) \cdot I(m_0,x)) e^{-(x-m_0)^2/2} dx \\ &= \frac{1}{2\sqrt{2\pi}} \int e^{-(m_1-x)^2/4} e^{-(m_0-x)^2/4} e^{-(x-m_0)^2/2} dx \end{aligned}$$

on trouve :

$$\int (K(x,y)d\mu_1(y) \cdot K(x,y)d\mu_0(y))d\mu_0(x) = \frac{1}{2\sqrt{2}} e^{-3(m_0 - m_1)^2/16} \quad (12)$$

$$\int (K(x,y)d\mu_1(y))^2 d\mu_0(x) = \frac{1}{2\pi\sqrt{2\pi}} \int I(m_1,x) e^{-(x-m_0)^2/2} dx$$

$$\int (K(x,y)d\mu_1(y))^2 d\mu_0(x) = \frac{1}{2\sqrt{2}} e^{-(m_0 - m_1)^2/4} \quad (13)$$

$$\int (K(x,y)d\mu_0(y))^2 d\mu_0(x) = \frac{1}{2\sqrt{2}} \quad (14)$$

$$\int (K(x,y)d\mu_1(y)) \cdot (K(x,y)d\mu_0(y))d\mu_1(x) = \frac{1}{2\sqrt{2}} e^{-3(m_0 - m_1)^2/16} \quad (15)$$

Les formules (4) à (10) donnent :

$$b = \frac{1}{2\sqrt{2}} - \frac{1}{2\sqrt{2}} e^{-(m_0 - m_1)^2/4} + e^{-(m_0 - m_1)^2/3} \cdot \left( \frac{1}{2\sqrt{2}} e^{-(m_0 - m_1)^2/4} - \frac{1}{2\sqrt{2}} \right)$$

$$b = \frac{1}{2\sqrt{2}} (1 - e^{-(m_0 - m_1)^2/4}) (1 - e^{-(m_0 - m_1)^2/3}) \quad (16)$$

$$c = \frac{1}{2\sqrt{2}} e^{-(m_0 - m_1)^2/4} - \frac{1}{\sqrt{2}} e^{-(m_0 - m_1)^2/6} \cdot e^{-3(m_0 - m_1)^2/16} + \frac{1}{2\sqrt{2}} e^{-(m_0 - m_1)^2/3}$$

$$c = \frac{1}{2\sqrt{2}} (e^{-(m_0 - m_1)^2/4} - 2 e^{-17(m_0 - m_1)^2/48} + e^{-(m_0 - m_1)^2/3}) \quad (17)$$

(11) et (16) donnent :

$$\boxed{-\frac{b}{2a} = \frac{3}{4\sqrt{2}} \cdot \frac{1-e}{1-e} \cdot \frac{-(m_0-m_1)^2/4}{-(m_0-m_1)^2/3}} \quad (18)$$

d'où :

$$\boxed{\text{var } \hat{\epsilon}_n = \frac{1}{n} \left[ -\epsilon^2 + \frac{3}{2\sqrt{2}} \cdot \frac{1-e}{1-e} \cdot \frac{-(m_0-m_1)^2/4}{-(m_0-m_1)^2/3} \epsilon + \frac{3}{2\sqrt{2}} \cdot \frac{e \cdot \frac{-(m_0-m_1)^2/4}{-2e} - \frac{17(m_0-m_1)^2/48}{+e}}{(1-e \cdot \frac{-(m_0-m_1)^2/3}{})^2} \right]} \quad (19)$$

d'autre part on peut remarquer que

$$\frac{1-e}{1-e} \cdot \frac{-(m_0-m_1)^2/4}{-(m_0-m_1)^2/3} \leq 1 \quad \text{donc d'après (18)} \quad 0 < -\frac{b}{2a} < 1$$

et par conséquent on a d'après la proposition II.3.1.

$$\boxed{n \sup_{0 < \epsilon < 1} \text{var } \hat{\epsilon}_n = \frac{9}{32} \left( \frac{1-e}{1-e} \cdot \frac{-(m_0-m_1)^2/4}{-(m_0-m_1)^2/3} \right)^2 + \frac{3}{2\sqrt{2}} \cdot \frac{e \cdot \frac{-(m_0-m_1)^2/4}{-2e} - \frac{17(m_0-m_1)^2/48}{+e}}{(1-e \cdot \frac{-(m_0-m_1)^2/3}{})^2}} \quad (20)$$

Remarque :

. Par la méthode des fonctions orthogonales et la méthode par projection sur les ensembles de mesures on peut au lieu de considérer le modèle  $\{(1-\epsilon) v_1 + \epsilon v_\epsilon, 0 < \epsilon < 1\}$ , considérer le modèle généralisé  $\{(1-\epsilon) v_1 + \sum_{i=1}^k \epsilon_{k_i} v_{k_i}, \text{ avec } \epsilon_{k_i} \geq 0 \forall i = 1, k \text{ et } \epsilon = \sum_{i=1}^k \epsilon_{k_i}\}$  et estimer les  $\epsilon_{k_i}$ .

III - AUTRES ESTIMATEURS DE  $\epsilon$ .-

Nous avons tout récemment découvert un article de M. Ahmed, N. Giri et B.K. Sinha [33] où d'autres estimateurs du degré de contamination  $\epsilon$  sont étudiés. A titre comparatif citons les principaux résultats de cet article et où les auteurs adoptent le critère suivant : un estimateur  $\hat{\epsilon}_n$  est dit optimal sur une classe d'estimateur  $C$  si  $\sup_{0 < \epsilon < 1} \text{var } \hat{\epsilon}_n = \inf_{\hat{r}_n \in C} \sup_{0 < \epsilon < 1} \text{var } \hat{r}_n$ .

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de densité  $f = (1-\epsilon) f_1 + \epsilon f_2$  avec  $f_1, f_2$  supposées connues et  $0 < \epsilon < 1$  inconnu.

Soit  $F_i$  la fonction de répartition de  $f_i, i = 1, 2$ .

a) L'estimateur de James :

Soit  $\Gamma = \{\lambda \mid F_1(\lambda) \neq F_2(\lambda)\}$  pour  $\lambda \in \Gamma$  l'estimateur de James est défini par :

$$\hat{\epsilon}_{n,\lambda} = \begin{cases} \cdot \epsilon_{n,\lambda}^* = \frac{F_n(\lambda) - F_2(\lambda)}{F_1(\lambda) - F_2(\lambda)}, & \text{si } 0 < \epsilon_{n,\lambda}^* < 1 \\ \cdot 0 & \text{si } \epsilon_{n,\lambda}^* \leq 0 \\ \cdot 0 & \text{si } \epsilon_{n,\lambda}^* \geq 1 \end{cases}$$

où  $F_n$  est la fonction de répartition empirique associée à

$$(X_1, \dots, X_n) : F_n(\lambda) = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, \lambda]}(X_i).$$

La variance de cet estimateur est :

$$\text{var } \hat{\epsilon}_{n,\lambda} = \frac{1}{n} \frac{(\epsilon(F_1(\lambda) + (1-\epsilon) F_2(\lambda)) (\epsilon \bar{F}_1(\lambda) + (1-\epsilon) \bar{F}_2(\lambda)))}{(F_1(\lambda) - F_2(\lambda))^2} = \frac{1}{n} \psi_\epsilon(\lambda)$$

où  $\bar{F}_i = 1 - F_i, i = 1, 2$ .

Proposition [33].-

$$\sup \psi_\varepsilon(\lambda) = \begin{cases} \cdot \frac{1}{4(F_1(\lambda)-F_2(\lambda))^2} & \text{si } F_1(\lambda) \leq \frac{1}{2} < F_2(\lambda) \quad \text{ou} \\ & F_1(\lambda) \leq \frac{1}{2} \leq F_2(\lambda) \quad \text{ou} \\ & F_2(\lambda) \leq \frac{1}{2} < F_1(\lambda) \quad \text{ou} \\ & F_2(\lambda) < \frac{1}{2} \leq F_1(\lambda) \\ \\ \cdot \frac{F_1(\lambda)\bar{F}_1(\lambda)}{(F_1(\lambda)-F_2(\lambda))^2} & \text{si } F_2(\lambda) > F_1(\lambda) \geq \frac{1}{2} \quad \text{ou} \\ & F_2(\lambda) < F_1(\lambda) \leq \frac{1}{2} \\ \\ \cdot \frac{F_2(\lambda)\bar{F}_2(\lambda)}{(F_1(\lambda)-F_2(\lambda))^2} & \text{si } F_1(\lambda) < F_2(\lambda) \leq \frac{1}{2} \quad \text{ou} \\ & F_1(\lambda) > F_2(\lambda) \geq \frac{1}{2} \end{cases}$$

Remarque : Trouver  $\gamma$  qui minimise  $\sup_{0 < \varepsilon < 1} \psi_\varepsilon(\gamma)$  n'est pas toujours chose facile, dans ce cas les auteurs proposent de choisir  $\gamma$  qui maximise  $(F_1(\gamma)-F_2(\gamma))^2$  puisque  $\forall \gamma F_i(\gamma) \bar{F}_i(\gamma) \leq \frac{1}{4}$ ,  $i = 1, 2$ .

Comme application a cette proposition les auteurs ont donné plusieurs exemples :

. Pour le modèle  $\{(1-\varepsilon) N(m_1, 1) + \varepsilon N(m_2, 1), 0 < \varepsilon < 1\}$ , la fonction

$\sup_{0 < \varepsilon < 1} \psi_\varepsilon(\lambda)$  est convexe, son minimum unique est atteint au point

$$\lambda_0 = \frac{m_1 + m_2}{2} \quad \text{et} \quad \sup_{0 < \varepsilon < 1} \psi_\varepsilon(\lambda_0) = \frac{\pi}{8 \int_0^{\lambda_0} e^{-y^2/2} dy}$$

. Pour le modèle  $\{(1-\varepsilon) N(0, \sigma_1) + \varepsilon N(0, \sigma_2), 0 < \varepsilon < 1\}$ .

L'optimum  $\lambda$  n'est pas finie. Dans ce cas M. Ahmed, N. Giri et B.K. Sinhh se servent de leur remarque pour dire que prendre

$\lambda$  telle que  $\exp\left(\left(\frac{\lambda}{\sigma_2}\right)^2 - \left(\frac{\lambda}{\sigma_1}\right)^2\right) = \frac{\sigma_1^2}{\sigma_2^2}$  est peut être un bon choix.



. Pour le modèle uniforme contaminé  $\{(1-\varepsilon) 1_{[0, \theta_1]} + \varepsilon 1_{[0, \theta_2]}\}$  ;  
 $0 < \varepsilon < 1$  ;  $\theta_1 < \theta_2 < 2\theta_1$  }  $\min_{\lambda} \sup_{0 < \varepsilon < 1} \psi_{\varepsilon}(\lambda) = \sup_{\varepsilon} \psi_{\varepsilon}(\theta_1) = \frac{1}{\theta_2 - \theta_1}$  c'est le  
 même résultat qu'on a trouvé avec la méthode des fonctions orthogonales.

Remarque :  $\sup_{0 < \varepsilon < 1} \psi_{\varepsilon}(\lambda)$  peut avoir plusieurs optimums (voir exemple 4 page 353 de [33]).

b) Estimateur basé sur les quantiles :

Soit  $\alpha \in ]0, 1[$  et  $X_{\alpha}$  son quantile inconnu correspondant.

On suppose que  $F_1(X_{\alpha}) \neq F_2(X_{\alpha})$ .

On pose  $\gamma = [n\alpha]$  la partie entière de  $n\alpha$ .

Comme estimateur de  $\varepsilon$  on pose :

$$\hat{\varepsilon}_n = \begin{cases} \cdot \varepsilon_n^* = \frac{\alpha - F_2(X_{(\gamma)})}{F_1(X_{(\gamma)}) - F_2(X_{(\gamma)})} & \text{si } \varepsilon_n^* \in ]0, 1[ \\ \cdot 0 & \text{si } \varepsilon_n^* \leq 0 \\ \cdot 1 & \text{si } \varepsilon_n^* \geq 1 \end{cases}$$

la variance asymptotique de  $\hat{\varepsilon}_n$  est :

$$\text{var } \hat{\varepsilon}_n = \frac{1}{n} \frac{\alpha(1-\alpha)}{(F_1(X_{\alpha}) - F_2(X_{\alpha}))^2} \left[ \frac{f_2(X_{\alpha})}{\varepsilon f_1(X_{\alpha}) + (1-\varepsilon)f_2(X_{\alpha})} \right]^2 = \psi_{\varepsilon}(\alpha)$$

trouver  $\alpha$  qui minimise  $\sup_{0 < \varepsilon < 1} \psi_{\varepsilon}(\alpha)$  n'est pas un problème facile.

On peut citer aussi un estimateur de  $\varepsilon$  basé sur les statistiques d'ordre et qui a été étudié en 1968 par Choi et Bulgren [34], cet estimateur est défini par :

$$\hat{\epsilon}_{n,c} = \begin{cases} \cdot \epsilon_n^* = \frac{\sum_{i=1}^n (\frac{i}{n} - F_2(X_{(i)}))(F_1(X_{(i)}) - F_2(X_{(i)}))}{\sum_{i=1}^n (F_1(X_{(i)}) - F_2(X_{(i)}))^2} & \text{si } \epsilon_n^* \in ]0,1[ \\ \cdot 0 & \text{si } \epsilon_n^* \leq 0 \\ \cdot 1 & \text{si } \epsilon_n^* \geq 1 \end{cases}$$

c) L'estimateur de classification.

Cet estimateur est basé sur l'idée de grouper les  $n$  observations  $X_1, \dots, X_n$  en deux classes : l'une est constituée par les observations qui proviennent de la loi contaminée, l'autre est constituée par les observations qui proviennent de loi contaminante.

Soit  $A \in \mathcal{B}_{\mathbb{R}}$  tel que  $P_1(A) \neq P_2(A)$ .

L'estimateur de classement de  $\epsilon$  est défini par

$$\epsilon_{n,A}^* = \begin{cases} \cdot \epsilon_{n,A}^* = \frac{n^{-1} \sum_{j=1}^n 1_A(X_j) - P_2(A)}{P_1(A) - P_2(A)} & , \quad \text{si } \epsilon_{n,A}^* \in ]0,1[ \\ \cdot 0 & \text{si } \epsilon_{n,A}^* \leq 0 \\ \cdot 1 & \text{si } \epsilon_{n,A}^* \geq 1 \end{cases}$$

Posons  $C_0 = \{A : P_1(A) \leq \frac{1}{2} < P_2(A) \text{ ou } P_1(A) < \frac{1}{2} \leq P_2(A) \text{ ou}$

$P_2(A) \leq \frac{1}{2} < P_1(A) \text{ ou } P_2(A) < \frac{1}{2} \leq P_1(A)\}$

$C_1 = \{A : P_1(A) < P_2(A) \leq \frac{1}{2} \text{ ou } P_1(A) > P_2(A) \geq \frac{1}{2}\}$

$C_2 = \{A : P_2(A) < P_1(A) \leq \frac{1}{2} \text{ ou } P_2(A) > P_1(A) \geq \frac{1}{2}\}$

$C = C_0 \cup C_1 \subset C_2$ .

Proposition.-

$$\text{var}_{\epsilon} \hat{\epsilon}_{n,A} = \frac{1}{n} \frac{(\epsilon P_1(A) + (1-\epsilon) P_2(A)) (\epsilon \bar{P}_1(A) + (1-\epsilon) \bar{P}_2(A))}{(P_1(A) - P_2(A))^2} = \frac{\psi_A(\epsilon)}{n}$$

$$\sup_{0 < \epsilon < 1} \psi_A(\epsilon) = \begin{cases} \cdot \frac{P_2(A) \bar{P}_2(A)}{(P_1(A) - P_2(A))^2} & \text{si } A \in C_1 \\ \cdot \frac{P_1(A) \bar{P}_1(A)}{(P_1(A) - P_2(A))^2} & \text{si } A \in C_2 \\ \cdot \frac{1}{4(P_1(A) - P_2(A))^2} & \text{si } A \in C_0 \end{cases}$$

Remarques :

1) Trouver  $A \in \mathbb{B}_{\mathbb{R}^k}$  tel que  $P_1(A) \neq P_2(A)$  et  $\sup_{0 < \epsilon < 1} \psi_{\epsilon}(A)$  soit minimum est un problème très difficile.

2) Si on prend  $A = ]-\infty, \gamma]$  on retrouve l'estimateur de James.

3) Choisir  $A$  qui maximise  $|P_1(A) - P_2(A)|$  et peut être un choix raisonnable du fait que  $P_i(A) \bar{P}_i(A) \leq \frac{1}{4}$ ,  $i = 1, 2$ .

Comparaison :

. Modèle  $\{(1-\epsilon) N(m_1, 1) + \epsilon N(m_2, 1)\}$   $m_1 < m_2$

. méthode des fonctions orthogonales :

$$A_n = \sup_{0 < \epsilon < 1} \text{var} \hat{\epsilon}_n = \frac{1}{n} \left[ \frac{e^{2(m_0 - m_1)^2}}{(n_0 - m_1)^2} \right]_{-1}^1$$

. par l'estimateur de James :

$$B_n = \inf_{\lambda} \sup_{0 < \epsilon < 1} \text{var} \hat{\epsilon}_n = \frac{1}{n} \left[ \frac{\pi}{8 \int_0^{(m_0 + m_1)/2} e^{-x^2/2} dy} \right]$$

$$\frac{A_n}{B_n} = \frac{8 e^{2(m_0 - m_1)^2} \cdot \int_0^{(m_0 + m_1)/2} e^{-x^2/2} dy}{\pi(e^{(m_0 - m_1)^2} - 1)}$$

$$\frac{A_n}{B_n} = \frac{8 e^{2(m_0 - m_1)^2} \cdot \int_0^{(m_0 + m_1)/2} e^{-x^2/2} dy}{\pi(e^{(m_0 - m_1)^2} - 1)}$$

CHAPITRE IV

SIMULATION.

I - INTRODUCTION.

Soient  $(\Omega, \mathcal{A}, P)$  un espace de probabilité,  $\mathcal{B}_{\mathbb{R}}$  la tribu borélienne de  $\mathbb{R}$  et  $E$  un intervalle de  $\mathbb{R}$ .

$X : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  variable aléatoire de loi  $P_X$  absolument continu par rapport à la mesure de Lebesgue  $\mu$  et telle que  $f = \frac{dP_X}{d\mu} \in L^2(\mu)$ .  
Grâce à une visualisation graphique on se propose de comparer trois méthodes d'estimation de  $f$ .

1) Méthode du noyau.

Méthode introduite par M. Rosenblatt en 1956 [34]

$$\hat{f}_{n,1}(x) = \frac{1}{n \delta_n} \sum_{j=1}^n K\left(\frac{x - X_j}{\delta_n}\right) \text{ avec } K \text{ une application mesurable de } \mathbb{R} \text{ dans } \mathbb{R}$$

qui est en général une densité de probabilité et  $\delta_n$  une suite de nombres réels positifs tendant vers 0 et telle que  $n \delta_n \xrightarrow[n \rightarrow \infty]{} \infty$ .

Le choix optimal de  $K$  a été résolu par V.A. Epanechnikov (voir [36]). Cependant il est important de souligner que le choix de  $K$  ne présente pas une grande importance, alors que le choix de  $\delta_n$  représente une importance capitale voir à ce sujet par exemple [37].

De nombreux auteurs se sont intéressés au choix pratique de  $\delta_n$  voir par exemple W. Silverman [38].

2) Méthode des fonctions orthogonales.

$(e_i)_{i \in \mathbb{N}}$  étant une base de  $L^2(\mu)$ .

$$\hat{f}_{n,2}(x) = \sum_{i=0}^{q(n)} \hat{a}_{i,n} e_i(x) \text{ avec } \hat{a}_{i,n} = \frac{1}{n} \sum_{j=1}^n e_i(X_j) \text{ et } q(n) \rightarrow \infty.$$

La mise en oeuvre de cette méthode a été effectuée par R. Kronmal M. Tarter [39] et par M. Ouledcheikh [40]. Le procédé consiste pour un système  $(e_i)_{i \in \mathbb{N}}$  donné à choisir  $q(n)$  qui minimise l'erreur quadratique intégrée  $J_2(n, q(n))$ .

3) Méthode par projection sur les ensembles de mesures.

Soient :

-  $(\mu_i)_{i \in \mathbb{N}^*}$  une suite de mesures absolument continues par rapport à  $\mu$

et telle que  $\frac{d\mu_i}{d\mu} \in L^2(\mu)$ ,  $\forall i \in \mathbb{N}^*$  pour  $P_X \in P(E) \cap (\cup_{i=1}^q \mu_i)$

on a  $\hat{f}_{n,3}(x) = \frac{d}{d\mu} (\text{proj}_{e(\mu_1, \dots, \mu_q)} (\frac{1}{n} \sum_{j=1}^n \delta_{X_j}))$ , pour  $n$  assez grand

$$\hat{f}_{n,3}(x) = \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} (-1)^{k+i} \frac{\det M_{q(n),ki}}{\det M_{q(n)}} \left( \frac{1}{n} \sum_{j=1}^n \int K(X_j, y) d\mu_k(y) \right) \right] \frac{d\mu_i}{d\mu}(x)$$

la détermination du rang  $q(n)$  qui minimise l'erreur quadratique est dans ce cas très difficile aussi nous allons nous contenter d'essayer plusieurs rangs et de choisir le rang qui rapproche le mieux la courbe estimée de la courbe théorique.

4) On va prendre la moyenne des estimateurs  $(\hat{f}_{n,1} + \hat{f}_{n,2} + \hat{f}_{n,3})/3$ .

II - METHODE PAR PROJECTION SUR LES ENSEMBLES DE MESURES : LE CALCUL COMMUN  
A TOUTES LES LOIS QU'ON VA TRAITER PAR LA SUITE.-

Il est important de souligner que pour cette méthode les calculs sont extrêmement longs et délicats comme on va le constater par la suite.

Dans cette partie, on prend  $E = [a, b]$  compact de  $\mathbb{R}$ .

Comme système de mesure  $(\mu_i)_{i \in \mathbb{N}^*}$  :  $\mu_i = \frac{d\mu_i}{d\mu} = x^{i-1}$ ,  $i \in \mathbb{N}^*$ . On montre que les  $\mu_i$  sont linéairement indépendants.

$$f_{n,3}(x) = \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} (-1)^{k+i} \frac{\det M_{q(n),ki}}{\det M_{q(n)}} \left( \frac{1}{n} \sum_{j=1}^n \int K(X_j, y) y^{k-1} dy \right) \right] x^{i-1}$$

Calcul de l'erreur quadratique intégrée :

$$J_3(n, q(n)) = E \int_a^b (\hat{f}_{n,3}(x) - f(x))^2 dx = E \int_a^b (\hat{f}_{n,3}(x) - E\hat{f}_{n,3}(x))^2 dx + \int_a^b (E\hat{f}_{n,3}(x) - f(x))^2 dx$$

Posons

$$\begin{cases} I_3(n, q(n)) = E \int_a^b (\hat{f}_{n,3}(x) - f_{n,3}(x))^2 dx & \text{avec } f_{n,3}(x) = E\hat{f}_{n,3}(x) \\ L_3(n, q(n)) = \int_a^b (f_{n,3}(x) - f(x))^2 dx \end{cases}$$

Pour alléger l'écriture posons  $\hat{f}_{n,3} = \hat{f}_n$ ,  $f_{n,3} = f_n$  et

$$A_{q(n),ki} = (-1)^{i+k} \frac{\det M_{q(n),ki}}{\det M_{q(n)}}$$

Calculons  $I_3(n, q(n))$  :

$$I_3(n, q(n)) = E \int (\hat{f}_n - f_n)^2 dx = \int E(\hat{f}_n - f_n)^2 dx$$

$$I_3(n, q(n)) = \int E \left( \frac{1}{n} \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n),ki} \left( \sum_{j=1}^n \int K(X_j, y) y^{k-1} dy \right) \right] x^{i-1} \right)^2 dx$$

$$\sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n),ki} \int \int K(x, y) f(x) y^{k-1} dx dy \right] x^{i-1} \right)^2 dx$$

$$\begin{aligned}
 I_3(n, q(n)) &= \int E \left( \frac{1}{n} \sum_{j=1}^n \left[ \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n), ki} \left( \int K(X_j, y) y^{k-1} dy \right. \right. \right. \right. \right. \\
 &\quad \left. \left. \left. \left. - \int \int K(x, y) f(y) y^{k-1} dx dy \right) \right] x^{i-1} \right] \right)^2 dx \\
 &= \frac{1}{n} \int E \left( \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n), ki} \left( \int K(X_1, y) y^{k-1} dy \right) \right] x^{i-1} \right)^2 dx - \\
 &\quad \frac{1}{n} \int \left( \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n), ki} \left( \int \int K(x, y) f(x) y^{k-1} dx dy \right) \right] x^{i-1} \right)^2 dx
 \end{aligned}$$

Posons

$$F_{q(n)}(x) = E \left( \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n), ki} \left( \int K(X_1, y) y^{k-1} dy \right) \right] x^{i-1} \right)^2$$

$$G_{q(n)}(x) = \left( \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n), ki} \left( \int \int K(x, y) f(x) y^{k-1} dx dy \right) \right] x^{i-1} \right)^2$$

on a  $J(n, q(x)) = \frac{1}{n} \int (F_{q(n)}(x) - G_{q(n)}(x)) dx$

$$F_{q(n)}(x) = E \left( \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n), ki} \left( \int K(X_1, y) y^{k-1} dy \right) \right]^2 x^{2(i-1)} \right) +$$

$$\begin{aligned}
 2E \left( \sum_{1 \leq i < i' \leq q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n), ki} \left( \int K(X_1, y) y^{k-1} dy \right) \right] \right. \\
 \left. \left[ \sum_{k=1}^{q(n)} A_{q(n), ki'} \left( \int K(X_1, y) y^{k-1} dy \right) \right] x^{i+i'-2} \right)
 \end{aligned}$$

$$\begin{aligned}
 F_{q(n)}(x) &= \sum_{i=1}^{q(n)} x^{2(i-1)} E \left( \sum_{k=1}^{q(n)} (A_{q(n), ki})^2 \left( \int K(X_1, y) y^{k-1} dy \right)^2 \right) \\
 &+ 2 \sum_{1 \leq k < k' \leq q(n)} A_{ki} A_{k'i} \left( \int K(X_1, y) y^{k-1} dy \right) \left( \int K(X_1, y) y^{k'-1} dy \right) \\
 &+ 2 \sum_{1 \leq i < i' \leq q(n)} \left[ \sum_{\substack{1 \leq k \leq q(n) \\ 1 \leq k' \leq q(n)}} A_{ki} A_{k'i} E \left( \int K(X_1, y) y^{k-1} dy \right) \left( \int K(X_1, y) y^{k'-1} dy \right) \right] x^{i+i'-2}
 \end{aligned}$$



$$G_{q(n)}(x) = \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n),ki} \left( \iint K(x,y) f(x) y^{k-1} dx dy \right) \right]^2 x^{2(i-1)} +$$

$$2 \sum_{1 \leq i < i' \leq q(n)} \left[ \sum_{k=1}^{q(n)} A_{q(n),ki} \left( \iint K(x,y) f(x) y^{k-1} dx dy \right) \right] \left[ \sum_{k=1}^{q(n)} A_{q(n),k'i'} \left( \iint K(x,y) f(x) y^{k-1} dx dy \right) \right] x^{i+i'-2}$$

$$G_{q(n)}(x) = \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} (A_{q(n),ki})^2 \left( \iint K(x,y) f(x) y^{k-1} dx dy \right)^2 \right] x^{2(i-1)} +$$

$$\sum_{i=1}^{q(n)} \left[ 2 \sum_{1 \leq k < k' \leq q(n)} A_{q(n),ki} \cdot A_{q(n),k'i'} \left( \iint K(x,y) f(x) y^{k-1} dx dy \right) \left( \iint K(x,y) f(x) y^{k'-1} dx dy \right) \right] x^{2(i-1)} +$$

$$2 \sum_{1 \leq i < i' \leq q(n)} \left[ \sum_{\substack{1 \leq k \leq q(n) \\ 1 \leq k' \leq q(n)}} A_{q(n),ki} A_{q(n),k'i'} \left( \iint K(x,y) f(x) y^{k-1} dx dy \right) \left( \iint K(x,y) f(x) y^{k'-1} dx dy \right) \right] x^{i+i'-2}$$

$$n I_3(n, q(n)) = \int_a^b (F_{q(n)}(x) - G_{q(n)}(x)) dx$$

$$n I_3(n, q(n)) = \sum_{i=1}^{q(n)} \left[ \sum_{k=1}^{q(n)} (A_{q(n),ki})^2 \left( E \left( \int K(X_1, y) y^{k-1} dy \right)^2 - \left( \iint K(x,y) f(x) y^{k-1} dx dy \right)^2 \right) \frac{(b^{2i-1} - a^{2i-1})}{(2i-1)} \right] +$$

$$2 \sum_{i=1}^{q(n)} \left[ \sum_{1 \leq k < k' \leq q(n)} A_{q(n),ki} \cdot A_{q(n),k'i'} \left( E \left( \int K(X_1, y) y^{k-1} dy \right) \left( \int K(X_1, y) y^{k'-1} dy \right) - \left( \iint K(x,y) f(x) y^{k-1} dx dy \right) \left( \iint K(x,y) f(x) y^{k'-1} dx dy \right) \right) \right] \times \frac{b^{2i-1} - a^{2i-1}}{(2i-1)}$$

$$+ 2 \sum_{1 \leq i < i' \leq q(n)} \left[ \sum_{\substack{1 \leq k \leq q(n) \\ 1 \leq k' \leq q(n)}} A_{q(n),ki} \cdot A_{q(n),k'i'} \left( E \left( \int K(X_1, y) y^{k-1} dy \right) \left( \int K(X_1, y) y^{k'-1} dy \right) - \left( \iint K(x,y) f(x) y^{k-1} dx dy \right) \left( \iint K(x,y) f(x) y^{k'-1} dx dy \right) \right) \right] \frac{b^{i+i'-1} - a^{i+i'-1}}{(i+i'-1)}$$

Considérons le noyau reproduisant

$$K(x,y) = \frac{2c^2}{2c^2-xy} = \sum_{i=0}^{\infty} \frac{1}{2^i} \left(\frac{x}{c}\right)^i \left(\frac{y}{c}\right)^i$$

avec  $c = \sup(|a|, |b|)$ .

Pour l'écriture posons :

- $h_k(x) = \int_a^b y^{k-1} \text{Log}(2c^2-xy) dy \quad k \geq 1$
- $g_k(x) = \int_a^b \frac{2c^2}{2c^2-xy} y^{k-1} dy$
- $\alpha(i,k) = \int_a^b \int_a^b \frac{2c^2}{2c^2-xy} x^{i-1} y^{k-1} dx dy = \int_a^b g_k(x) x^{k-1} dx$
- $\beta(k) = \int_a^b \int_a^b \text{Log}(2c^2-xy) y^{k-1} x^{k-1} dx dy = \int_a^b h_k(x) x^{k-1} dx$
- $U(k) = \int_a^b \left( \int_a^b \frac{2c^2}{2c^2-xy} y^{k-1} dy \right)^2 f(x) dx$
- $V(k) = \int_a^b \int_a^b \frac{2c^2}{2c^2-xy} y^{k-1} f(x) dy dx$
- $W(i,k) = \int_a^b \left( \int_a^b \frac{2c^2}{2c^2-xy} y^{i-1} dy \right) \left( \int_a^b \frac{2c^2}{2c^2-xy} y^{k-1} dy \right) f(x) dx$
- $T(i,k) = W(i,k) - V(i) V(k)$
- $s(k) = u(k) - V^2(k)$

Calcul de  $h(x)$  :

pour  $x \neq 0$  :

- $h_1(x) = \int_a^b \text{Log}(2c^2-xy) dy = - \frac{(2c^2-xy)\text{Log}(2c^2-xy)}{x} - y \Big|_a^b$
- $h_1(x) = \frac{(2c^2-ax)\text{Log}(2c^2-ax) - (2c^2-ax)\text{Log}(2c^2-bx)}{x} + (a-b)$
- $h_k(x) = \int_a^b y^{k-1} \text{Log}(2c^2-xy) dy$   
 $k > 1$

$$h_k(x) = y^{k-1} \left( -\frac{(2c^2-xy)}{x} \text{Log}(2c^2-xy) - y \right) \Big|_a^b + (k-1) \int_a^b y^{k-2} \left( \frac{(2c^2-xy)}{x} \text{Log}(2c^2-xy) + y \right) dy$$

$k > 1$

$$h_k(x) = a^{k-1} \left( \frac{(2c^2-ax)}{x} \text{Log}(2c^2-ax) + a \right) - b^{k-1} \left( \frac{(2c^2-bx)}{x} \text{Log}(2c^2-bx) + b \right) +$$

$$(k-1) \left[ \int_a^b \frac{y^{k-2}}{x} (2c^2-xy) \text{Log}(2c^2-xy) dx + \int_a^b y^{k-1} dy \right]$$

$k > 1$

Après tout calcul on trouve :

$$k_k(x) = \frac{1}{kx} \left[ a^{k-1} (2c^2-ax) \text{Log}(2c^2-ax) - b^{k-1} (2c^2-bx) \text{Log}(2c^2-bx) \right] -$$

$$\frac{(b^k - a^k)}{k^2} + 2c^2 \frac{(k-1)}{kx} h_{k-1}(x)$$

$k > 1$

Pour  $x = 0$  :

- $h_1(0) = (b-a) \text{Log}(2c^2)$
- $h_k(0) = \frac{(b^k - a^k)}{k} \text{Log} 2c^2$

$k > 1$

Calcul de  $g_k(x)$  :

Pour  $x \neq 0$  :

- $g_1(x) = 2c^2 \int_a^b \frac{1}{2c^2-xy} dy = \frac{2c^2}{x} [\text{Log}(2c^2-ax) - \text{Log}(2c^2-bx)]$
- $g_k(x) = 2c^2 \int_a^b \frac{y^{k-1}}{2c^2-xy} dy$

$$= 2c^2 \left[ -y^{k-1} \frac{\text{Log}(2c^2-xy)}{x} \Big|_a^b + \frac{k-1}{x} \int_a^b y^{k-2} \text{Log}(2c^2-xy) dy \right]$$

$$g_k(x) = \frac{2c^2}{x} \left[ a^{k-1} \text{Log}(2c^2-ax) - b^{k-1} \text{Log}(2c^2-bx) \right] + \frac{2c^2}{x} (k-1) h_{k-1}(x)$$

$k > 1$

Pour  $x = 0$  :

$$g_k(0) = \frac{b^k - a^k}{k} \quad \forall k \in \mathbb{N}^*$$

III - LOI POLYNOMIALE :  $x + \frac{3}{2} x^2$  sur  $[0,1]$ .-

$$\text{On considère } \frac{dP_X}{dx}(x) = \begin{cases} x + \frac{3}{2} x^2 & \text{pour } x \in [0,1] \\ 0 & \text{ailleurs .} \end{cases}$$

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de loi  $P_X$  pour simuler cet échantillon on utilise le lemme suivant :

Lemme.-

Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F$  alors  $Y = F(X)$  suit une loi de probabilité uniforme sur  $[0,1]$ .

La fonction de répartition  $F$  de  $P_X$  est  $F(x) = \frac{x^2 + x^3}{2}$  strictement croissante sur  $[0,1]$ , pour calculer  $X = F^{-1}(Y)$  on utilise la méthode de Newton.

a) Méthode des fonctions orthogonales.

On utilise les polynômes de Legendre :

$$P_n(x) = \frac{1}{2^n n!} D^n [(x^2-1)^n], \quad n = 0, 1, \dots \quad x \in [-1, +1]$$

dans le chapitre II on a vu que le système  $\{((2n+1)/2)^{1/2} P_n\}$  est orthonormale et complet dans  $L^2 [-1, +1]$  une simple transformation nous permet de dire que  $\{Q_n(x) = \sqrt{2n+1} Y_n(x)\}$  est complet et orthonormal dans  $L^2 [0,1]$  avec  $Y_n(x) = \frac{1}{n!} D^n [x^n(x-1)^n]$ .

On a :

$$\begin{aligned} Q_0(x) &= 1 \\ Q_1(x) &= \sqrt{3} (2x-1) \\ Q_2(x) &= \sqrt{5} (6x^2-6x+1) \\ &\vdots \\ &\dots\dots \end{aligned}$$

$$f(x) = a_0 Q_0(x) + a_1 Q_1(x) + a_2 Q_2(x) \quad \text{avec} \quad a_i = \int_0^1 f(x) Q_i(x) dx$$

$\hat{f}_{n,2}(x) = \sum_{i=0}^2 \left[ \frac{1}{n} \sum_{j=1}^n Q_i(X_j) \right] Q_i(x)$  est un estimateur sans biais de  $f$  et  $r^*(n)$

le rang qui minimise  $I_2(n, r(n)) = E \int_0^1 (\hat{f}_{n,2} - f)^2(x) dx$  est  $\leq 2$ .

On a :

$$\bullet I_2(n, 2) = E \int_0^1 (\hat{f}_{n,2}(x) - f(x))^2 dx = \frac{1}{n} \sum_{i=0}^2 \text{var } Q_i(X)$$

$$\bullet I_2(n, 1) = \frac{1}{n} \sum_{i=0}^1 \text{var } Q_i(X) + a_2^2$$

$$\begin{aligned} I_2(n, 2) - I_2(n, 1) &= \frac{1}{2} \text{var } Q_2(X) - a_2^2 \\ &= \frac{1}{n} [E_f(Q_2(X))^2 - (E_f(Q_2(X)))^2] - a_2^2 \end{aligned}$$

Or  $a_2 = \int_0^1 \sqrt{5(6x^2 - 6x + 1)} \left(x + \frac{3}{2}x^2\right) dx = \frac{\sqrt{5}}{20}$

$$E_f(Q_2(X))^2 = \int_0^1 5(6x^2 - 6x + 1)^2 \left(x + \frac{3}{2}x^2\right) dx = \frac{15}{14}$$

$$I_2(n, 2) - I_2(n, 1) = \frac{1}{n} \left[ \frac{15}{14} - (n+1) \frac{1}{80} \right] \leq 0 \quad \text{pour } n \geq 84$$

$$I_2(n, 1) - I_2(n, 0) = \frac{1}{n} \sum_{i=0}^1 \text{var } Q_i(X) + a_2^2 - \left( \frac{1}{n} \text{var } Q_0(X) + a_1^2 + a_2^2 \right)$$

$$I_2(n, 1) - I_2(n, 0) = \frac{1}{n} [E_f(Q_1(X))^2 - (n+1) a_1^2]$$

Or  $a_1 = \int_0^1 Q_1(X) \left(x + \frac{3}{2}x^2\right) dx = \frac{5\sqrt{3}}{12}$  et  $E_f(Q_1(X))^2 = 3 \int_0^1 (2x-1)^2 \left(x + \frac{3}{2}x^2\right) dx = \frac{11}{10}$

donc  $I_2(n, 1) - I_2(n, 0) = \frac{1}{n} \left[ \frac{11}{10} - (n+1) \frac{25}{48} \right] \leq 0 \quad \text{pour } n \geq 2$

donc pour

$$\begin{aligned} n \geq 84 \quad & I_2(n, 2) \leq I_2(n, 1) \leq I_2(n, 0) \\ 2 \leq n \leq 84 \quad & I_2(n, 1) \leq I_2(n, 0) \\ & \leq I_2(n, 2) \end{aligned}$$

Pour  $n \geq 84$  :

$$\hat{f}_{n,2}(x) = 1 + 3\left(\frac{1}{n} \sum_{j=1}^n (2X_j - 1)(2x - 1)\right) + 5\left(\frac{1}{n} \sum_{j=1}^n (6X_j^2 - 6X_j + 1)\right)(6x^2 - 6x + 1)$$

$$\hat{f}_{n,2}(x) = 9 + 30x^2 - 36x + \left(\frac{1}{n} \sum_{j=1}^n X_j\right) (-180x^2 + 192x - 36) + 30\left(\frac{1}{n} \sum_{j=1}^n X_j^2\right) (6x^2 - 6x + 1)$$

Pour  $n < 84$  :

$$\hat{f}_{n,2}(x) = 1 + 3\left(\frac{1}{n} \sum_{j=1}^n (2X_j - 1)(2x - 1)\right)$$

$$\hat{f}_{n,2}(x) = 4 + 6\left(\frac{1}{n} \sum_{j=1}^n X_j\right)(2x - 1) - 6x$$

Après tout calcul on trouve :

$$\begin{aligned} \cdot \quad I_2(n,2) &= \frac{1}{n} \cdot \frac{52}{35} \\ \cdot \quad \hat{A}_2(n,2) &= \int_0^1 \hat{f}_{n,2}(x) dx = 1 \\ \cdot \quad I_2(n,1) &= \frac{1}{n} \cdot \frac{139}{240} + \frac{1}{80} \quad ; \quad \hat{A}_2(n,1) = \int_0^1 \hat{f}_{n,2}(x) dx = 1 \end{aligned}$$

b) Méthode par projection sur les ensembles de mesures.

$$P_X \in e(\mu_1, \mu_2, \mu_3) \quad (P_X : \frac{dP_X}{dx}(x) = x + \frac{3}{2}x^2 \text{ sur } [0,1])$$

$$K(x,y) = \frac{2}{2-xy}$$

$$\hat{f}_{n,3}(x) = \frac{1}{n} \sum_{i=1}^3 \left[ \sum_{k=1}^3 (-1)^{i+k} \frac{\det M_{3,ki}}{\det M_3} \left( \sum_{j=1}^n \int \frac{2}{2-X_j y} y^{k-1} dy \right) \right] x^{i-1}$$

est un estimateur sans biais de  $f(x) = x + \frac{3}{2}x^2$ . On a :

Pour  $x \neq 0$  :

$$h_1(x) = \frac{2 \operatorname{Log} 2 - (2-x) \operatorname{Log}(2-x)}{x} - 1 = \frac{2}{x} \operatorname{Log} \frac{2}{2-x} + \operatorname{Log}(2-x) - 1$$

$$h_k(x) = \frac{1}{kx} [(x-2) \operatorname{Log}(2-x)] - \frac{1}{k^2} + \frac{2(k-1)}{xk} h_{k-1}(x)$$

$k > 1$

$$g_1(x) = \frac{2}{x} \operatorname{Log} \frac{2}{2-x}$$

$$g_k(x) = -\frac{2}{x} \operatorname{Log}(2-x) + \frac{2}{x} (k-1) h_{k-1}(x)$$

$k > 1$

Pour  $x = 0$  :

$$h_k(0) = \frac{\operatorname{Log} 2}{k} \quad \forall k \in \mathbb{N}^*$$

$$g_k(0) = \frac{1}{k} \quad \forall k \in \mathbb{N}^*$$

Calcul des éléments  $\alpha(i, k)$  de  $M_3$ .

$$\alpha(i, k) = \int_0^1 g_k(x) x^{i-1} dx = \int_0^1 \int_0^1 \frac{2 x^{i-1} y^{k-1}}{2 - xy} dx dy$$

$$\alpha(1, 1) = \int_0^1 \int_0^1 \frac{2}{2-xy} dx dy = \int_0^1 g_1(x) dx$$

$$\alpha(1, i) = \lim_{\substack{i > 1 \\ \varepsilon > 0}} \int_{\varepsilon}^1 g_1(x) x^{i-1} dx = \lim_{\varepsilon \rightarrow 0} 2 \left( \int_{\varepsilon}^1 x^{i-2} \operatorname{Log} 2 dx - \int_{\varepsilon}^1 x^{i-2} \operatorname{Log}(2-x) dx \right)$$

$$\alpha(1, i) = 2(h_{i-1}(0) - h_{i-1}(1))$$

$i > 1$

$$\alpha(2,2) = \int_0^1 g_2(x) x dx = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^1 g_2(x) x dx$$

Après tout calcul on trouve

$$\alpha(2,2) = 2 (\alpha(1,1) - 1)$$

$$\beta(1) = \int_0^1 h_1(x) dx = \alpha(1,1) + h_1(1) - 1$$

$$\beta(1) = \alpha(1,1) + 2(\text{Log } 2 - 2)$$

$$\beta(k) = \int_0^1 h_k(x) x^{k-1} dx = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^1 h_k(x) x^{k-1} dx$$

Après deux pages de calcul, on trouve la relation de récurrence :

$$\beta(k) = \frac{2}{k} \left[ (k-1) \beta(k-1) - \frac{h_{k-1}(1)}{k} - \frac{1}{k^2} \right]$$

$$\alpha(k,k) = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^1 g_k(x) x^{k-1} dx$$

On trouve à la fin :

$$\alpha(k,k) = 2 \left[ (k-1) \beta(k-1) h_{k-1}(1) \right]$$

$$\alpha(i,k) = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^1 g_k(x) x^{i-1} dx$$

$i > 1$   
 $i \neq k$   
 $k > 1$

on trouve :

$$\alpha(i,k) = \frac{2}{(i-k)} \left[ (k-1) h_{k-1}(1) - (i-1) h_{i-1}(1) \right]$$

$i \neq k$   
 $i, k > 1$



$$\hat{A}_3(n,3) = \int_0^1 \hat{f}_{n,3}(x) dx = \frac{1}{n} \sum_{i=1}^3 \left[ \sum_{k=1}^3 (-1)^{k+i} \frac{\det M_{3,ki}}{\det M_3} \left( \sum_{j=1}^n g_k(x_j) \right) \right] \frac{1}{i}$$

Calcul de l'erreur quadratique intégrée :

$\hat{f}_{n,3}$  est un estimateur sans biais donc l'erreur quadratique intégrée est :  $I_3(n,3) = E \int_0^1 (\hat{f}_{n,3}(x) - f_{n,3}(x))^2 dx = J_3(n,3)$  .

Calculons  $W(i,k)$  :

$$W(i,k) = \int_0^1 \left( \int_0^1 \frac{2}{2-xy} y^{i-1} dy \right) \left( \int_0^1 \frac{2}{2-xy} y^{k-1} dy \right) \left( x + \frac{3}{2} x^2 \right) dx$$

$$W(i,k) = \int_0^1 \left( \int_0^1 \frac{2}{2-xy} y^{i-1} dy \right) \left( \int_0^1 \frac{2}{2-xy} y^{k-1} dy \right) x dx + \frac{3}{2} \int_0^1 \left( \int_0^1 \frac{2}{2-xy} y^{i-1} dy \right) \left( \int_0^1 \frac{2}{2-xy} y^{k-1} dy \right) x^2 dx$$

$$W(i,k) = \int_0^1 g_i(x) g_k(x) x dx + \frac{3}{2} \int_0^1 g_i(x) g_k(x) x^2 dx$$

$$W(1,1) = \int_0^1 g_1^2(x) x dx + \frac{3}{2} \int_0^1 g_1^2(x) x^2 dx$$

$$\begin{aligned} W(1,k)_{k>1} &= \int_0^1 g_1(x) g_k(x) x dx + \frac{3}{2} \int_0^1 g_1(x) g_k(x) x^2 dx \\ &= \int_0^1 g_1(x) \left[ -\frac{2}{x} \text{Log}(2-x) + \frac{2}{x} (k-1) h_{k-1}(x) \right] x dx \\ &+ \frac{3}{2} \int_0^1 g_1(x) \left[ -\frac{2}{x} \text{Log}(2-x) + \frac{2}{x} (k-1) h_{k-1}(x) \right] x^2 dx \end{aligned}$$

$$\begin{aligned} W(1,k) &= \int_0^1 g_1(x) \left[ -2 \text{Log}(2-x) + 2(k-1) h_{k-1}(x) \right] dx \\ &+ \frac{3}{2} \int_0^1 g_1(x) \left[ -2 \text{Log}(2-x) + 2(k-1) h_{k-1}(x) \right] x dx \end{aligned}$$

$$W(1,k) = \int_0^1 g_1(x) [-2 \operatorname{Log}(2-x)] dx + 2(k-1) \int_0^1 g_1(x) h_{k-1}(x) dx$$

$$+ \frac{3}{2} \int_0^1 g_1(x) [-2 \operatorname{Log}(2-x)] x dx + 3(k-1) \int_0^1 g_1(x) h_{k-1}(x) dx$$

c) Méthode du noyau.

Soit  $g$  un noyau symétrique, tel que  $\int g(x) dx = 1$  et  $\int x^2 g(x) dx < \infty$

$\hat{f}_{n,1}(x) = \frac{1}{n \delta_n} \sum_{j=1}^n g\left(\frac{X_j - x}{\delta_n}\right)$  avec  $\delta_n \xrightarrow[n \rightarrow \infty]{} 0$ ,  $n \delta_n \xrightarrow[n \rightarrow \infty]{} 0$ , l'erreur quadratique intégrée est équivalente pour  $n$  assez grand à

$$I_1(n, \delta_n) \approx \frac{1}{n \delta_n} \int g^2(t) dt + \frac{\delta_n^4}{4} \left( \int (f''(x))^2 dx \right) \left( \int t^2 f(t) dt \right)^2$$

$\hat{\delta}_n$  qui minimise cette erreur  $I_1(n, \delta_n)$  est :

$$\hat{\delta}_n = \left( \frac{\int g^2(t) dt}{\left( \int t^2 g(t) dt \right)^2} \right)^{1/5} \frac{1}{\left( \int (f''(x))^2 dx \right)^{1/5} n^{1/5}}$$

$$I_1(n, \hat{\delta}_n) \approx \frac{5}{4} \left( \frac{\left( \int g^2(t) dt \right)^4 \cdot \left( \int t^2 g(t) dt \right) \cdot \int (f''(x))^2 dx}{n^4} \right)^{1/5}$$

pour  $g(x) = \frac{1}{2} 1_{[-1, +1]}(x)$  on trouve  $\hat{\delta}_n = \frac{1}{(2n)^{1/5}}$  et  $I_1(n, \hat{\delta}_n) = \frac{5}{4} (2n)^{-4/5}$

$g(x) = 1_{[-1/2, +1/2]}(x)$  on trouve  $\hat{\delta}_n = \left(\frac{16}{n}\right)^{1/5}$ .

IV - LOI EXPONENTIELLE NEGATIVE  $f(x) = \begin{cases} \cdot e^{-x} & \text{pour } x \in \mathbb{R}^+ \\ \cdot 0 & \text{sinon} \end{cases}$

a) Méthode du noyau.

On considère comme noyau le noyau  $g(x) = 1_{[-1/2, 1/2]}(x)$  et comme fenêtre mobile  $\delta_n = (\frac{288}{n})^{1/5}$  qui est optimale d'après P. Hominal et P. Deheuvels (voir [35] page 57).

$$f_{n,1}(x) = \frac{1}{n \delta_n} \sum_{j=1}^n g\left(\frac{X_j - x}{n}\right) .$$

b) Méthode des fonctions orthogonales : les fonctions de Laguerre.

Les fonctions de Laguerre  $(\ell_j)_{j \geq 0}$  qui constituent un système orthonormal et complet de  $L^2(\mathbb{R}_+)$  satisfait aux équations :

$$x D^2 \ell_n + D \ell_n + (n + \frac{1}{2} - \frac{x}{4}) \ell_n = 0 \quad n \in \mathbb{N} .$$

Leur expression en termes de polynômes de Laguerre  $L_n$  est la plus courante  $\ell_n(x) = e^{-x/2} L_n(x)$  avec  $L_n(x) = \frac{e^x}{n!} D^n(x^n e^{-x})$ . On construit les polynômes de Laguerre  $L_n$  à partir de la relation de récurrence :

$$(n+1) L_{n+1} - (2n + 1 - x) L_n + n L_{n-1} = 0 \quad \text{avec :}$$

$$L_0(x) = 1$$

$$L_1(x) = 1 - x$$

$$L_2(x) = \frac{1}{2} (3-x) (1-x) - 1 = \frac{x^2}{2} - 2x + \frac{1}{2}$$

donc on a :

$$\ell_0(x) = e^{-x/2}$$

$$\ell_1(x) = e^{-x/2} (1-x)$$

$$\ell_2(x) = e^{-x/2} \left( \frac{x^2}{2} - 2x + \frac{1}{2} \right) = \frac{1}{2} e^{-x/2} (x^2 - 4x + 1)$$

$$\hat{f}_{n,2}(x) = \sum_{i=0}^{q(n)} \left( \frac{1}{n} \sum_{j=1}^n \ell_i(X_j) \right) \ell_i(x) .$$

Dans [40] on trouve pour ce système le choix optimal de  $q(n)$  pour différentes tailles de l'échantillon simulé.

Pour  $n = 50$  et  $100$  on a  $q_0(n) = 1$

$$\hat{f}_{n,2}(x) = \frac{1}{n} \sum_{i=0}^1 \left( \sum_{j=1}^n \ell_i(X_j) \ell_i(x) \right)$$

$$\hat{f}_{n,2}(x) = \frac{1}{n} \left[ \sum_{j=1}^n e^{-X_j/2} (1 + (1-X_j)(1-x)) \right] e^{-x/2}$$

c) Méthode par projection sur les ensembles de mesures.

On prend  $E = [0, b]$  compact de  $\mathbb{R}$ .

Estimer la loi exponentielle négative sur  $[0, b]$  c'est commettre une erreur de  $e^{-b}$  donc pour l'erreur soit faible il suffit de prendre  $b$  assez grand.

Comme noyau reproduisant on prend  $K(x, y) = \frac{2b^2}{2b^2 - xy}$ .

On a alors :

Pour  $x \neq 0$  :

$$h_1(x) = \frac{2b^2 \text{Log } 2b^2 - (2b^2 - bx) \text{Log}(2b^2 - bx)}{x} - b$$

$$h_k(x) = -\frac{b^k}{k} \left[ \frac{(2b-x) \text{Log}(2b^2 - bx)}{x} + \frac{1}{k} \right] + 2b^2 \frac{(k-1)}{k \cdot x} h_{k-1}(x)$$

$k > 1$

$$g_1(x) = \frac{2b^2}{x} \text{Log } \frac{2b}{2b-x}$$

$$g_k(x) = -\frac{2b^{k+1}}{x} \text{Log } (2b^2 - bx) + \frac{2b^2}{x} (k-1) h_{k-1}(x)$$

$k > 1$

Pour  $x = 0$  :

$$h_k(x) = \frac{b^k}{k} \text{Log } 2b^2 \quad \forall k \in \mathbb{N}^*$$

$$g_k(x) = \frac{b^k}{k} \quad \forall k \in \mathbb{N}^*$$

Calcul des éléments de la matrice  $M_{q(n)}$  :  $\alpha(i,k) = \int_0^b \int_0^b \frac{2b^2}{2b^2 - xy} x^{i-1} y^{k-1} dx dy$

$$\alpha(1,1) = \int_0^b \int_0^b \frac{2b^2}{2b^2 - xy} dx dy = \int_0^b g_1(x) dx$$

$$\alpha(1,i) = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^b g_1(x) x^{i-1} dx \quad . \quad \text{On trouve à la fin}$$

$$\alpha(1,i) = 2b^2 [h_{i-1}(0) - h_{i-1}(b)] = \alpha(i,1) \quad \text{pour } i > 1$$

$$\alpha(2,2) = \int_0^b g_2(x) x dx = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^b g_2(x) x dx \quad ,$$

on trouve :

$$\alpha(2,2) = 2b^2(\alpha(1,1) - b^2)$$

$$\beta(1) = \int_0^b h_1(x) dx = \alpha(1,1) + b h_1(b) - b^2$$

$$\beta(k) = \int_0^b x^{k-1} h_k(x) dx = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^b x^{k-1} h_k(x) dx$$

$$\alpha(k,k) = \lim_{\substack{k>2 \\ \epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^b g_k(x) x^{k-1} dx$$

$$\alpha(k,k) = - 2b^{k+1} h_{k-1}(b) + 2b^2(k-1) \beta(k-1) \quad k>2$$

$$\alpha(i,k) = \lim_{\substack{i>1 \\ i>k \\ i \neq k \\ \epsilon \rightarrow 0 \\ \epsilon > 0}} \int_{\epsilon}^b g_k(x) x^{i-1} dx \quad ,$$

$$\alpha(i,k) = \frac{2b^2}{i-k} [(k-1) b^{i-1} h_{k-1}(b) - (i-1) b^{k-1} h_{i-1}(b)] \quad \begin{matrix} i>1 \\ k>1 \\ i \neq k \end{matrix}$$

Remarque : On peut utiliser le développement de  $K(x,y)$  en série infinie pour le calcul des éléments  $\alpha(i,k)$  en effet ;

$$K(x,y) = \frac{2b^2}{2b^2 - xy} = \sum_{i=0}^{\infty} \frac{1}{2^i} \left(\frac{x}{b}\right)^i \left(\frac{y}{b}\right)^i$$

$$\alpha(i,k) = \sum_{h=0}^{\infty} \frac{1}{2^h} \left( \int_0^b \left(\frac{x}{b}\right)^h x^{i-1} dx \right) \left( \int_0^b \left(\frac{y}{b}\right)^h y^{k-1} dy \right)$$

$$\alpha(i,k) = b^{i+k} \sum_{h=0}^{\infty} \frac{1}{2^h (h+i) (h+k)}$$

$$g_k(x) = b^k \sum_{h=0}^{\infty} \left(\frac{x}{2b}\right)^h \frac{1}{(h+k)}$$

mais pour les calculs numériques, on est obligé de tronquer ces séries en fixant un seuil d'erreur.

V - LOI NORMALE  $N(0,1)$ .

a) Méthode du noyau :

$$f_{n,1}(x) = \frac{1}{n \delta_n} \sum_{j=1}^n K\left(\frac{X_j - x}{\delta_n}\right) \quad \text{avec} \quad K(x) = 1_{[-1/2, 1/2]}(x)$$

- pour  $n = 100$  on prend  $\delta_n = 1,510$  qui est optimal d'après P. Deheuvels ([37], page 39).

- pour  $n \neq 100$ , on prend  $\delta_n = \frac{(3,685)S}{n^{1/5}}$  avec  $S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$   
 et  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  (voir [41], p. 29).

b) Méthode des fonctions orthogonales : les fonctions d'Hermite.

Les fonctions d'Hermite  $h_j$ ,  $j = 0, 1, \dots$  sont définies par :

$$h_j(x) = \frac{e^{-x^2/2}}{(2^j j! \sqrt{\pi})^{1/2}} H_j(x) \quad \text{et où} \quad H_j(x) = e^{x^2} D^j(e^{-x^2}), \quad j \in \mathbb{N}.$$

Le nombre  $q(n)$  optimal pour ce système est  $m_0 = 0$  d'après M. Ouledcheikh ([40]).

$$\hat{f}_{n,2}(x) = \left(\frac{1}{n}\right) \sum_{j=1}^n h_0(X_j) h_0(x)$$

$$\hat{f}_{n,2}(x) = \frac{1}{\sqrt{\pi}} \left(\frac{1}{n}\right) \sum_{j=1}^n e^{-(X_j^2 + x^2)/2}$$



c) Méthode par projection sur les ensembles de mesures.

On prend  $E = [-b, +b]$ .

Pour  $x \neq 0$  :

$$h_1(x) = b \left[ \frac{2b+x}{2b-x} \text{Log} \frac{2b+x}{2b-x} + \text{Log} b^2(4b^2 - x^2) - 2 \right]$$

$$h_k(x) = \frac{1}{kx} \left[ (-b)^{k-1} (2b^2 + bx) \text{Log}(2b^2 + bx) - b^{k-1} (2b^2 - bx) \text{Log}(2b^2 - bx) \right] - \frac{b^k - (-b)^k}{k^2} + 2 b^2 \frac{(k-1)}{kx} \cdot h_{k-1}(x)$$

$$h_{2p}(x) = -b^{2p+1} \frac{\text{Log } b^2(4b^2-x^2)}{px} + \frac{b^{2p}}{2p} \text{Log } \frac{2b-x}{2b+x} + \frac{b^2(2p-1)}{px} h_{2p-1}(x)$$

$$h_{2p-1}(x) = \frac{b^{2p-1}}{(2p-1)} \left[ \frac{2b}{x} \text{Log } \frac{2b+x}{2b-x} + \text{Log } b^2(4b^2-x^2) - \frac{2}{(2p-1)} \right]$$

$$+ 4b^2 \frac{(p-1)}{x \cdot (2p-1)} h_{2(p-1)}(x)$$

$$g_1(x) = \frac{2b^2}{x} \text{Log } \frac{2b+x}{2b-x}$$

$$g_k(x) = \frac{2b^2}{x} \left[ (-b)^{k-1} \text{Log}(2b^2+bx) - b^{k-1} \text{Log}(2b^2-bx) \right] + \frac{2b^2}{x} (k-1) h_{k-1}(x)$$

$k > 1$

$$g_{2p}(x) = -\frac{2b^{2p+1}}{x} \text{Log } b^2(4b^2-x^2) + \frac{2b^2}{x} (2p-1) h_{2p-1}(x)$$

$$g_{2p-1}(x) = \frac{2b^{2p}}{x} \text{Log } \frac{2b+x}{2b-x} + \frac{4b^2}{x} (p-1) h_{2(p-1)}(x)$$

Pour  $x = 0$  :

$$h_{2p}(0) = 0$$

$$h_{2p-1}(0) = \frac{2b^{2p-1}}{2p-1} \text{Log } 2b^2$$

$$g_{2p}(0) = 0$$

$$g_{2p-1}(0) = \frac{2b^{2p-1}}{2p-1}$$



Calcul des éléments de la matrice  $M_{q(n)} = (\alpha(i,k))$

$$\alpha(1,1) = 2b^2 \int_{-b}^b \frac{1}{x} \text{Log} \frac{2b+x}{2b-x} dx$$

$$\alpha(2,2) = 2 b^2 \alpha(1,1) - 8 b^4$$

$$\beta(1) = \alpha(1,1) + b h_1(b) + b h_1(-b) - 4b^2$$

$$\beta(1) = \alpha(1,1) + 2b h_1(b) - 4b^2 \quad (h_1 \text{ est paire})$$

$$\beta(k) = \frac{1}{k} \left[ 2b^2 ((-b)^{k-1} h_{k-1}(-b) - b^{k-1} h_{k-1}(b)) + b^k h_k(b) - (-b)^k h_k(-b) - \left( \frac{b^k - (-b)^k}{k} \right)^2 \right]$$

$$+ 2 b^2 \left( \frac{k-1}{k} \right) \beta(k-1) .$$

Remarque : On montre que  $h_{2p}$  est impaire et  $h_{2p+1}$  est paire

$$\beta(2p) = \frac{b^{2p}}{p} [h_{2p}(b) - 2b h_{2p-1}(b)] + 2b^2 \frac{(2p-1)}{2p} \beta(2p-1)$$

$$\beta(2p-1) = \frac{2b^{2p-1}}{2p-1} [h_{2p-1}(b) - 2b h_{2(p-1)}(b) - \frac{2b^{2p-1}}{(2p-1)^2}] + 2b^2 \frac{2(p-1)}{2p-1} \beta(2(p-1))$$

$$\alpha(k,k) = 2b^2 [(-b)^{k-1} h_{k-1}(-b) - b^{k-1} h_{k-1}(b)] + 2b^2 (k-1) \beta(k-1)$$

$k > 2$

$$\alpha(2p, 2p) = -4 b^{2p+1} h_{2p-1}(b) + 2b^2 (2p-1) \beta(2p-1)$$

$$\alpha(2p-1, 2p-1) = -4 b^{2p} h_{2(p-1)}(b) + 4 b^2 (p-1) \beta(2(p-1))$$

$$\alpha(i,k) = 2b^2 \frac{2(i-1)}{i-k} [(-b)^{k-1} h_{i-1}(-b) - b^{k-1} h_{i-1}(b)] + 2b^2 \frac{(k-1)}{i-k} (b^{i-1} h_{k-1}(b) - (-b)^{i-1} h_{k-1}(-b))$$

$i > 1$   
 $k > 1$   
 $i \neq k$

pour  $i = 2p, k = 2r \quad r \neq p$  on trouve :

$$\alpha(2p, 2r) = \frac{2}{p-r} [(2r-1) b^{2p+1} h_{2r-1}(b) - (2p-1) b^{2r+1} h_{2p-1}(b)]$$

Pour  $i = 2p+1$ ,  $k = 2r+1$   $r \neq p$

$$\alpha(2p+1, 2r+1) = \frac{4}{p-r} [r b^{2(p+1)} h_{2r}(b) - p b^{2(r+1)} h_{2p}(b)]$$

Pour  $i = 2p+1$ ,  $k = 2r$ , on trouve

$$\alpha(2p+1, 2r) = 0$$

$$\alpha(1, i) = 2 b^2 [h_{i-1}(b) - h_{i-1}(b)]$$

$i > 1$

Pour  $i = 2p$  : on a

$$\alpha(1, 2p) = 0$$

( $h_{2p-1}$  est paire) .

Pour  $i = 2p+1$  :

$$\alpha(1, 2p+1) = -4b^2 h_{2p}(b)$$

Remarque : On peut aussi utiliser le développement de  $K$  en série infinie :  $K(x, y) = \sum_{i=0}^{\infty} \frac{1}{2^i} \left(\frac{x}{b}\right)^i \left(\frac{y}{b}\right)^i$

$$\begin{aligned} g_k(x) &= \int_{-b}^{+b} K(x, y) y^{k-1} dy = \sum_{i=0}^{\infty} \frac{1}{2^i b^{2i}} (x^i) \int_{-b}^{+b} y^i y^{k-1} dy \\ &= b^k \sum_{i=1}^{\infty} \left(\frac{x}{2b}\right)^i \frac{1}{i+k-1} (1 - (-1)^{i+k-1}) \end{aligned}$$

Pour  $k = 2p$  :

$$g_{2p}(x) = 2 b^{2p} \sum_{i=1}^{\infty} \left(\frac{x}{2b}\right)^{2i-1} \frac{1}{2i+2p-1}$$

Pour  $k = 2p-1$  :

$$g_{2p+1}(x) = 2 b^{2p-1} \sum_{i=1}^{\infty} \left(\frac{x}{2b}\right)^{2(i-1)} \frac{1}{2p+2i-3}$$

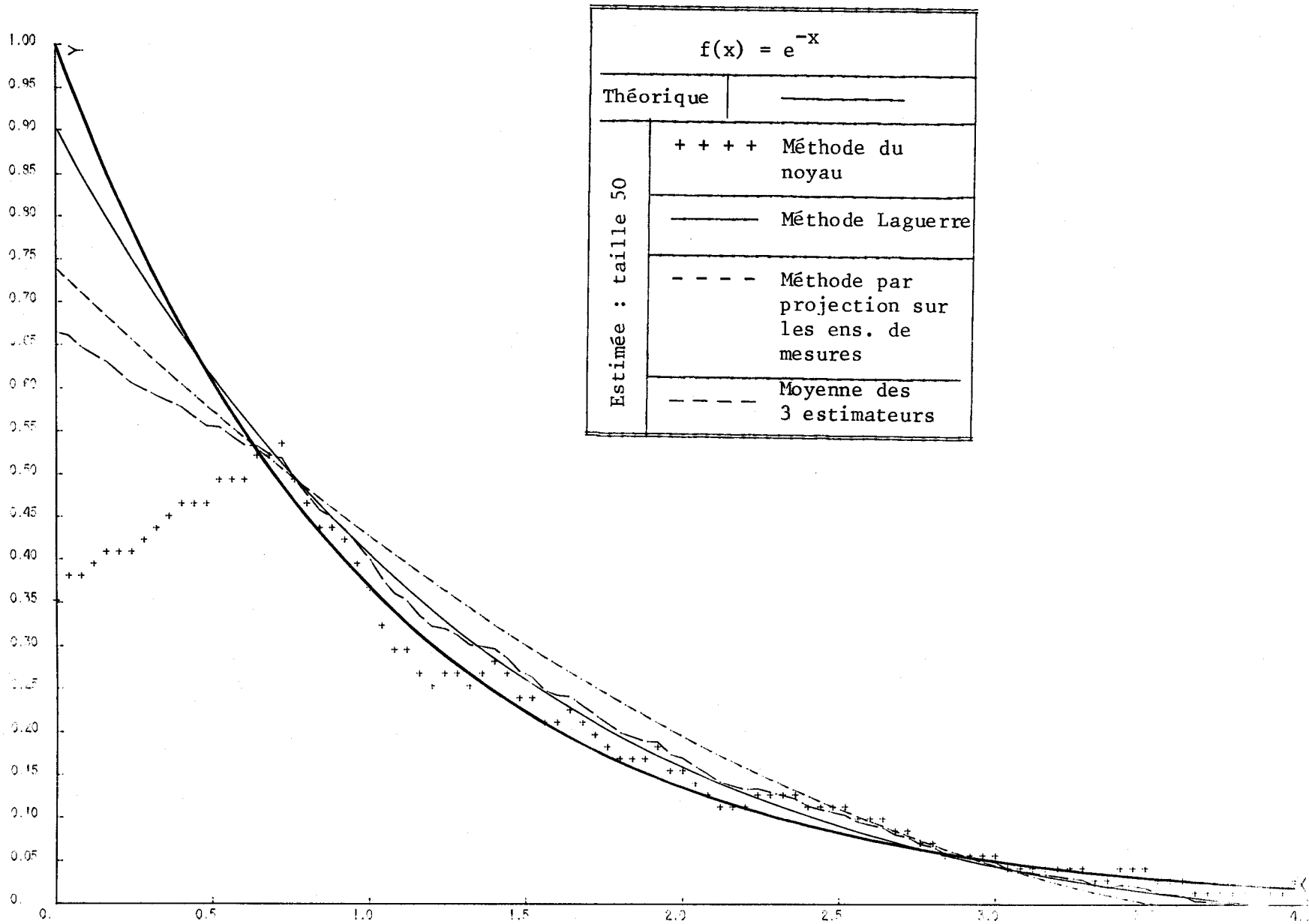
Calcul des  $\alpha(i,k)$ .

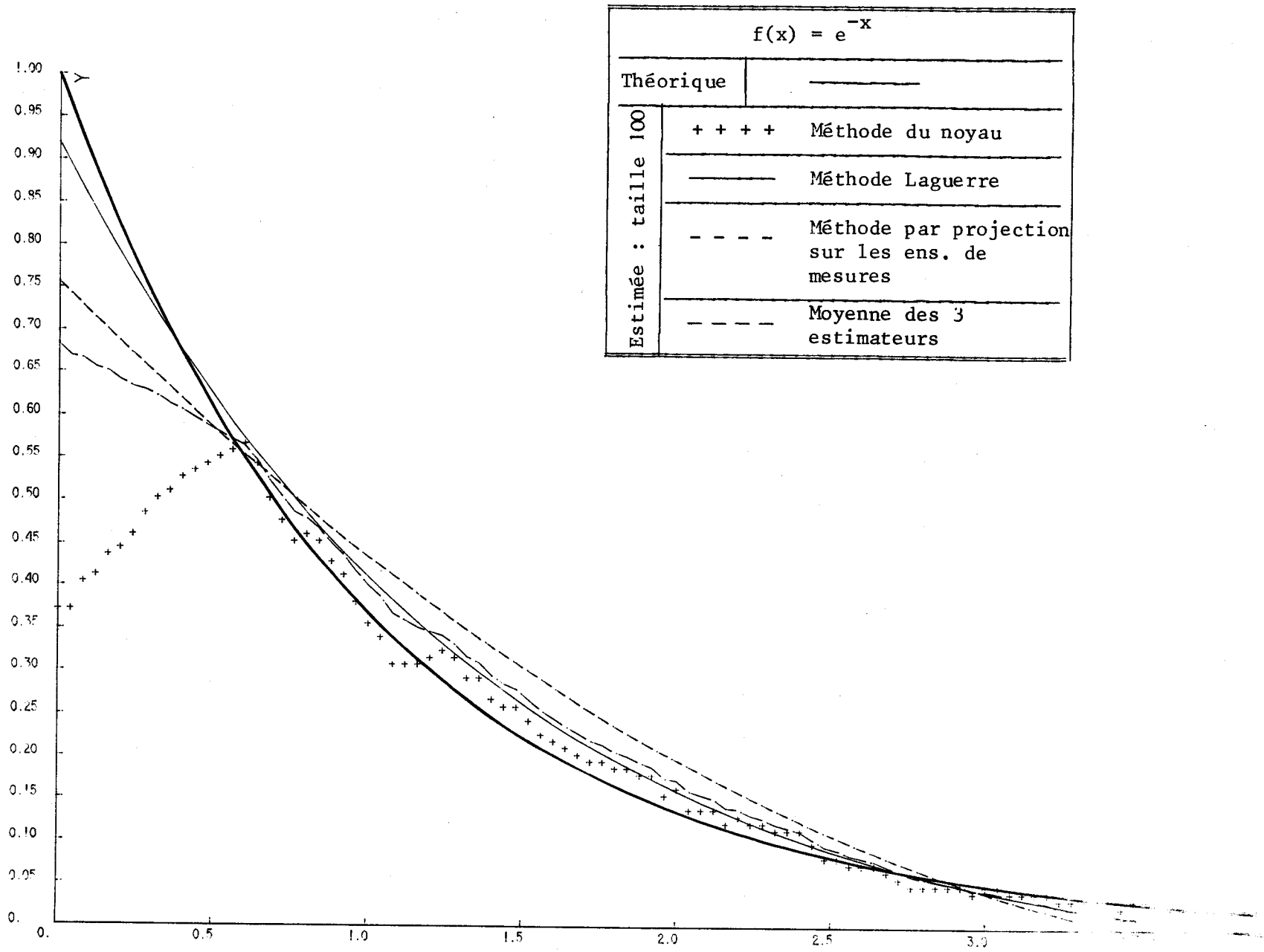
$$\alpha(i,k) = \sum_{k=0}^{\infty} \frac{1}{2^h} \left( \int_{-b}^{+b} \left(\frac{x}{b}\right)^h x^{i-1} dx \right) \left( \int_{-b}^{+b} \left(\frac{y}{b}\right)^h y^{k-1} dy \right)$$

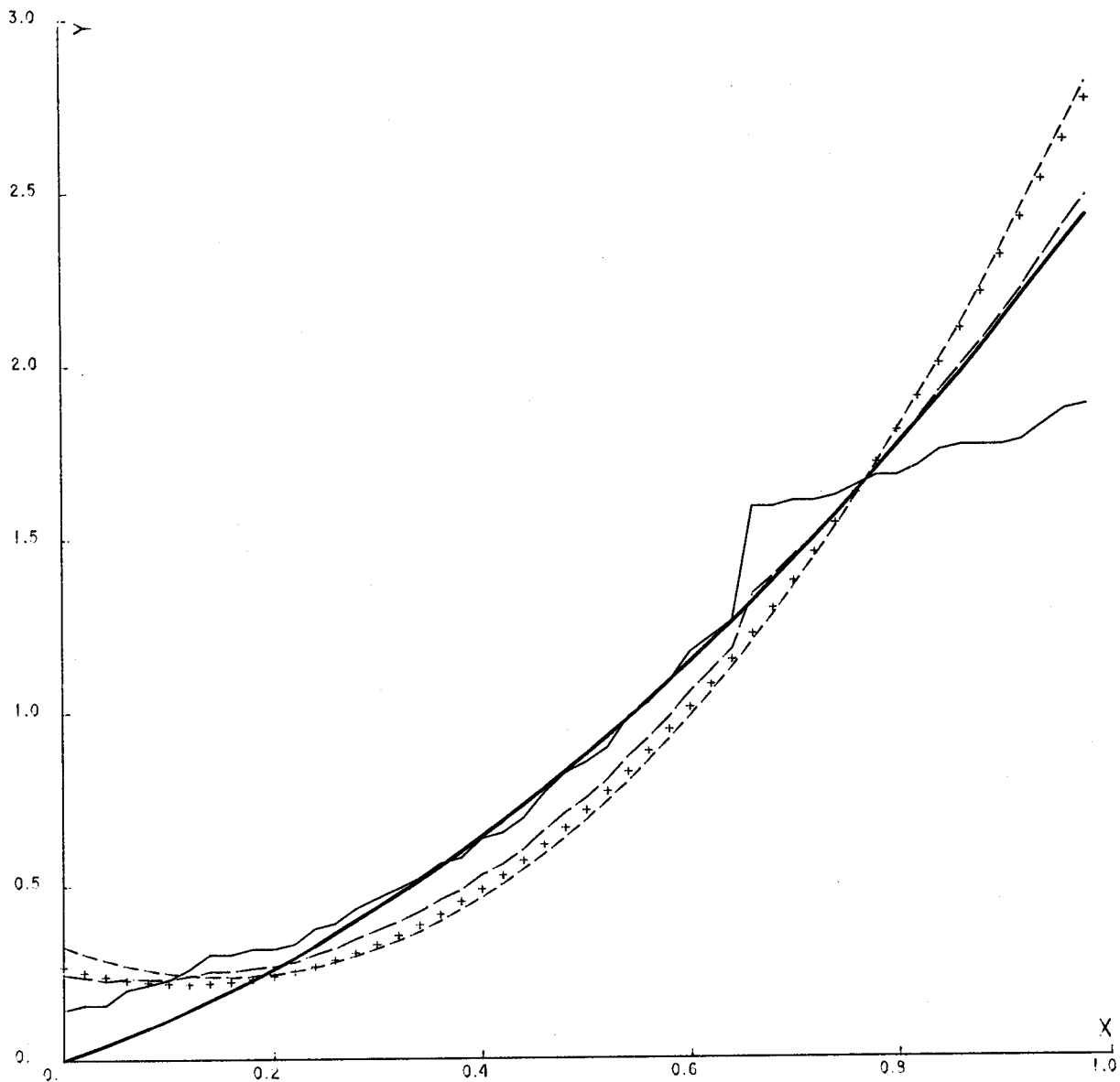
$$\alpha(2p, 2q) = 4 b^{2p+2q} \cdot \sum_{h=1}^{\infty} \frac{1}{2^{2h-1}} \cdot \frac{1}{(2p+2h-1)} \cdot \frac{1}{2q+2h-1}$$

$$\alpha(2p-1, 2q-1) = 4 b^{2(p+q-1)} \sum_{h=1}^{\infty} \frac{1}{2^{2(h-1)}} \cdot \frac{1}{(2p+2h-3)(2q+2h-3)}$$

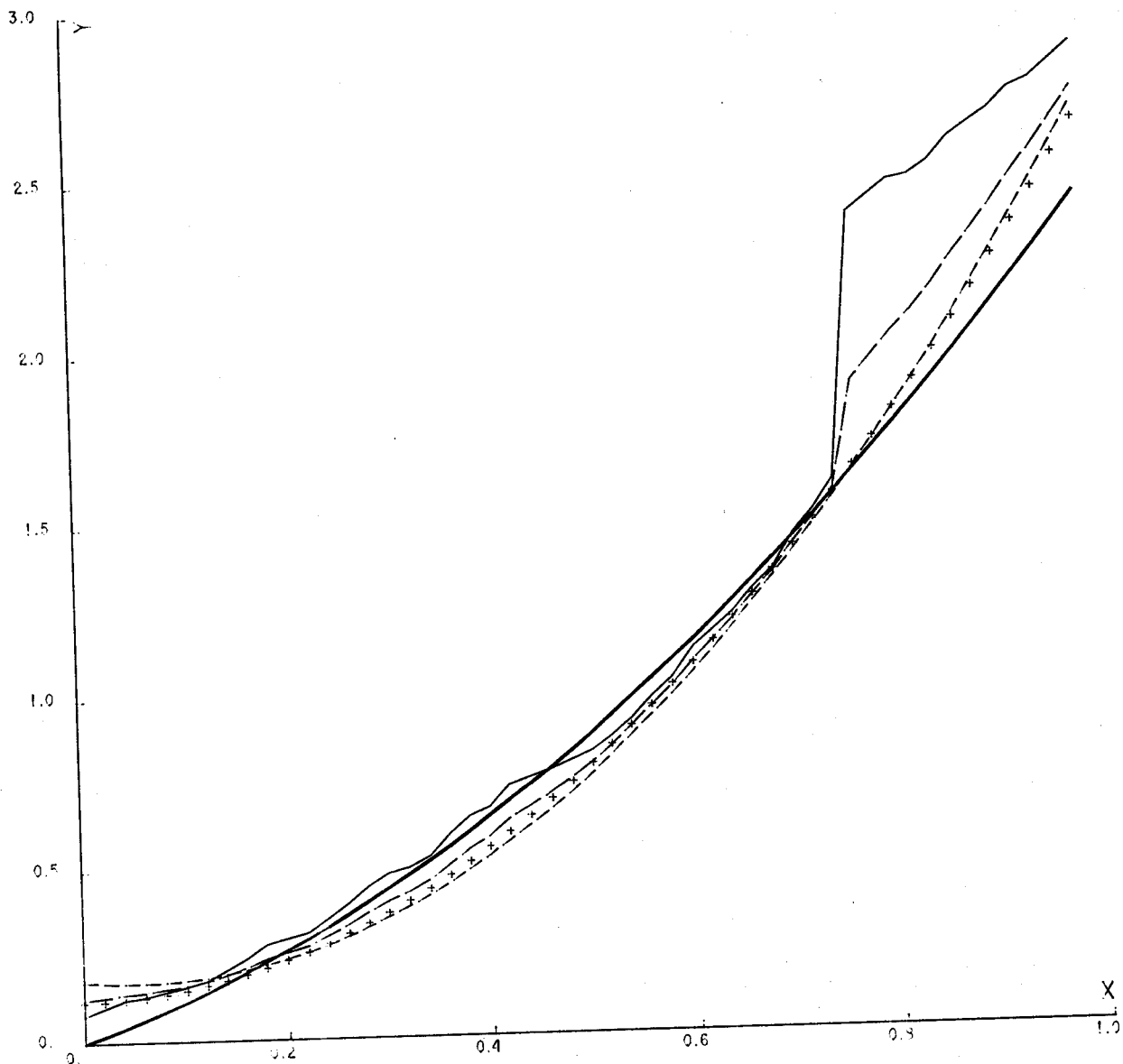
$$\alpha(2p, 2q-1) = 0$$







$f(x) = x + \frac{3}{2} x^2$	
Théorique	—————
Estimée : taille 100	————— Méthode du noyau
	- - - - Méthode par projection sur les ens. de mesures
	+ + + + Méthode Legendre
	- - - - Moyenne des trois estimateurs



$f(x) = x + \frac{3}{2}x^2$	
Théorique	—————
Estimée : taille 500	————— Méthode du noyau
	+ + + + Méthode Legendre
	- - - - Méthode par projection sur les ens. de mesures
	- - - - Moyenne des trois estimateurs

B I B L I O G R A P H I E

-----

- [1] ARONSAJN N. - La théorie des noyaux reproduisants et applications  
Proceedings of the Cambridge Philosophical Society  
(Vol. 39), 1943 (pages 133-153).
- [2] ARONSAJN N. - Theory of reproducing Kernels, dans Transactions of  
the American Mathematical Society n° 68, 1950  
(pages 307-404).
- [3] SCHWARTZ L. - Sous-espaces hilbertiens d'espaces vectoriels topolo-  
giques et noyaux associés (noyaux reproduisants),  
J. Anal. Math. of Jerusalem XIII (1964) (p. 115-256).
- [4] DUC-JACQUET M. - Approximation des fonctionnelles linéaires sur les  
espaces hilbertiens autoreproduisants,  
Thèse Université Scientifique et Médicale de Grenoble, 1973.
- [5] SHONKWILLER R. - Dilations of reproducing kernels.  
Lecture notes in Math.
- [6] NEVEU J. - Processus aléatoires gaussiens,  
Presses de l'Université de Montréal, 1968.
- [8] GUILBART C. - Etude des produits scalaires sur l'espace des mesures.  
Estimation par projections. Tests à noyaux.  
Thèse Université des Sciences et Techniques de Lille, 1980.
- [9] BERLINET A. - Espaces autoreproduisants et mesure empirique. Méthodes  
splines en estimation fonctionnelle.  
Thèse de 3ème cycle, Université des Sciences et  
Techniques de Lille, 1980.
- [10] RAJPUT B.S. et - Gaussian processes and Gaussian measures.  
CAMBANIS Ann. of Math. Stat., 1972, 43 (pages 1944-52).

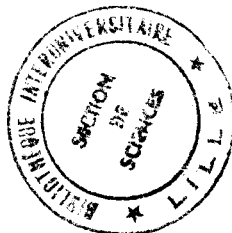


- [11] AHMAD S. - Eléments aléatoires dans les espaces vectoriels topologiques.  
Ann. Inst. Henri Poincaré Section B, Vol. II n° 2, 1965 p. 95-135.
- [12] J. GEFFROY  
H. ZEBoulON - Sur certaines convergences stochastiques des mesures aléatoires et des processus ponctuels.  
C.R. Acad. Sc. Paris, t. 280, série A-291-293 (1975)
- [13] I. GIHMAN et  
A.V. SHOROHOD - The theory of stochastic Processes, Springer-Verlag, 1974.
- [14] D. BOSQ - Cours de probabilités et intégration.  
Université des Sciences et Techniques de Lille.
- [15] BILLINGSLEY P. - Convergence of probability measures.  
J. Wiley, New-York, 1968.
- [16] PARTHASARATHY K.R. - Probability measures on metric spaces,  
Academic press 1967.
- [17] HOEFFDING W. - Probability inequalities for sums of bounded random variables,  
J.A.S.A. 58 (1963), pages 13-30.
- [18] BOSQ D. et BLEUEZ J. - Etude d'une classe d'estimateurs non paramétrique de la densité,  
Ann. Inst. Henri Poincaré, Vol. XIV, n° 4, p. 479-498.
- [19] BOSQ D. - Sur l'estimation sans biais d'un paramètre fonctionnel.  
Pub. Int. de l'U.E.R. de Mathématiques Pures et Appliquées de Lille I, n° 104, mars 1977.
- [20] BLEUEZ J. - Conditions nécessaires et suffisantes de convergence pour une classe d'estimateurs de la densité.  
Thèse de 3ème cycle (Université de Lille I), février 1976.
- [21] BOSQ D. - Contribution à la théorie de l'estimation fonctionnelle.  
Thèse publication de l'Inst. de Statistique de l'Université de Paris - octobre 1970.

- [22] LOEVE M. - Probability theory I  
4<sup>th</sup> edition.
- [23] TIAGO de OLIVEIRA J. - Estatística de densidades. Resultados assintoticos,  
Rev. Fac. Ci. Univ. Lisboa, A, 9 p. 165-171 (1961).
- [24] CENCOV N.N. - Evaluation of an unknown distribution density from  
observations, Soviet Maths 3, pp. 1559-62 (1962).
- [25] BOSQ D. - Estimation de la densité par projection sur un  
sous-espace de dimension finie.  
Portugaliae Mathematica Vol. 37 Fasc. 1-2 - 1978.
- [26] NEVEU J. - Martingales à temps discret.  
Masson 1972.
- [27] HUBER P.J. - Robust estimation of a location parameter.  
Ann. Math. Stat. 35, 73-101 (1964).
- [28] HUBER P.J. - Théorie de l'inférence statistique robuste.  
Séminaire de Mathématiques Supérieures.  
Université de Montréal.
- [28] HUBER P.J. - Robust statistics.  
Wiley Interscience 1981.
- [29] TUKEY J.W. - A survey of sampling from contaminated distributions.  
Contrib. to Prob. and Stat. (ed Olkin), pp. 448-485,  
Stanford University press (1960).
- [30] HUBER Catherine - Théorie de la robustesse et estimation d'un paramètre.  
Société Mathématique de France. Astérisque 43-44  
(1977) Chapitre VI.
- [31] BOSQ D. - Cours de statistique mathématique.  
4ème partie 1982-1983. Université de Lille I
- [32] BICKEL P.J. - Unbiased estimation in convex families.  
LEHMANN E.L. Ann. Math. Stat. vol. 40 n° 5, p. 1523-35 (1969).
- [33] AHMAD M. - GIRI N. - Estimation of the mixing proportion of two known  
and SINHA B.K. distributions.  
Sankhya : 1983, volume 45, séries A, Pt. 3, pp. 357-371.

Simulation :

- [34] ROSENBLATT M. - Remarks on some non-parametric estimates of a density fonction.  
Ann. Math. Stat. 27 (1956), 832-837.
- [35] HONIMAL P. et DEHEUVELS P. - Estimation non paramétrique de la densité compte-tenu d'information sur le support.  
Revue de Statistique appliquée, 1979, Vol. XXVII, n° 3.
- [36] EPANECHNIKOV V.A. - Non-parametric estimation of a multivariate probability density.  
Teor. Prob. Appl. 14 p. 153-158.
- [37] DEHEUVELS P. - Estimation non paramétrique de la densité par histogrammes généralisés.  
Revue de Statistique Appliquée, 1977 vol. XXV n° 3.
- [39] KRONMAL R. TARTER M. - The estimation of probability densities and cumulatives by Fourier series methods.  
J. Amer. Stat. Assoc. 63, pp. 925-52 (1968).
- [40] OULEDCHEIKH M. - Construction d'estimateurs de la densité par la méthode des fonctions orthogonales,  
thèse 3ème cycle, Université Lille I (1979).



## RÉSUMÉ

-----

Ce travail est essentiellement consacré à l'estimation d'un paramètre fonctionnel par projection sur les ensembles de mesures.

La convergence stochastique des mesures aléatoires considérées comme des variables à valeurs dans un espace de Hilbert à noyau reproduisant est étudiée.

La densité d'une loi de probabilité est estimée. Plusieurs modes de convergence sont étudiés. La loi limite de l'estimateur est fournie.

Pour une densité  $f = (1-\epsilon) g + \epsilon h$  appartenant à un modèle contaminé, le degré de contamination  $\epsilon$  est estimé. La borne de Cramer-Rao et la loi limite de l'estimateur sont fournies.

Enfin, des exemples, et des simulations illustrent cette étude.

MOTS CLÉS : ESTIMATION FONCTIONNELLE, DENSITE.  
ESTIMATION, DEGRE CONTAMINATION.  
CONVERGENCE STOCHASTIQUE.  
SIMULATION.