



**HAL**  
open science

# Contribution à la théorie de l'estimation fonctionnelle. Chap. I-III

D. Bosq

► **To cite this version:**

D. Bosq. Contribution à la théorie de l'estimation fonctionnelle. Chap. I-III. Annales de l'ISUP, 1970, XIX (2), pp.1-96. hal-04083374

**HAL Id: hal-04083374**

**<https://hal.science/hal-04083374v1>**

Submitted on 27 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## TABLE DES MATIÈRES

### CONTRIBUTION A LA THEORIE DE

Pages

INTRODUCTION 1

Chapitre I - GÉNÉRALITÉS SUR L'ESTIMATION  
L'ESTIMATION FONCTIONNELLE 3

I - Généralités sur l'estimation statistique 3

II - Estimation de paramètres fonctionnels 5

III - Estimation sans biais 9

IV - Estimation convergente 23

V - Estimation asymptotique 36

D. BOSQ

Chapitre II - ESTIMATION DE LA DENSITÉ DE  
PROBABILITÉ 43

I - Généralités 43

II - Convergence en moyenne quadratique intégrée 45

III - Convergence uniforme en moyenne quadratique  
et convergence uniforme, presque sûre 51

IV - loi limite de l'estimateur 64

V - Applications 68

Chapitre III - ESTIMATION DES DÉRIVÉES DE LA DENSITÉ 73

I - Généralités 73

II - Théorèmes asymptotiques 74

III - Etude de cas particuliers 80

CONTRIBUTION A LA THEORIE DE

L'ESTIMATION FONCTIONNELLE

D. BOSQ

# INTRODUCTION

## TABLE DES MATIÈRES

	Pages
INTRODUCTION	1
Chapitre I - GÉNÉRALITÉS SUR L'ESTIMATION FONCTIONNELLE	3
I - Généralités sur l'estimation statistique	3
II - Estimation de paramètres fonctionnels	5
III - Estimation sans biais	9
IV - Estimation convergente	23
V - Estimation symétrique	36
Chapitre II - ESTIMATION DE LA DENSITÉ DE PROBABILITÉ	43
I - Généralités	43
II - Convergence en moyenne quadratique intégrée	45
III - Convergence uniforme en moyenne quadratique et convergence uniforme presque sûre	51
IV - Loi limite de l'estimateur	64
V - Applications	68
Chapitre III - ESTIMATION DES DÉRIVÉES DE LA DENSITÉ	72
I - Généralités	72
II - Théorèmes asymptotiques	74
III - Etude de cas particuliers	80

# TABLE DES MATIÈRES

Pages	
1	INTRODUCTION
3	Chapitre I - GÉNÉRALITÉS SUR L'ESTIMATION FONCTIONNELLE
3	I - Généralités sur l'estimation statistique
5	II - Estimation de paramètres fonctionnels
9	III - Estimation sans biais
23	IV - Estimation convergente
30	V - Estimation symétrique
43	Chapitre II - ESTIMATION DE LA DENSITÉ DE PROBABILITÉ
43	I - Généralités
45	II - Convergence en moyenne quadratique intégrée
51	III - Convergence uniforme en moyenne quadratique et convergence uniforme presque sûre
64	IV - Loi limite de l'estimateur
68	V - Applications
73	Chapitre III - ESTIMATION DES DÉRIVÉES DE LA DENSITÉ
73	I - Généralités
74	II - Théorèmes asymptotiques
80	III - Étude de cas particuliers

## INTRODUCTION

L'objet de ce travail est l'étude de certains Problèmes d'estimation statistique dans lesquels le paramètre est à valeurs dans un espace fonctionnel.

Le premier chapitre est consacré à des généralités sur l'estimation fonctionnelle : nous étudions notamment les conditions d'existence d'estimateurs sans biais. Dans le cas où de tels estimateurs existent, et où l'espace des paramètres est un espace de Hilbert réel, nous montrons l'existence d'un estimateur sans biais uniformément optimal.

Cependant on constate que, dans de nombreux Problèmes d'estimation fonctionnelle, il n'est pas possible d'obtenir un estimateur sans biais (notamment dans le cas où le paramètre est « local » en un sens que nous précisons. On doit alors se limiter à une étude asymptotique du Problème : ceci nous amène à donner le principe d'une construction d'un estimateur convergent, dans le cas où le paramètre est à valeurs dans un espace métrique.

Les chapitres II et III sont consacrés à l'estimation de la densité et de ses dérivées ; la méthode employée est celle des fonctions orthogonales. Cette méthode nous semble, en effet, plus simple et plus lisse que la méthode du noyau. Nous étudions essentiellement les propriétés asymptotiques de l'estimateur, puis nous appliquons les résultats obtenus aux estimateurs construits à partir de deux systèmes orthogonaux importants : les fonctions trigonométriques et les fonctions d'Hermite.

Dans le chapitre IV, nous montrons comment, à partir d'un estimateur de la densité, on peut construire des estimateurs pour la densité conditionnelle et les fonctions de régression.

Enfin, les deux derniers chapitres sont consacrés à des questions liées plus ou moins directement à l'estimation du support d'une loi de probabilité :

Dans le chapitre V, on généralise des résultats de M. GEFFROY sur l'estimation de la fonction frontière et, dans le chapitre VI, on étudie les éléments extrémaux d'un échantillon bivarié et on construit, à l'aide des éléments maximaux de l'échantillon, un estimateur de l'ensemble des éléments maximaux du support.

Nous ne voudrions pas terminer cette présentation sans remercier vivement Monsieur le Professeur DUGUÉ de nous avoir fait l'honneur de présider le jury de cette thèse.

Nous tenons à exprimer notre profonde gratitude à Monsieur le Professeur GEFROY pour ses encouragements et la constante bienveillance qu'il a bien voulu nous manifester ; il nous a tout particulièrement aidé en nous communiquant certains de ses résultats inédits dont il a été tiré un grand parti dans le chapitre VI.

Nous remercions également Monsieur le Professeur MALLIAVIN d'avoir bien voulu faire partie de notre jury et de nous avoir proposé un deuxième sujet.

# I

## GÉNÉRALITÉS SUR L'ESTIMATION FONCTIONNELLE

### I - GÉNÉRALITÉS SUR L'ESTIMATION STATISTIQUE.

#### 1. Schéma d'un problème d'estimation.

Si l'on se limite au cas d'observations indépendantes et équidistribuées, on peut schématiser un problème d'estimation statistique de la façon suivante :

Soit  $\mathcal{A}$  une tribu sur un ensemble  $\Omega$  et soit  $\mathcal{P} = \mathcal{P}(\Omega, \mathcal{A})$  l'espace des mesures de probabilité sur l'espace mesurable  $(\Omega, \mathcal{A})$ .

Soit  $\mathcal{D}$  un sous-espace de  $\mathcal{P}$ , soit  $\Theta$  un ensemble, a priori quelconque, appelé espace des paramètres. Soit, enfin,  $\varphi$  une application de  $\mathcal{D}$  dans  $\Theta$ .

Le problème est d'estimer  $\theta = \varphi(P)$  à partir d'observations qui proviennent d'une loi de probabilité inconnue  $P \in \mathcal{D}$ . L'application qui associe à ces observations la valeur estimée du paramètre s'appellera un *estimateur*. Plus précisément :

#### DÉFINITION 1 :

Un *estimateur d'ordre*  $n$  est une application de  $\Omega^n$  dans  $\Theta$ .

#### NOTATIONS :

Dans ce chapitre, sauf avis contraire, on désignera par  $T_n$  un estimateur d'ordre  $n$  et par  $T = (T_1, \dots, T_n, \dots)$  une suite d'estimateurs.

#### 2. Définitions.

Pour évaluer  $\theta$ , on cherchera, évidemment, à obtenir des estimateurs possédant certaines propriétés de régularité et pour lesquels l'erreur



commise dans l'estimation (cette notion étant à préciser) est aussi petite que possible.

Nous donnons, ci-dessous, les définitions d'un certain nombre de propriétés que l'on peut exiger d'un estimateur.

DÉFINITION 2 :

Si  $\Theta$  est muni d'une tribu  $\mathcal{B}$ , on dira que  $T_n$  est **mesurable** s'il est  $(\mathcal{Q}^n - \mathcal{B})$  mesurable.

DÉFINITION 3 :

On dira que  $T_n$  est **strict** si  $T_n(\Omega^n)$  est inclus dans  $\varphi(\mathcal{D})$ .

DÉFINITION 4 :

Si  $\Theta$  est un espace topologique muni de sa tribu Borélienne, on dira que  $T$  **converge en probabilité** si  $\forall P \in \mathcal{D}$ ,  $\forall U$  voisinage ouvert de  $\varphi(P)$ , on a :

$$\lim_{n \rightarrow \infty} P^n^* [ T_n \in U ] = 0 \quad (1)$$

$P^n^*$  désignant la mesure extérieure associée à la mesure de probabilité  $P^n$  définie sur la tribu produit  $\mathcal{Q}^n$ .

DÉFINITION 5 :

Avec les mêmes hypothèses sur  $\Theta$ , on dira que  $T$  **converge presque sûrement** si

$$P^{\infty*} [ T_n \not\rightarrow \varphi(P) ] = 0 \quad \forall P \in \mathcal{D} \quad (2)$$

$P^{\infty*}$  désignant la mesure extérieure associée à la mesure de probabilité produit  $P^\infty$  définie sur la tribu produit  $\mathcal{Q}^\mathbb{N} = \mathcal{Q}^\infty$ .

DÉFINITION 6 :

On suppose que  $\Omega$  est un espace topologique localement compact et que  $\Theta$  est un espace vectoriel topologique localement convexe séparé.

On dira que  $T_n$  est un **estimateur sans biais** de  $\varphi(P)$  si,  $\forall P \in \mathcal{D}$  :

1)  $T_n$  est scalairement essentiellement  $P^n$ -intégrable et admet une intégrale par rapport à  $P^n$ .

$$2) E_P [T_n] = \int_{\Omega^n} T_n(\omega) dP^n(\omega) = \varphi(P) \quad (3)$$

DÉFINITION 7 :

Soit  $\mathcal{T}_n$  une famille d'estimateurs d'ordre  $n$  et soit  $R(T_n, P)$  une fonction de risque.

On dira que  $T_n^0$  est **uniformément optimal** (dans la famille  $\mathcal{T}_n$  et pour la fonction de risque  $R$ ) si,  $\forall P \in \mathcal{D}$ ,  $\forall T_n \in \mathcal{T}_n$

$$R(T_n^0, P) \leq R(T_n, P) \quad (4)$$

Il n'est pas possible, en général, de construire un estimateur vérifiant toutes les propriétés précédentes.

Nous verrons que, sous des conditions assez larges, on peut obtenir un estimateur strict, mesurable et convergent (§IV). Au contraire, l'existence d'un estimateur sans biais n'est vérifiée que dans des cas particuliers (cf §III) ; il en est de même pour l'existence d'un estimateur uniformément optimal (cf §V).

## II - ESTIMATION DE PARAMÈTRES FONCTIONNELS.

### 1. Exemples.

Dans la plupart des problèmes classiques d'estimation, l'espace  $\Theta$  est un sous-ensemble de  $\mathbf{R}$  ou de  $\mathbf{R}^k$ . Dans ce travail, nous étudierons principalement des problèmes d'estimation où l'espace  $\Theta$  est soit un espace vectoriel topologique (en général normé), soit un espace métrique (ou pseudo-métrique).

Indiquons quelques exemples de problèmes de ce type.

Exemple 1 :

$\Omega = \mathbf{R}^k$ ,  $\mathcal{A} = \mathcal{B}_k$  (tribu Borélienne de  $\mathbf{R}^k$ ),

$\mathcal{D} = \mathcal{P}(\mathbf{R}^k, \mathcal{B}_k)$ ,  $\Theta = B(\mathbf{R}^k, \mathbf{R})$  (espace des applications bornées de  $\mathbf{R}^k$  dans  $\mathbf{R}$  muni de la topologie de la convergence uniforme).

$\varphi(P) = F$  fonction de répartition de  $P$  :

$$F(x_1, \dots, x_k) = P \left[ (-\infty, x_1 [ , (-\infty, x_2 [ , \dots, (-\infty, x_k [ ] \right]$$

$$\forall (x_1, \dots, x_k) \in \mathbf{R}^k .$$

Exemple 2 :

$\mu$  mesure sur  $(\Omega, \mathcal{A})$  quelconque.

$\mathcal{D} =$  Ensemble des lois de probabilités sur  $(\Omega, \mathcal{A})$  admettant, par rapport à  $\mu$ , une densité de Radon-Nikodym de carré  $\mu$ -intégrable.

$\Theta =$  espace de Hilbert  $L^2(\mu)$

$$\varphi(P) = \frac{dP}{d\mu} .$$

Exemple 3 :

$\Omega = \mathbf{R}$ ,  $\mathcal{A} = \mathcal{B}$  (tribu Borélienne de  $\mathbf{R}$ )<sup>(1)</sup>,

$\mathcal{D} =$  ensemble des lois de probabilité  $P$  absolument continues par rapport à la mesure de Lebesgue  $\lambda$  sur  $\mathbf{R}$  et possédant une version continue de la densité :  $f_p$ .

$\Theta = C(\mathbf{R})$  (Espace des fonctions continues bornées sur  $\mathbf{R}$  muni de la topologie de la convergence uniforme).

$$\varphi(P) = f_p .$$

(1) Dans la suite, nous supposons, sauf avis contraire, que les espaces topologiques sont munis de leur tribu Borélienne.

Exemple 4 :

$\Omega = \mathbf{R}^2$ ,  $\mathcal{D}$  = Ensemble des mesures de probabilité absolument continues, à densité  $f(x, y)$  continue, telles que

$$\int_{\mathbf{R}} y f(x, y) dy$$

existe et soit continue  $\forall x \in \mathbf{R}$  et telles que

$$\int_{\mathbf{R}} f(x, y) dy$$

soit continue et positive  $\forall x \in \mathbf{R}$ .

$\Theta$  = espace des fonctions continues sur  $\mathbf{R}$ , muni de la topologie de la convergence uniforme sur tout compact.

$\varphi(P)$  = **Fonction de régression** de  $y$  en  $x$  =

$$\int_{\mathbf{R}} y f(x, y) dy / \int_{\mathbf{R}} f(x, y) dy$$

Exemple 5 :

$\Omega = \mathbf{R}^2$ ,  $\mathcal{D}$  = Ensemble des mesures de probabilité absolument continues, dont le support  $S$  est de la forme :

$$S = \{ (x, y) \in \mathbf{R}^2 \mid x \in [0, 1], 0 \leq y \leq s_p(x) \}$$

où  $s_p$  est une fonction définie et continue sur  $[0, 1]$ .

$\Theta = C[0, 1]$  (espace des fonctions continues sur  $[0, 1]$  muni de la topologie de la convergence uniforme),

$\varphi(P) = s_p$  (on dira que  $s_p$  est la **fonction frontière** de  $P$ ).

Exemple 6 :

$\Omega = \mathbf{R}^2$ ,  $\mathcal{D}$  = ensemble des lois de probabilité absolument continues, à

support compact.

$\Theta$  = Ensemble des compacts de  $\mathbf{R}^2$  muni de la topologie induite par la pseudo-métrie :

$$d(A,B) = \lambda^2 (A \Delta B) \quad \forall A, B \in \Theta$$

où  $\lambda^2$  désigne la mesure de Lebesgue sur  $\mathbf{R}^2$  et  $\Delta$  l'opération différence symétrique.

$\varphi(P)$  = Support de P.

Nous reviendrons sur ces exemples dans la suite de ce travail.

## 2. Choix d'une fonction de coût.

Pour comparer l'efficacité des diverses méthodes d'estimation, il est nécessaire de choisir une fonction de coût.

Si  $\Theta$  est un espace normé, il semble naturel de choisir la fonction

$$l [T_n(\omega_1, \dots, \omega_n) ; \varphi(P)] = \| T_n(\omega_1, \dots, \omega_n) - \varphi(P) \|^2 \quad (5)$$

$$\forall (\omega_1, \dots, \omega_n) \in \Omega^n, \forall P \in \mathcal{D}$$

La fonction de risque correspondante s'écrit alors :

$$R(T_n, \varphi(P)) = E_P \| T_n(\omega_1, \dots, \omega_n) - \varphi(P) \|^2 \quad \forall P \in \mathcal{D} \quad (6)$$

Elle généralise le MISE (Mean integrated square error) introduit par PARZEN [1] ; en effet, le MISE est défini par :

$$J_n = E \int_{\mathbf{R}} [\hat{f}_n(x) - f(x)]^2 dx \quad f \in L^2, \hat{f}_n \in L^2 \quad (7)$$

où  $f$  est le paramètre ( $\Theta = L^2$ ) et  $\hat{f}_n$  l'estimation de  $f$ .

*Propriétés de  $\ell$*

Si  $\Theta$  vérifie les hypothèses convenables, la fonction  $\ell(q) = \|q\|^2$

a les propriétés usuelles d'une fonction de coût :

$$1) \quad q = 0 \implies \ell(q) = 0 \quad (8)$$

$$2) \quad \ell(-q) = \ell(q) \quad (9)$$

3) Dans un treillis vectoriel normé :

$$|q_2| \geq |q_1| \implies \ell(|q_2|) \geq \ell(|q_1|) \quad (10)$$

4) Dans un espace de Hilbert :

$$\ell(\alpha q_1 + (1-\alpha)q_2) \leq \alpha \ell(q_1) + (1-\alpha) \ell(q_2) \quad \forall \alpha \in [0,1] \quad (11)$$

#### REMARQUES.

a) La définition de  $\ell$  s'étend sans difficulté au cas où  $\Theta$  est un espace pseudo-métrique.

b) La fonction  $\ell$  a l'avantage d'être maniable (surtout dans un espace de Hilbert), mais elle ne donne pas une idée entièrement satisfaisante de la précision d'une estimation. Dans  $L^2$  (par exemple),  $\ell$  ne tient pas compte de la corrélation entre  $\hat{f}_x(x)$  et  $\hat{f}_n(x')$  ( $x \neq x'$ ). Toutefois, il semble difficile de définir une fonction de coût plus satisfaisante que  $\ell$ , et qui reste facilement utilisable dans la pratique.

### III - ESTIMATION SANS BIAIS.

Dans ce paragraphe, nous donnons quelques propriétés des estimateurs sans biais et étudions des conditions d'existence pour de tels estimateurs.

#### 1. Propriétés d'un estimateur sans biais.

a) Rappelons que si  $T_n$  est scalairement essentiellement  $P^n$ -intégrable et admet une intégrale par rapport à  $P^n$ , cette intégrale est caractérisée par la relation :

$$\left\langle \int_{\Omega^n} T_n(\omega) dP^n(\omega), \theta' \right\rangle = \int_{\Omega^n} \langle T_n(\omega), \theta' \rangle dP^n(\omega) \quad (12)$$

$\forall \theta' \in \Theta'$  dual topologique de  $\Theta$ . Compte tenu de la définition 6 (I - 2), le fait que  $T_n$  est sans biais s'écrira donc :

$$\int_{\Omega^n} \langle T_n(\omega), \theta' \rangle dP^n(\omega) = \langle \varphi(P), \theta' \rangle \quad \forall \theta' \in \Theta', \forall P \in \mathcal{D} \quad (13)$$

b) Donnons, maintenant, une généralisation d'une propriété bien connue de la variance d'un estimateur sans biais.

**THÉORÈME 1 :**

Si  $\Theta$  est un espace de Hilbert réel, si  $T_n$  est essentiellement  $P^n$ -intégrable et si  $E_P(T_n) \in \varphi(\mathcal{D}), \forall P \in \mathcal{D}$ . alors, pour que  $T_n$  soit sans biais, il faut et il suffit que :

$$E_P \|T_n - \varphi(P)\|^2 < E_P \|T_n - \theta\|^2 \quad \forall \theta \in \varphi(\mathcal{D}), \forall P \in \mathcal{D} \quad (14)$$

**DÉMONSTRATION.**

Si  $T_n$  est essentiellement  $P^n$ -intégrable, on sait (cf, COURREGE [2] p, 227 th. 1A) que  $T_n$  est scalairement essentiellement  $P^n$ -intégrable et qu'il admet une intégrale par rapport à  $P^n$ .

D'autre part, il est facile de voir que  $\|T_n - \theta\|^2$  est  $P^n$ -mesurable,  $\forall \theta \in \varphi(\mathcal{D})$  (cf, COURREGE [2] p. 227 prop. 4) donc que les deux membres de (14) ont un sens (leurs valeurs étant finies ou infinies).

Ceci dit, on peut écrire,  $\forall \theta \in \varphi(\mathcal{D})$  :

$$\begin{aligned} \|T_n - \theta\|^2 &= \|T_n - E_P(T_n)\|^2 + \|E_P(T_n) - \theta\|^2 + 2 \langle T_n - E_P(T_n), E_P(T_n) - \theta \rangle, \dots \\ &\dots E_P(T_n) - \theta > \end{aligned} \quad (15)$$

En écrivant (15) pour  $\theta = \varphi(P)$  et en faisant la différence, il vient :

$$\begin{aligned} \|T_n - \theta\|^2 - \|T_n - \varphi(P)\|^2 &= \|E_P(T_n) - \theta\|^2 - \|E_P(T_n) - \varphi(P)\|^2 \\ &+ 2 \langle T_n - E_P(T_n), \varphi(P) - \theta \rangle \end{aligned} \quad (16)$$

D'où, en passant aux espérances et en tenant compte de (12) :

$$\begin{aligned} E_P \|T_n - \theta\|^2 - E_P \|T_n - \varphi(P)\|^2 &= \|E_P(T_n) - \theta\|^2 - \|E_P(T_n) - \varphi(P)\|^2 \\ &\dots \end{aligned} \quad (17)$$

Donc (14)  $\Leftrightarrow \|E_P(T_n) - \theta\|^2 \geq \|E_P(T_n) - \varphi(P)\|^2$

$$\forall \theta \in \varphi(\mathcal{D}), \forall P \in \mathcal{D} \quad (18)$$

Si  $T_n$  est sans biais, on a (18) donc (14) ; réciproquement, si (14) est vérifié, on a (18) et en prenant  $\theta = E_P(T_n)$  dans (18), on en déduit que  $T_n$  est sans biais.

REMARQUES :

a) si  $\varphi(\mathcal{D})$  est un cône convexe complet, si  $T_n$  est strict et si l'application  $\omega \rightarrow T_n(\omega)$  est continue à support compact, alors  $T_n$  est  $P^n$ -intégrable et  $E_P(T_n) \in \varphi(\mathcal{D}), \forall P \in \mathcal{D}$ . (cf. COURREGE [2] p.216, cor. 1).

b) LEHMANN [3] a proposé une définition de la notion d'estimateur sans biais <sup>(1)</sup> valable pour un  $\Theta$  quelconque :  $R$  désignant une fonction de risque, on dit que  $T_n$  est sans biais si :

$$R(T_n, \varphi(P)) \leq R(T_n, \theta) \quad \forall \theta \in \varphi(\mathcal{D}) \quad (19)$$

Le théorème 1 montre que si  $R$  désigne la fonction de risque définie par (6) et si  $\Theta$  est un espace de Hilbert, la définition 6 coïncide avec celle de Lehmann.

(1) et plus généralement de la notion de procédure statistique sans biais.



## 2. Cas d'un paramètre fonctionnel.

Les paramètres que nous voulons estimer étant en général des fonctions, il est intéressant de savoir dans quelle mesure un estimateur sans biais « point par point » est sans biais au sens de la définition 6.

Plus précisément, si  $\Theta$  est un espace de fonctions  $x \rightarrow f(x)$  définies sur un ensemble  $E$  quelconque et à valeurs réelles, et si l'on note  $T_n(\omega, x)$  un estimateur de  $\varphi(P)$ , est-ce que l'on a l'équivalence

$$T_n \text{ sans biais} \Leftrightarrow E_P(T_n(\omega, x)) = \varphi(P)(x), \forall x \in E ? \quad (20)$$

Nous allons voir que la réponse est positive pour les espaces  $L^2$  et  $C[0,1]$ , des hypothèses convenables étant faites sur  $T_n$ .

LEMME 1 :

si  $\Omega = \mathbf{R}$ ,  $\Theta = L^2$  et si  $T_n(\omega, x)$  est de carré  $P^n$   $x$   $\lambda$ -intégrable alors,  $T_n$  est sans biais si et seulement si

$$E_P [T_n^*(\omega, x)] = \varphi(P)^*(x) \quad \text{p.p.} \quad (21)$$

DÉMONSTRATION :

$T_n$  étant toujours supposé scalairement essentiellement  $P^n$  intégrable et intégrable dans  $L^2$ , on a, d'après (12) :

$$\int_{\mathbf{R}} \Psi(P)^*(x) g^*(x) dx = \int_{\mathbf{R}} dP^n(\omega) \int_{\mathbf{R}} T_n^*(\omega, x) g^*(x) dx \quad (22)$$

$$\forall g \in L^2$$

où l'on a posé  $\Psi(P) = \int_{\Omega^n} T_n(\omega) dP^n(\omega)$

(1) si  $f \in L^2$ ,  $f^*$  désigne une version de  $f$ .

en appliquant le théorème de Fubini, il vient, en supposant (21) vérifié :

$$\int_{\mathbf{R}} \Psi(P)^*(x) \cdot g^*(x) dx = \int_{\mathbf{R}} g^*(x) dx \int_{\mathbf{R}^n} T_n^*(\omega, x) dP_n(\omega) \quad (23)$$

$$= \int_{\mathbf{R}} \varphi(P)^*(x) g^*(x) dx \quad \forall g \in L^2 \quad (24)$$

on en tire  $\Psi(P)^*(x) = \varphi(P)^*(x)$  p.p. donc  $T_n$  est sans biais.

Réciproquement : si  $T_n$  est sans biais, on peut écrire :

$$\int_{\Omega^n} dP^n(\omega) \int_{\mathbf{R}} T_n^*(\omega, x) g(x) dx = \int_{\mathbf{R}} \varphi(P)^*(x) g^*(x) dx \quad \forall g \in L^2 \quad (25)$$

d'où, d'après le théorème de Fubini :

$$\int_{\mathbf{R}} g^*(x) dx \int_{\Omega^n} T_n^*(\omega, x) dP^n(\omega) = \int_{\mathbf{R}} \varphi(P)^*(x) g^*(x) dx \quad \forall g \in L^2 \quad (26)$$

donc  $\int_{\Omega} T_n^*(\omega, x) dP^n(\omega) = \varphi(P)^*(x)$  p.p. (27)

C.Q.F.D.

LEMME 2 :

Si  $\Omega = \mathbf{R}$ ,  $\Theta = C[0,1]$ , si  $T_n(\omega, x)$  est  $P^n_{x\mu}$ -intégrable,  $\forall \mu$  mesure sur  $[0,1]$  et si  $\int_{\mathbf{R}^n} T_n(\omega, x) dP_n(\omega)$  existe  $\forall x \in [0,1]$ ,  $\forall P \in \mathcal{D}$  ; alors,  $T_n$  est sans biais si et seulement si :

$$E_P [T_n(\omega, x)] = \varphi(P)(x) \quad \forall x \in [0,1] \quad (28)$$

DÉMONSTRATION :  $T_n$  étant encore supposé scalairement essentiellement  $P^n$ -intégrable et intégrable dans  $C[0,1]$ ,  $\forall P \in \mathcal{D}$  ; la

démonstration est analogue à celle du lemme 1 (utilisation de (12) et du théorème de Fubini).

Ces deux lemmes ramènent, dans les espaces de fonctions usuels, la recherche d'estimateurs sans biais à la recherche d'estimateurs sans biais « point par point ».

#### REMARQUE :

La situation est plus compliquée dans  $C(\mathbf{R})$  car  $\Theta'$  est alors l'espace vectoriel des fonctions d'ensembles additives, régulières et bornées, définies sur l'Algèbre engendrée par les fermés de  $\mathbf{R}^{(1)}$ .

La démonstration du lemme 2 permet seulement de montrer que (28) est une condition nécessaire pour que  $T_n$  soit sans biais.

### 3. Existence d'un estimateur sans biais.

Les conditions d'existence d'un estimateur sans biais ont été étudiées principalement par HALMOS [5] et par BICKEL et LEHMANN [6]. Nous complétons leurs résultats en étudiant l'estimation sans biais de paramètres que nous appellerons « paramètres locaux ».

#### α) LE THÉORÈME DE BICKEL-LEHMANN.

##### DÉFINITION 8 :

*On dit que  $\varphi$  est estimable s'il existe un entier  $n$  pour lequel  $\varphi$  admet un estimateur sans biais d'ordre  $n$ .*

*Si  $\varphi$  est estimable, le degré de  $\varphi$  est le plus petit entier  $n$  pour lequel  $\varphi$  est estimable.*

##### EXEMPLES :

a)  $\Omega = \mathbf{R}$ ,  $\mathcal{D}$  = ensemble des mesures de probabilité sur  $(\mathbf{R}, \mathcal{B})$  à variance finie,  $\varphi(P) = \text{variance de } P$  :  $\varphi$  est estimable de degré 2.

(1) cf. DUNFORD-SCHWARTZ [4] I p. 261-62.

b) Mêmes hypothèses,  $\varphi(P) = \text{écart type de } P$  :  $\varphi$  n'est pas estimable.

c) Exemple 1 du §II : la **fonction de répartition** est estimable de degré 1, point par point.

d)  $\Omega = \mathbf{R}$ ,  $\mathcal{D} = \mathcal{P}(\mathbf{R})$ ,  $\Theta = C(\mathbf{R}, C)$  (espace des applications continues bornées de  $\mathbf{R}$  dans  $C$ ),  $\varphi(P) = \int_{\mathbf{R}} e^{itx} dP(x)$   $t \in \mathbf{R}$  : la **fonction caractéristique** est estimable de degré 1, point par point.

THÉORÈME 2 (BICKEL-LEHMANN).

Pour  $\Theta = \mathbf{R}^{(1)}$ , si  $\mathcal{D}$  est convexe et si  $\varphi$  est estimable, pour que  $\varphi$  soit de degré  $n$ , il faut et il suffit que :

1°)  $\forall P, Q \in \mathcal{D}$   $\varphi[\alpha P + (1-\alpha)Q]$  est un polynôme de degré  $\leq n$  en  $\alpha$  ( $\alpha \in [0,1]$ ).

2°)  $\forall P \in \mathcal{D}$ ,  $\exists Q \in \mathcal{D}$  tel que  $\varphi[\alpha P + (1-\alpha)Q]$  est un polynôme de degré  $n$  en  $\alpha$  ( $\alpha \in [0,1]$ ).

DÉMONSTRATION.

Voir BICKEL-LEHMANN [6] p, 1525-26.

APPLICATION :

Il est clair que, si  $\varphi[\alpha P + (1-\alpha)Q]$  n'est pas un polynôme, le théorème 2 permet d'affirmer que  $\varphi$  n'est pas estimable.

EXEMPLES :

e) Si on reprend l'exemple 4 (II - 1), on voit immédiatement que  $\mathcal{D}$  est convexe et que  $\varphi$  n'est pas un polynôme : la **fonction de régres - sion n'est pas estimable point par point.**

(1) Ce théorème se généralise immédiatement au cas où  $\Theta$  est un espace de Banach.

f) En reprenant les hypothèses de l'exemple 4 (II - 1) et en posant  $\varphi(P) = f(x,y) / \int_{\mathbf{R}} f(x,y)dy$  (densité de  $y$  conditionnée par  $x$ ), on en déduit **que la densité conditionnelle n'est pas estimable point par point.**

g) Si on reprend l'exemple 5 (II - 1), on vérifie de même que  $\mathcal{D}$  est convexe et que  $\varphi$  n'est pas un polynôme d'où, en appliquant le lemme 2 (III - 2)<sup>(1)</sup> :

**la fonction frontière n'est pas estimable.**

h)  $\Omega = \mathbf{R}$ ,  $\mathcal{D} = \mathcal{P}(\mathbf{R})$ ,  $\Theta =$  Ensemble des applications bornées de  $\mathbf{R}^+$  dans  $\mathbf{R}$  muni de la topologie de la convergence uniforme,  $\varphi(P) = q$  avec

$q(\ell) = \sup_{z \in \mathbf{R}} P([z, z+\ell])$ ,  $\forall \ell \geq 0$  : **la fonction de concentration n'est pas**

**estimable point par point.** (On vérifie facilement que  $\varphi(\alpha \delta_a + (1-\alpha) \delta_b)$ , où  $\delta_a$  et  $\delta_b$  sont les mesures de Dirac aux points  $a$  et  $b$  ( $a \neq b$ ), n'est pas un polynôme en  $\alpha$ ).

### β) CAS D'UN PARAMÈTRE LOCAL.

Dans ce paragraphe, on suppose que  $\Omega$  est un espace topologique localement compact métrisable et que  $\mu$  est une mesure  $\sigma$ -finie sur  $\Omega$ .

#### DÉFINITION 9 :

Soit  $F$  un fermé de  $\Omega$ , on dira que  $\varphi$  est localisé par  $F$  si,  $\forall U$  voisinage ouvert de  $F$ ;  $\forall P, Q \in \mathcal{D}$ ,

$$P \upharpoonright U = Q \upharpoonright U \implies \varphi(P) = \varphi(Q)$$

$P \upharpoonright U$  et  $Q \upharpoonright U$  désignant les restrictions de  $P$  et  $Q$  à  $U$ .

On dira que  $\varphi$  est local s'il est localisé par un point.

On dira que  $\varphi$  est  $\mu$ -local s'il est localisé par un fermé de  $\mu$ -mesure nulle.

(1) Ou directement le théorème 2.

Enfin, on dira que  $\varphi$  est global s'il n'est localisé que par  $F = \Omega$ .

EXEMPLES :

i)  $\Omega = \Theta = \mathbf{R}$ ,  $\mathcal{D}$  = ensemble des mesures de probabilité à fonction de répartition strictement croissante,  $\varphi(P)$  = médiane de  $P$ ,  $\varphi$  est un paramètre global.

j)  $\Omega = \Theta = \mathbf{R}$ ,  $\mathcal{D} = \mathcal{P}$ ,  $\varphi(P) = P(-\infty, a]$ ,  $a$  donné  $\in \mathbf{R}$ .  $\varphi$  est localisé par  $(-\infty, a]$ . En revanche, la fonction de répartition (cf, exemple 1 (II - 1)) est évidemment un paramètre global.

k)  $\Omega = \Theta = \mathbf{R}$ ,  $\mathcal{D} = \mathcal{P}$ ,  $\varphi(P) = P(\{a\})$ ,  $a$  donné  $\in \mathbf{R}$ .  $\varphi$  est local, Si  $\mu$  est la mesure de Lebesgue sur  $\mathbf{R}$ ,  $\varphi$  est aussi  $\mu$ -local.

DÉFINITION 10 :

$\Theta$  étant un espace vectoriel, on dira que  $\varphi$  est un paramètre **homogène sur**  $F$  (fermé de  $\Omega$ ) si,  $\forall U$  voisinage ouvert de  $F$  et  $\forall P \in \mathcal{D}$  :

$$1^{\circ}) P(U) > 0 \Rightarrow P_U \in \mathcal{D}^{(1)} \text{ et } \varphi(P_U) = \frac{\varphi(P)}{P(U)} \quad (30)$$

$$2^{\circ}) P(U) = 0 \Rightarrow \varphi(P) = 0$$

EXEMPLES :

l)  $\Omega = \mathbf{R}^p$ ,  $\Theta = \mathbf{R}$ ,  $\mathcal{D}$  = ensemble des mesures de probabilité  $P$  absolument continues par rapport à  $\lambda^p$  et à densité  $f_p$  continue dans un voisinage du point  $a$ ,  $a$  donné  $\in \mathbf{R}$ .

$\varphi(P) = f_p(a)$ .  $\varphi$  est un paramètre homogène sur  $\{a\}$ .

m)  $\Omega = \Theta = \mathbf{R}$ ,  $\mathcal{D}$  = ensemble des mesures de probabilité  $P$  absolument continues, à densité  $f_p$   $k$  fois continuellement différentiable dans un voisinage du point  $a \in \mathbf{R}$ .

$\varphi(P) = f_p^{(k)}(a)$ .  $\varphi$  est un paramètre homogène sur  $\{a\}$ .

(1)  $P_U$  désigne la mesure de probabilité induite par  $P$  sur  $U$ .



- Remarquons qu'un paramètre homogène sur  $F$  est localisé par  $F$ . En effet :

$$P \uparrow U = Q \uparrow U \implies P(U) = Q(U)$$

si  $P(U) > 0$ , on a aussi  $P_U = Q_U$ , donc  $\varphi(P_U) = \varphi(Q_U)$

d'où  $\varphi(P) = \varphi(Q)$ . Si  $P(U) = 0$ , on a immédiatement  $\varphi(P) = \varphi(Q) = 0$ ;

La réciproque n'est pas vraie : la fonction frontière au point  $a$  ( $a \in ]0, 1[$ ) est localisée par la droite  $x = a$  mais n'est pas homogène sur cette droite.

Donnons maintenant deux résultats relatifs à l'estimation sans biais des paramètres locaux.

**THÉORÈME 3 :**

*Si  $\varphi$  est un paramètre à valeurs réelles, non constant, homogène sur un fermé  $F$  et tel que,  $\forall P \in \mathcal{D}$ ,  $F$  soit de  $P$ -mesure nulle, alors  $\varphi$  est de degré supérieur à 1<sup>(1)</sup>.*

**DÉMONSTRATION.**

Nous allons raisonner par l'absurde.

Soit  $T$  un estimateur de  $\varphi$ , sans biais d'ordre 1, on a :

$$\int_{\Omega} T(\omega) dP(\omega) = \varphi(P) \quad \forall P \in \mathcal{D} \tag{32}$$

Soit  $(U_k)_{k \in \mathbb{N}}$  une suite décroissante de voisinages ouverts de  $F$ , telle que  $U_k \downarrow F$  quand  $k \rightarrow \infty$ .

Si  $P(U_k) > 0$ , on peut écrire, compte tenu de (30) :

$$\varphi(P_{U_k}) = \int_{\Omega} T(\omega) dP_{U_k}(\omega) = \int_{U_k} T(\omega) \frac{dP(\omega)}{P(U_k)} = \frac{\varphi(P)}{P(U_k)} \tag{33}$$

(1) Si  $\varphi$  n'est pas estimable, nous posons par convention degré de  $\varphi = +\infty$ .

On en tire :

$$\int_{U_k} T(\omega) dP(\omega) = \varphi(P) \quad (34)$$

si  $P(U_k) = 0$ , on a, d'après (31) :

$$\int_{U_k} T(\omega) dP(\omega) = \varphi(P) = 0. \quad (35)$$

Donc, en tenant compte de (32), on peut écrire dans tous les cas :

$$\int_{U_k^c} T(\omega) dP(\omega) = 0 \quad (36)$$

On peut donc construire une suite  $(T_k)$  d'estimateurs de  $\varphi$  sans biais d'ordre 1 en posant :

$$T_k(\omega) = T(\omega) I_{U_k}(\omega) \quad (1) \quad \forall \omega \in \Omega \quad (37)$$

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

Quand  $k \rightarrow \infty$ ,  $T_k(\omega) \rightarrow T(\omega) I_F(\omega) = T_0(\omega) \quad \forall \omega \in \Omega$

d'autre part.  $|T_k(\omega)| \leq |T(\omega)|, \quad \forall \omega \in \Omega$

on peut donc appliquer le théorème de Lebesgue et écrire :

$$\int_{\Omega} T_k(\omega) dP(\omega) \rightarrow \int_{\Omega} T_0(\omega) dP(\omega)$$

or

$$\int_{\Omega} T_k(\omega) dP(\omega) = \varphi(P) \quad \forall P \in \mathcal{D}$$

(1)  $I_{U_k}(\omega)$  désigne l'indicateur de  $U_k$ .



20

et

$$\int_{\Omega} T_o(\omega) dP(\omega) = \int_F T_o(\omega) dP(\omega) = 0 \quad \forall P \in \mathcal{D}$$

d'où :

$$\varphi(P) = 0 \quad \forall P \in \mathcal{D}$$

ce qui est contradictoire.  $\varphi$  est donc bien de degré supérieur à 1.

**COROLLAIRE 1 :**

*Si  $\varphi$  vérifie les hypothèses du théorème 3, si  $\mathcal{D}$  est convexe et si  $\varphi[\alpha P + (1 - \alpha) Q]$  est un polynôme de degré 1 en  $\alpha$ , alors  $\varphi$  n'est pas estimable.*

C'est une conséquence des théorèmes 2 et 3.

**APPLICATION :**

On peut appliquer le corollaire 1 aux paramètres des exemples l) et m), on obtient donc les résultats suivants :

- 1) **La densité multivariée n'est pas estimable point par point.**
- 2) **Les dérivées de la densité ne sont pas estimables point par point.**

**REMARQUE :**

Le fait que la densité univariée n'admet pas d'estimateur sans biais a déjà été démontré par ROSENBLATT [7] et par BICKEL et LEHMANN [6]. Les résultats ci-dessus sont plus généraux et, nous semble-t-il, ont été obtenus d'une manière plus naturelle.

Démontrons, pour terminer, le

**THÉORÈME 4 :**

*Soit  $\varphi$  un paramètre à valeurs réelles, non constant,  $\mu$ -local et tel que*

1°)  $P \in \mathcal{D} \Rightarrow P$  est absolument continue par rapport à  $\mu$

2°) Si  $F$  est un fermé qui localise  $\varphi$ , si  $U$  est un voisinage ouvert de  $F$  et si  $P$  est tel que  $0 < P(U) < 1$  alors, la famille  $\mathcal{D}(U, P)$  de mesures de probabilité  $Q \in \mathcal{D}$  telles que  $Q_U = P_U$  est  $\mu$ -complète (1).

3°)  $P(U) = 0 \Rightarrow \varphi(P) = 0$  quel que soit  $U$  voisinage ouvert de  $F$ .

Dans ces conditions,  $\varphi$  est de degré supérieur à 1.

DÉMONSTRATION :

Comme pour le théorème 3, nous allons procéder par l'absurde :

Soit  $(U_k)$  une suite d'ouverts de  $\Omega$ , décroissante et tendant vers  $F$  et soit  $P \in \mathcal{D}$ . On aura toujours, au moins pour  $k$  assez grand,  $P(U_k) < 1$ . En effet, si l'on avait  $P(U_k) = 1, \forall k \in \mathbb{N}$ , on en déduirait  $P(F) = 1$ , ce qui est contradictoire avec le fait que  $P$  est absolument continue par rapport à  $\mu$  et que  $\mu(F) = 0$ ;

D'autre part, on sait que si  $P(U_k) = 0, \varphi(P) = 0$ ; il suffit donc de montrer qu'il en est de même si  $0 < P(U_k) < 1$  pour obtenir la contradiction cherchée.

Soit  $T$  un estimateur de  $\varphi$ , sans biais, d'ordre 1; soit  $Q \in \mathcal{D}(U_k, P)$ . On a les deux égalités :

$$\int_{U_k} T(\omega) dP(\omega) + \int_{U_k^c} T(\omega) dP(\omega) = \varphi(P) \quad (38)$$

et

$$\int_{U_k} T(\omega) dP(\omega) + \int_{U_k^c} T(\omega) dQ(\omega) = \varphi(Q) \quad (39)$$

(1) Autrement dit :  $\int_{\Omega} T(\omega) dQ(\omega) = 0, \forall Q \in \mathcal{D}(U, P) \Rightarrow T(\omega) = 0$   
 $\mu$ -p.p.

Comme  $\varphi$  est  $\mu$ -local, on en tire :

$$\int_{U_k^c} T(\omega) dQ(\omega) = \int_{U_k^c} T(\omega) dP(\omega) = \alpha(P) \quad \forall Q \in \mathcal{D}(U_k, P) \quad (40)$$

Posons  $T_1(\omega) = T(\omega) I_{U_k^c}(\omega)$ , il vient :

$$\int_{\Omega} [T_1(\omega) - \alpha(P)] dQ(\omega) = 0 \quad \forall Q \in \mathcal{D}(U_k, P) \quad (41)$$

d'où, comme  $\mathcal{D}(U_k, P)$  est  $\mu$ -complète :

$$T_1(\omega) - \alpha(P) = 0 \quad \mu \text{ p.p.}$$

Or, comme  $P(U_k) > 0$ , on a aussi  $\mu(U_k) > 0$ , d'où

$$T_1(\omega) - \alpha(P) = 0 \quad \mu \text{ p.p. sur } U_k$$

c'est-à-dire que  $\alpha(P) = 0$ .

On en conclut que,  $\forall P \in \mathcal{D}$  tel que  $0 < P(U_k) < 1$

on a

$$\int_{U_k^c} T(\omega) dP(\omega) = 0$$

Comme  $U_k^c \uparrow \Omega - F$ , on en tire

$$\varphi(P) = \int_{\Omega} T(\omega) dP(\omega) = \int_{\Omega - F} T(\omega) dP(\omega) = 0 \quad \text{CQFD}$$

**COROLLAIRE 2 :**

Si  $\varphi$  vérifie les hypothèses du théorème 4, si  $\mathcal{D}$  est convexe et si  $\varphi[\alpha P + (1 - \alpha) Q]$  est un polynôme de degré 1 en  $\alpha$ ,  $\varphi$  n'est pas estimable.

C'est une conséquence des théorèmes 2 et 4 .

### γ) CONCLUSION

On voit que, dans de nombreux problèmes d'estimation de fonctions, il n'est pas possible d'obtenir un estimateur sans biais, On peut alors essayer de remplacer la condition (3) par une condition moins forte. Généralisant et modifiant légèrement la notion de « unbiased touch estimate » proposée par TUKEY [8], on peut donner la définition suivante :

DEFINITION 6 bis :

Soit  $f(x)$  une fonction continue et soit  $T_n(\omega, x)$  un estimateur de  $f(x)$  .  
On dira que  $T_n$  est  $\epsilon$ -sans biais si

$$\forall x \in \mathbf{R}, \exists x' \in [x - \epsilon, x + \epsilon] \text{ tel que } E_p[T_n(\omega, x')] = f(x') .$$

De nombreux paramètres fonctionnels sont  $\epsilon$ -sans biais estimables. Par exemple

$$\frac{F_n(x + \epsilon) - F_n(x - \epsilon)^{(1)}}{\epsilon}$$

est un estimateur  $\epsilon$ -sans biais de la densité si celle-ci est continue.

## IV - ESTIMATION CONVERGENTE.

Ce paragraphe est consacré à l'étude de l'existence et de la construction d'un estimateur convergent.

Nous procéderons par étapes en commençant par construire un estimateur de  $P (P \in \mathcal{D})$ , puis en en déduisant un estimateur de  $\varphi(P)$  .

(1)  $F_n$  désigne ici la fonction de répartition empirique.

## 1. Estimation de P.

## a) Préliminaires.

Nous supposons dans la suite, sauf avis contraire, que  $\Omega$  est un espace topologique localement compact à base dénombrable.

Soit  $\mathfrak{M}$  l'espace des mesures de Radon sur  $\Omega$ , muni de la topologie vague. On sait qu'on peut identifier l'espace  $\mathcal{P}$  des mesures de probabilité sur  $\Omega$  à un sous-espace topologique de  $\mathfrak{M}^{(1)}$ . De plus,  $\mathcal{P}$  est métrisable et séparable pour la topologie induite par la topologie vague<sup>(2)</sup>.

Commençons par donner un résultat préliminaire connu mais qui nous sera utile dans la suite.

LEMME 3 :

Soit  $P_n = \frac{1}{n} \sum_{i=1}^n \delta(\omega_i)^{(3)}$  la mesure empirique associée à un échantillon de taille  $n$  tiré suivant  $P \in \mathcal{P}$ . Soit  $V_0$  un voisinage fondamental de l'origine dans  $\mathfrak{M}$  muni de la topologie vague. On a l'inégalité :

$$P^\infty [P_n - P \in V_0, \forall n \geq N] \geq 1 - \frac{C}{N} \quad \forall P \in \mathcal{P} \quad (43)$$

où  $C$  est une constante positive qui ne dépend que de  $V_0$ .

## DÉMONSTRATION.

On sait que  $V_0$  est de la forme :

$$V_0 = \{ \mu \in \mathfrak{M} \mid |\mu(\varphi_i)| < \varepsilon, \quad 1 \leq i \leq k \}$$

où  $\varphi_1, \dots, \varphi_k \in K(\Omega)$  (espace des fonctions continues à support compact définies sur  $\Omega$ ) et où  $\varepsilon$  est un nombre réel positif.

(1) cf. COURREGE [2] théorème 4 p. 119.

(2) cf. COURREGE [2] exercice 3 p. 129.

(3)  $\delta(\omega_i)$  désigne la mesure de Dirac au point  $\omega_i$ .

L'application  $\omega \rightarrow \delta(\omega)$  étant continue, on en déduit que l'événement  $\{P_n - P \in V_0, \forall n \geq N\}$  est mesurable.  $\forall P \in \mathcal{P}, \forall N \in \mathbb{N}$ .

Soit alors  $M$  un majorant commun aux  $|\varphi_i|$ . D'après l'inégalité de Kolmogoroff, on peut écrire<sup>(1)</sup> :

$$P^\infty \left[ \text{Max}_{2^h \leq n < 2^{h+1}} |P_n(\varphi_i) - P(\varphi_i)| > \varepsilon \right] \leq \frac{2^{h+1} \cdot M^2}{2^{2h} \cdot \varepsilon^2} \quad i = 1, \dots, k$$

d'où

$$P^\infty \left[ \text{Max}_{n \geq 2^h} |P_n(\varphi_i) - P(\varphi_i)| > \varepsilon \right] \leq \frac{M^2}{2^{h-2} \cdot \varepsilon^2} \quad i = 1, \dots, k$$

le résultat s'en déduit aisément en posant  $C = \frac{8 k M^2}{2}$

et en choisissant  $h$  tel que  $2^h \leq N < 2^{h+1}$ .

REMARQUE :

L'inégalité (43) implique la convergence presque sûre de  $P_n$  vers  $P$ , uniformément sur  $\mathcal{P}$ , c'est-à-dire que

$$\sup_{P \in \mathcal{P}} P^\infty \left[ \bigcup_{n > N} P_n - P \notin V_0 \right] \rightarrow 0 \quad \text{avec} \quad \frac{1}{N}$$

$\forall V_0$  voisinage fondamental de l'origine.

b) Construction d'un estimateur strict.

Soit  $d$  une distance sur  $\mathfrak{M}^+$  compatible avec la topologie vague.  $\mathcal{P}$  étant séparable, on peut se donner une suite de recouvrements de

(1) cf. DUGUE [9] p, 85.

$P_n(\Omega^n)$  par un nombre dénombrable de boules ouvertes de même rayon  $r_n$  tel que  $\lim_{n \rightarrow \infty} r_n = 0$ .

Désignons par  $(B_{in})_{i \in \mathbb{N}}$  une telle suite et posons

$$d(B_{in} \cap P_n(\Omega^n), \mathcal{D}) = d_{in} \quad \forall i, n \in \mathbb{N} \tag{44}$$

Soit alors  $(\gamma_n)$  une suite de nombres positifs telle que  $\lim_{n \rightarrow \infty} \gamma_n = 0$  et soit  $(P_{in})$  une suite double d'éléments de  $\mathcal{D}$  telle que :

$$d(B_{in} \cap P_n(\Omega^n), P_{in}) < d_{in} + \gamma_n \quad \forall i, n \in \mathbb{N} \tag{45}$$

On définira l'estimateur  $P_n^*$  de la façon suivante :

Soit  $(\omega_1, \dots, \omega_n) \in \Omega^n$ , il existe  $i \in \mathbb{N}$  tel que

$$P_n = \frac{1}{n} \sum_{i=1}^n \delta(\omega_i) \in B_{in}.$$

Parmi tous les  $B_{in}$  contenant  $P_n$  soit  $B_{jn}$  celui qui a le plus petit indice, on posera :

$$P_n^*(\omega_1, \dots, \omega_n) = P_{jn} \quad \forall (\omega_1, \dots, \omega_n) \in \Omega^n, \forall n \in \mathbb{N} \tag{46}$$

c) Convergence de  $P_n^*$ .

THÉORÈME 5 :

$P_n^*$  est un estimateur de  $P$  strict, mesurable, et qui converge presque sûrement, uniformément sur  $\mathcal{D}$ .

DÉMONSTRATION.

1°)  $P_n^*$  est strict puisque  $P_{in} \in \mathcal{D} \quad \forall i, n \in \mathbb{N}$ .

2°) Montrons que  $P_n^*$  est mesurable : pour cela il suffit de montrer que les ensembles

$$A_{jn} = \{ \omega \in \Omega^n \mid P_n^*(\omega) = P_{jn} \}$$

sont mesurables  $\forall j, n \in \mathbb{N}$ .

Or, si l'on pose

$$D_{jn} = \{ \omega \in \Omega^n \mid P_n \in B_{jn} \} \quad \forall j, n \in \mathbb{N}$$

on a, en tenant compte de (46) :

$$A_{jn} = D_{1n}^c \cap D_{2n}^c \cap \dots \cap D_{j-1n}^c \cap D_{jn} \quad (47)$$

et  $D_{jn}$  étant mesurable ( $\forall j, n \in \mathbb{N}$ ).  $A_{jn}$  l'est aussi. CQFD.

3°) Montrons enfin la convergence de  $P_n$ .

D'après (45), il existe  $Q_{jn} \in B_{jn}$   $P_n(\Omega^n)$  telle que :

$$d(Q_{jn}, P_{jn}) < d_{jn} + \gamma_n \quad \forall j, n \in \mathbb{N} \quad (48)$$

d'autre part

$$d_{jn} \leq d(P_n, P) \quad \forall j, n \in \mathbb{N} \quad (49)$$

Or

$$d(P_{jn}, P) \leq d(P_{jn}, Q_{jn}) + d(Q_{jn}, P_n) + d(P_n, P) \quad (50)$$

$$\forall j, n \in \mathbb{N}$$

Donc

$$d(P_{jn}, P) \leq \gamma_n + 2r_n + 2d(P_n, P) \quad \forall j, n \in \mathbb{N} \quad (51)$$

c'est-à-dire que,  $\forall \alpha > 0$ ,  $\forall \omega \in \Omega^n$ , et pour  $n$  assez grand :



$$d(P_n(\omega), P) < \frac{\alpha}{3} \Rightarrow d(P_n^*(\omega), P) < \alpha \quad (52)$$

compte tenu de (43), on en tire aisément :

$$\sup_{P \in \mathcal{D}} P^\infty \left[ \bigcup_{n \geq N} d(P_n^*, P) \geq \alpha \right] \leq \frac{C(\alpha)}{N} \quad (53)$$

$\forall \alpha > 0$  et pour  $N \geq n_0$  entier ne dépendant que de  $\alpha^{(1)}$ .

(53) entraîne évidemment la convergence presque sûre de  $P_n^*$  vers  $P$ , uniformément sur  $\mathcal{D}$ . CQFD.

Le théorème 5 peut se généraliser de la façon suivante :

**COROLLAIRE 3 :**

*Si  $\mathcal{P}$  est muni d'une topologie métrisable et séparable, si  $Q_n$  est un estimateur de  $P$ , mesurable et qui converge presque sûrement, uniformément sur  $\mathcal{P}$ , alors, il existe un estimateur de  $P$ , strict, mesurable, et qui converge presque sûrement, uniformément sur  $\mathcal{D}$ .*

**DÉMONSTRATION.**

Analogue à celle du théorème 5.

**Remarques.**

1. Le corollaire 3 reste valable pour un paramètre à valeurs dans un espace métrique séparable quelconque ; il reste encore valable dans un espace métrique non séparable à condition de supposer que  $Q_n(\Omega^n)$  est séparable (démonstrations analogues)<sup>(2)</sup>.

(1) Dans (53),  $C(\alpha)$  est la constante associée, dans (43), à un voisinage fondamental de l'origine contenu dans la boule de centre 0 et de rayon  $\alpha/3$ .

(2) JILOVEC [10] p. 436 a énoncé un résultat très voisin.

2. Tous les résultats de ce paragraphe restent valables pour  $\Omega$  non localement compact s'il est possible de munir  $\mathcal{P}$  d'une topologie métrisable telle que :

- l'application  $\omega \rightarrow \delta_{(\omega)}$  est mesurable ;
- $P_n(\Omega^n)$  est séparable ;
- $\forall P \in \mathcal{D}$ ,  $P_n$  tend vers  $P$  presque sûrement et cette convergence est uniforme sur  $\mathcal{D}$ .

*Exemple :* Les conditions précédentes sont vérifiées si  $\Omega$  étant un espace Polonais on munit  $\mathcal{P}$  de la topologie faible.

## 2. Estimation de $\varphi(P)$ .

### a) Préliminaires.

Commençons par indiquer deux lemmes très simples qui sont souvent utiles dans les problèmes d'estimation fonctionnelle.

LEMME 4 :

Soit  $(u_{nk})$  une suite double de nombres réels telle que :

$$\lim_{n \rightarrow \infty} u_{nk} = 0 \quad \forall k \in \mathbb{N}.$$

Alors il existe une suite d'entiers  $k(n)$  telle que

$$\lim_{n \rightarrow \infty} k(n) = +\infty \quad \text{et} \quad \lim_{n \rightarrow \infty} u_{nk(n)} = 0$$

DÉMONSTRATION.

Elémentaire.

LEMME 5 :

Si  $\Theta$  est un espace topologique quelconque et si  $\varphi$  est continue

$\varphi(P_n^*)^{(1)}$  est un estimateur de  $\varphi(P)$  strict, mesurable et qui converge presque sûrement, uniformément, sur  $\mathcal{D}$ .

DÉMONSTRATION :

Il est clair que  $\varphi(P_n^*)$  est strict et mesurable. La deuxième partie est une conséquence de l'implication

$$\bigcap_{n \geq N} \{ P_n^* \in \varphi^{-1}(U) \} \implies \bigcap_{n \geq N} \{ \varphi(P_n) \in U \}$$

$\forall U$  voisinage ouvert de  $\varphi(P)$ ,  $\forall P \in \mathcal{D}$  ; et de (53) .

Donnons quelques exemples d'application du lemme 5 .

Exemples :

1.  $\Omega = \mathbf{R}^p$ ,  $\mathcal{D} = \mathcal{P}$ ,  $\Theta = C(K, \mathbf{C})$  ( $K$  compact de  $\mathbf{R}^p$ ) muni de la topologie de la convergence uniforme.

$\varphi(P)$  = restriction à  $K$  de la fonction caractéristique de  $P$  . On voit, d'après le théorème de Paul Lévy, que le lemme 5 s'applique :

$$\sup_{t \in K} \left| \frac{1}{n} \sum_{j=1}^n e^{i \langle t, X_j \rangle} - \varphi(P)(t) \right| \xrightarrow[p.s.]{} 0 \quad (54)$$

uniformément sur  $\mathcal{P}^{(2)}$  .

2.  $\Omega = \mathbf{R}$ ,  $\mathcal{D} = \mathcal{P}$ ,  $\Theta$  = espace des fonctions de répartition à une variable, muni de la distance  $\delta$  de Paul Lévy.  $\varphi(P) = F$  fonction de répartition de  $P$  . Alors,  $F_n$  désignant la fonction de répartition empirique :

$$\delta(F_n, F) \xrightarrow[p.s.]{} 0 \text{ uniformément sur } \mathcal{P}^{(3)} \quad (55)$$

(1) ou  $\varphi(P_n)$  si  $\mathcal{D}$  contient les mesures de probabilités discrètes.

(2)  $(X_n) \in \mathbf{N}$  désigne une suite de variables aléatoires à valeurs dans  $\mathbf{R}^p$ , indépendantes et de même loi  $P$  .

(3) C'est aussi une conséquence du lemme 3 .

3. Exemple h du §III. En prenant  $\mathcal{D} = \mathcal{P}_c$  (espace des lois de probabilité sur  $\mathbf{R}$  à f.d.r. continue), on voit, à partir d'un théorème de Polya<sup>(1)</sup>, que  $\varphi$  est continue : le lemme 5 s'applique à la *fonction de concentration*.

**b) Construction d'un estimateur.**

Nous supposons maintenant que  $\varphi(P)$  prend ses valeurs dans un espace métrique et nous faisons sur  $\varphi$  l'hypothèse suivante :  $\varphi$  est une limite simple d'une suite  $(\varphi_k)$  d'applications uniformément continues de  $\mathcal{D}$  dans  $\Theta$ .

Soit  $(k(n))$  une suite d'entiers telle que  $\lim_{n \rightarrow \infty} k(n) = +\infty$  ; on définira une suite  $(T_n)$  d'estimateurs de  $\varphi(P)$

en posant :

$$T_n = \varphi_{k(n)}(P_n^*), \quad \forall n \in \mathbf{N} \quad (56)$$

Nous obtenons un estimateur très général qui est couramment utilisé dans les Problèmes d'estimation fonctionnelle.

**REMARQUE :**

Dans la pratique, on effectuera souvent un « lissage » préalable de  $P_n^*$  pour améliorer la précision de l'estimation.

**c) Convergence de  $T_n$ .**

**THEOREME 6 :**

Si  $\Theta$  est un espace métrique et si  $\varphi$  est limite simple d'une suite  $(\varphi_k)$  d'applications uniformément continues de  $\mathcal{D}$  dans  $\Theta$  ; alors,  $T_n$  est mesurable, et l'on peut choisir la suite  $k(n)$  de telle sorte que  $T_n$  converge en probabilité vers  $\varphi(P)$  ; de plus, si la convergence de  $(\varphi_k)$  est uniforme sur  $\mathcal{D}'\mathcal{C}\mathcal{D}$ , il en est de même pour la convergence de  $T_n$ .

(1) cf. DUGUE [9] p. 32.

DÉMONSTRATION.

Tout d'abord,  $T_n$  est évidemment mesurable comme composée de deux applications mesurables.

Montrons maintenant la convergence :

Comme  $\varphi_k$  est uniformément continue, on peut écrire :

$$\forall \varepsilon > 0, \quad \alpha(k, \varepsilon) > 0 \quad \text{tel que}$$

$$d(P, Q) < \alpha(k, \varepsilon) \implies d'[\varphi_k(P), \varphi_k(Q)] < \varepsilon \quad \forall P, Q \in \mathcal{D} \quad (57)$$

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

où  $d'$  est la distance définie sur  $\Theta$ .

De (57), on tire :

$$d(P_n^*, P) < \alpha(k, \frac{1}{k}) \implies d'[\varphi_k(P_n^*), \varphi_k(P)] < \frac{1}{k} \quad (58)$$

$$\forall P \in \mathcal{D}, \quad \forall \omega \in \Omega^n, \quad \forall k \in \mathbb{N}$$

d'où, en tenant compte de (53) :

$$\sup_{P \in \mathcal{D}} \text{Prob} \left[ d'(\varphi_k(P_n^*), \varphi_k(P)) \geq \frac{1}{k} \right] \leq \frac{C'(k)}{n} \quad (59)$$

où l'on a posé  $C'(k) = C \left[ \alpha(k, \frac{1}{k}) \right]$

en appliquant le lemme 4 à la suite  $\frac{C'(k)}{n}$ , on en déduit qu'il existe une suite d'entiers  $k(n)$  telle que

$$\lim_{n \rightarrow \infty} k(n) = +\infty \quad \text{et}$$

$$\lim_{n \rightarrow \infty} \sup_{P \in \mathcal{D}} \text{Prob} \left[ d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) \geq \frac{1}{k(n)} \right] = 0 \quad (60)$$

L'inégalité

$$d'(\varphi_{k(n)}(P_n^*), \varphi(P)) \leq d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) + \frac{d'(\varphi_{k(n)}(P), \varphi(P))}{k} \quad (61)$$

et le fait que  $\varphi_{k(n)}$  tend vers  $\varphi$  permettent de conclure.

La dernière partie du théorème est une conséquence immédiate de (60) et (61).

REMARQUE :

Si  $\Theta$  est séparable ou si  $\mathcal{D}$  est séparable,  $\forall K \in \mathbf{N}$  on en déduit, à partir d'un résultat de JILOVEC. ([10] p, 366) *l'existence d'un estimateur mesurable presque sûrement convergent*. De plus, compte tenu de la remarque du § 1, c), on peut imposer à cet estimateur d'être strict.

EXEMPLE :

$\Omega = \mathbf{R}$ ,  $\mathcal{D}$  = espace des mesures de probabilités sur  $\mathbf{R}$  absolument continues et à densité uniformément continue,  $\Theta = C(\mathbf{R})$  muni de la topologie de la convergence uniforme.  $\varphi(P) = f$  densité de  $P$

Soit  $F$  la f.d.r. de  $P$  et soit  $(h_k)$  une suite qui tend vers 0 avec  $\frac{1}{k}$ , posons :

$$f_k(x) = \frac{F(x+h_k) - F(x)}{h_k} \quad \forall x \in \mathbf{R}, \quad \forall k \in \mathbf{N}$$

On définit ainsi une suite  $\varphi_k$  d'applications de  $\mathcal{D}$  dans  $C(\mathbf{R})$ ; ces applications sont uniformément continues et  $\varphi_k$  tend vers  $\varphi$  dans  $C(\mathbf{R})$  quand  $k$  tend vers l'infini. Donc, le théorème 6 s'applique et l'on en déduit qu'il existe  $(k(n))$  telle que  $\lim_{n \rightarrow \infty} k(n) = +\infty$

et

$$\sup_{x \in \mathbf{R}} \left| f(x) - \frac{F_n^*(x + h_{k(n)}) - F_n^*(x)}{h_{k(n)}} \right| \xrightarrow{P} 0 \quad (62)$$

$F_n^*$  désignant la f.d.r. de  $P_n^*$ .

DISCUSSION.

LE CAM-LORRAINE SCHWARTZ [11] et JILOVEC [10] ont obtenu, pour  $\Omega$  quelconque, des théorèmes d'existence analogues au théorème 6. Les premiers ont traité le cas d'un espace métrique séparable, le second a traité le cas d'un espace de Hilbert.

Pour  $\Omega$  localement compact, la structure uniforme définie dans ces deux articles et la structure uniforme associée à la topologie vague ne sont pas comparables en général.

La méthode que nous employons ici a l'avantage d'être simple et de fournir, à la fois, une condition d'existence et une méthode de construction d'un estimateur convergent et mesurable.

3. Cas particulier important :  $\Omega = \mathbf{R}$ .

Nous allons voir que, pour  $\Omega = \mathbf{R}$ , on peut améliorer le théorème 6.

Munissons  $\mathcal{P}(\mathbf{R})$  de la distance de la convergence uniforme sur les fonctions de répartition. Pour des raisons de commodité, nous noterons encore  $d$  cette distance <sup>(1)</sup>

Il est bien connu que  $P_n$  est alors une application mesurable <sup>(2)</sup> de  $\mathbf{R}^n$  dans  $\mathcal{P}(\mathbf{R})$ ; de plus,  $P_n(\mathbf{R}^n)$  est séparable <sup>(3)</sup>.

Commençons par énoncer un lemme, analogue au lemme 3, mais plus précis :

(1) Bien qu'elle n'induisse pas la topologie vague sur  $\mathcal{P}(\mathbf{R})$ .

(2) Mais non continue.

(3) On peut recouvrir  $P_n(\mathbf{R}^n)$  avec les boules ouvertes centrées sur les  $P_n(\omega_1, \dots, \omega_n)$  correspondant aux points de coordonnées rationnelles, et de rayon  $2/n$ .

LEMME 6 :

(DVORETZKY - KIEFER - WOLFOWITZ - SCHUSTER) :

On a l'inégalité

$$P^\infty [d(P_n, P) > \varepsilon] \leq C e^{-2n\varepsilon^2}, \quad \forall P \in \mathcal{P} \quad \forall \varepsilon > 0 \quad (63)$$

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

où  $C$  est une constante universelle.

DÉMONSTRATION.

Voir DVORETZKY - KIEFER - WOLFOWITZ [13] et SCHUSTER [14].

On a alors le résultat suivant :

THÉORÈME 7 :

Si  $\Omega = \mathbf{R}$ , si  $\Theta$  est un espace métrique et si  $\varphi$  est limite simple d'une suite  $(\varphi_k)$  d'applications uniformément continues de  $\mathcal{D}$  dans  $\Theta$ , alors,  $T_n$  est mesurable, et l'on peut choisir la suite  $k(n)$  de telle sorte que  $T_n$  converge presque complètement sûrement<sup>(1)</sup> vers  $\varphi(P)$ .

DÉMONSTRATION.

La construction de  $P_n^*$  est identique à celle du §1 b) ; l'implication (52) reste valable, ce qui permet d'écrire, en tenant compte de (63) :

$$\sup_{P \in \mathcal{D}} \text{Prob} [d(P_n^*, P) > \alpha] \leq C e^{-\frac{2}{3} n \alpha^2}, \quad \forall n \in \mathbb{N}, \quad \forall \alpha > 0 \quad (64)$$

alors, (58) et (64) entraînent

(1) Donc presque sûrement.



$$\sup_{P \in \mathcal{F}} \text{Prob} [d'(\varphi_k(P_n^*), \varphi_k(P)) > \frac{1}{k}] \leq C e^{-\frac{2}{3} n \alpha(k, \frac{1}{k})^2} \quad (65)$$

$$\forall k, n \in \mathbb{N} .$$

D'après le lemme 4, on peut choisir  $(k(n))$  telle que  $\lim_{n \rightarrow \infty} k(n) = +\infty$  et

$$\lim_{n \rightarrow \infty} \frac{1}{[\sqrt{n} \alpha(k(n), \frac{1}{k(n)})]^2} = 0$$

On aura donc, pour  $n$  assez grand :

$$\sup_{P \in \mathcal{F}} \text{Prob} \left[ d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) > \frac{1}{k(n)} \right] \leq C e^{-\frac{2}{3} \cdot \frac{\sqrt{n}}{\delta}} \quad (66)$$

( $\delta > 0$  donné), on en déduit que

$$\sum_{n=1}^{\infty} \text{Prob} [d'(\varphi_{k(n)}(P_n^*), \varphi_{k(n)}(P)) > \varepsilon] < +\infty \quad (67)$$

$\forall P \in \mathcal{F}, \forall \varepsilon > 0$  ; d'où le résultat compte tenu de (61).

EXEMPLE :

Dans (62) on peut remplacer la convergence en probabilité par la convergence presque complètement sûre,  $k(n)$  étant convenablement choisi.

## V - ESTIMATION SYMÉTRIQUE.

Nous terminons ce chapitre en généralisant les résultats de HALMOS [5] sur les estimateurs symétriques sans biais.

## 1. Unicité.

Soit  $T_n$  un estimateur d'ordre  $n$  de  $\varphi(P)$ ; on sait que l'estimateur *symétrisé* de  $T_n$  est défini par :

$$T_n^{[n]} = S_n = \frac{1}{N} \sum T_n(\omega_{i_1}, \dots, \omega_{i_n}) \quad \forall (\omega_1, \dots, \omega_n) \in \Omega^n \quad (68)$$

où la somme porte sur le nombre de permutations distinctes des  $\omega_i$  et où  $N$  est le nombre de ces permutations.

Il est clair que, si  $T_n$  est sans biais,  $S_n$  est également sans biais : on peut donc toujours, quand on connaît un estimateur sans biais de  $\varphi(P)$  construire un estimateur *symétrique* sans biais de  $\varphi(P)$ . Nous allons voir que, sous des conditions assez générales, cet estimateur symétrique sans biais est unique.

Donnons, tout d'abord, deux définitions et un résultat préliminaire qui généralise des résultats de HALMOS et de FRASER :

## DÉFINITION 11 :

On dira que l'ensemble  $\mathcal{D}$  est de type  $F$  s'il contient les mesures de probabilité à support fini.

## DÉFINITION 11 bis :

Pour  $\Omega = \mathbf{R}$ , on dira que l'ensemble  $\mathcal{D}$  est de type  $C$  s'il contient les lois absolument continues à densité uniforme sur chaque intervalle d'une famille finie d'intervalles disjoints.

## LEMME 7 :

$\Omega$  étant localement compact et  $\Theta$  étant un E. V. T. localement convexe séparé soit  $t(\omega_1, \dots, \omega_n)$  une application symétrique,  $P^n$ -intégrable  $\forall P \in \mathcal{D}$  de  $\Omega^n$  dans  $\Theta$ , alors :

1° - Si  $\mathcal{D}$  est de type  $F$  :

$$E_P [t(\omega_1, \dots, \omega_n)] = 0, \quad \forall P \in \mathcal{D} \Rightarrow t(\omega_1, \dots, \omega_n) = 0 \\ \forall (\omega_1, \dots, \omega_n) \in \Omega^n. \quad (69)$$

2) - Si  $\mathcal{D}$  est de type C, et si  $t$  est continue :

$$E_P [t(\omega_1, \dots, \omega_n)] = 0, \quad \forall P \in \mathcal{D} \Rightarrow t(\omega_1, \dots, \omega_n) = 0 \quad (70) \\ \forall (\omega_1, \dots, \omega_n) \in \mathbf{R}^n.$$

### DÉMONSTRATION.

Remarquons que le lemme 1 de [5] reste valable pour un polynôme à coefficients dans un espace vectoriel. La démonstration de la première partie du lemme 7 est donc tout à fait analogue à celle du lemme 2 de [5].

Pour démontrer la deuxième partie, nous procéderons de la façon suivante : Soit  $\omega^0 = (\omega_1^0, \dots, \omega_n^0)$  un point de  $\mathbf{R}^n$  à coordonnées deux à deux distinctes et soit  $h > 0$  assez petit pour que les intervalles

$$I_j = [\omega_j^0 - h, \omega_j^0 + h], \quad j = 1, \dots, n$$

soient deux à deux disjoints. Soient  $p_1, \dots, p_n$   $n$  nombres positifs tels que :

$$\sum_{j=1}^n p_j = 1;$$

Enfin, soit  $P_h$  la mesure de probabilité à densité uniforme et égale à

$$\frac{p_j}{2h} \text{ sur } I_j, \quad j = 1, \dots, n. \text{ Alors :}$$

$$E_{P_j} [t(\omega_1, \dots, \omega_n)] = \sum_{j_1, \dots, j_n} \int_{I_{j_1} \times \dots \times I_{j_n}} t(\omega_1, \dots, \omega_n) \frac{p_{j_1} \dots p_{j_n}}{(2h)^n} d\omega_1 \dots d\omega_n$$

est un polynôme homogène par rapport aux variables  $p_1, \dots, p_n$  ; d'après l'hypothèse faite sur  $t$  et le lemme 1 de [5] il est identiquement nul. D'autre part, comme  $t$  est symétrique, le coefficient de  $p_1 \dots p_n$  s'écrit :

$$\frac{n!}{(2h)^n} \int_{I_1 \times \dots \times I_n} t(\omega_1, \dots, \omega_n) d\omega_1, \dots, d\omega_n$$

donc :

$$\int_{I_1 \times \dots \times I_n} t(\omega_1, \dots, \omega_n) d\omega_1 \dots d\omega_n = \vec{0}^{(1)} \quad (71)$$

Ce qui, d'après (12) implique :

$$\int_{I_1 \times \dots \times I_n} \langle t(\omega_1, \dots, \omega_n), \theta' \rangle dP_h^n = 0, \quad \forall \theta' \in \Theta' \quad (72)$$

Ce qui s'écrit encore :

$$\int_{I_1 \times \dots \times I_n} (\theta' \circ t)(\omega_1, \dots, \omega_n) dP_h^n(\omega_1, \dots, \omega_n) = 0 \quad \forall \theta' \in \Theta' \quad (73)$$

$t$  étant continue  $\theta' \circ t$  l'est également; donc, il existe  $\omega^h$  (dépendant de  $\theta'$ ),  $\in I_1 \times \dots \times I_n$  et tel que  $(\theta' \circ t)(\omega^h) = 0$ . Le raisonnement précédent étant valable  $\forall h > 0$  assez petit, on peut donc construire une suite  $(\omega^{h(k)})_{k \in \mathbb{N}}$  qui tend vers  $\omega^0$  et telle que  $(\theta' \circ t)(\omega^{h(k)}) = 0, \forall k \in \mathbb{N}$ , on en déduit :

$$(\theta' \circ t)(\omega^0) = 0 \quad \forall \theta' \in \Theta'$$

d'où l'on tire <sup>(2)</sup>

$$t(\omega^0) = \vec{0}.$$

(1)  $\vec{0}$  désignant ici l'élément neutre de  $\Theta$ .

(2) C'est une conséquence du théorème de Hahn-Banach. cf, BOURBAKI, [15] Chapitre IV, p. 48.

Donc  $t$  s'annule en tout point de  $\mathbf{R}^n$  à coordonnées deux à deux distinctes et comme  $t$  est continue, il vient :

$$t(\omega) \equiv \vec{0}, \quad \text{CQFD}^{(1)}.$$

Nous pouvons donner maintenant une généralisation du théorème 3 de [5].

#### THÉORÈME 8 :

Si  $\Theta$  est un E.V.T. localement convexe séparé, si  $\mathcal{D}$  est de type F (resp. de type C) et si  $T_n$  est un estimateur de  $\varphi(P)$  sans biais (resp. continu et sans biais) alors, le seul estimateur de  $\varphi(P)$  symétrique et sans biais (resp. symétrique, continue et sans biais) est  $T_n^{[n]}$ .

#### DÉMONSTRATION.

$S_n$  désignant un estimateur symétrique, sans biais (resp. symétrique, continu et sans biais) de  $\varphi(P)$ , on voit que  $S_n - T_n^{[n]}$  vérifie les conditions du lemme 7, 1°) (resp. du lemme 7, 2°), donc  $S_n = T_n^{[n]}$ ,

CQFD.

## 2. Optimalité.

Nous allons voir, maintenant, qu'en symétrisant un estimateur, on diminue le risque associé (la fonction de risque étant définie par la relation (6)), Le théorème ci-dessous est une généralisation du théorème 5 de [5].

#### THÉORÈME 9 :

Sous les conditions du théorème 8 et si  $\Theta$  est un espace de Hilbert réel, l'estimateur symétrique sans biais (resp. symétrique continu et sans biais) est uniformément optimal dans la famille des estimateurs sans

(1) Pour  $\Theta = \mathbf{R}$ , on peut s'affranchir de la condition de continuité. Cf, FRASER [16] p, 28-30,

sans biais (resp, continu et sans biais) pour la fonction de risque  $R(T_n, \varphi(P)) = E_P \|T_n - \varphi(P)\|^2$ .

DÉMONSTRATION.

Si  $\Theta$  est un Hilbert réel, on peut écrire :

$$\begin{aligned} E_P \|T_n - E_P(T_n)\|^2 &= E_P \langle T_n - E_P(T_n), T_n - E_P(T_n) \rangle \\ &= E_P \|T_n\|^2 - \|E_P(T_n)\|^2 \end{aligned} \quad (74)$$

Comparer les risques associés à deux estimateurs sans biais revient donc à comparer les moments d'ordre 2 de leurs normes.

Or, si  $T_n$  désigne un estimateur sans biais (resp, continu et sans biais) de  $\varphi(P)$  et si  $S_n$  désigne l'estimateur symétrique sans biais (resp, symétrique continu et sans biais) de  $\varphi(P)$ , on peut écrire, d'après (78) :

$$S_n(\omega_1, \dots, \omega_n) = \frac{1}{N} \sum T_n(\omega_{i_1}, \dots, \omega_{i_n}) = \frac{1}{N} \sum_{i=1}^N T_{ni} \quad (75)$$

$$\text{d'où } \|S_n\| \leq \frac{1}{N} \sum_{i=1}^N \|T_{ni}\| \quad (76)$$

$$\text{soit } \|S_n\|^2 \leq \left(\frac{1}{N} \sum_{i=1}^N \|T_{ni}\|\right)^2 \leq \frac{1}{N} \sum_{i=1}^N \|T_{ni}\|^2 \quad (77)$$

$$\text{d'où } E_P \|S_n\|^2 \leq \frac{1}{N} \sum_{i=1}^N E_P \|T_{ni}\|^2 = E_P \|T_n\|^2 \quad (78)$$

C. Q. F. D.

Exemple :

$\Omega = [0,1]$  ;  $\Theta = L^2[0,1]$ , la f.d.r. empirique  $F_n$  est un estimateur uniformément optimal dans la famille des estimateurs sans biais de la f.d.r. Un calcul facile montre que :

$$R(F_n, F) = E \int_0^1 [F_n(x) - F(x)]^2 dx = \frac{1}{n} \int_0^1 F(x) [1 - F(x)] dx \quad (79)$$

COROLLAIRE 4 :

Si  $\Theta$  est un espace de Hilbert réel et si  $T_n$  est un estimateur de  $\varphi(P)$ ,  $P^n$ -intégrable,  $\forall P \in \mathcal{D}$ , alors  $T_n^{[n]}$  est  $P^n$ -intégrable,  $\forall P \in \mathcal{D}$  et

$$R [T_n^{[n]}, \varphi(P)] \leq R [T_n, \varphi(P)], \quad \forall P \in \mathcal{D} \quad (80)$$

DÉMONSTRATION :

Analogue à celle du théorème 9.

REMARQUES :

1. Le corollaire 4 justifie le choix d'estimateurs symétriques par rapport aux observations ; il est clair que ce choix est naturel puisque le numérotage des observations est complètement arbitraire. Il semble intéressant de poser le problème de l'estimation d'un paramètre (à partir d'un échantillon de taille  $n$ ) d'une manière qui ne fasse pas intervenir le numérotage des observations : on peut, par exemple, utiliser la notion de répartition ponctuelle aléatoire<sup>(1)</sup>. Nous comptons revenir sur cette question.

2. On pourrait retrouver le théorème 9 à partir d'une généralisation des théorèmes de RAO-BLACKWELL et LEHMANN-SCHEFFÉ. L'examen des démonstrations de ces théorèmes (cf. FRASER [16] p. 57-58 et p. 61) montre en effet qu'on pourrait les généraliser au cas où  $\Theta$  est un E.V.T. localement convexe séparé.

(1) cf. LANNUZEL [17].

## II

## ESTIMATION DE LA DENSITÉ DE PROBABILITÉ.

## 1 - GÉNÉRALITÉS.

## 1. La méthode du noyau.

Le problème de l'estimation de la densité de probabilité a été l'objet de nombreux travaux au cours des dernières années.

Le premier article important sur cette question semble être celui de ROSENBLATT [7] (1956) ; dans cet article, l'auteur propose une méthode d'estimation qui a été très étudiée par la suite : *la méthode du noyau*. Cette méthode consiste à prendre comme estimateur de la densité la fonction aléatoire  $\hat{f}_n$  définie de la façon suivante :

$$\hat{f}_n(\omega; \omega_1, \dots, \omega_n) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{\omega - \omega_i}{h}\right), \quad \forall (\omega, \omega_1, \dots, \omega_n) \in \Omega^{n+1} \quad (1)$$

où le noyau  $K$  est souvent une densité possédant certaines propriétés de régularité et  $h$  un nombre positif qui dépend de  $n$ .

Cet estimateur a été étudié notamment par PARZEN [29] (1962), NADARAYA [18] (1965) et WOODROOFE [19] (1967) pour  $\Omega = \mathbf{R}$  ; par CACOULOS [20] (1966) et VAN RYZIN [21] (1969) pour  $\Omega = \mathbf{R}^p$  et par CRASWELL [22] (1965) dans un groupe topologique.

LEADBETTER et WATSON [23] (1963) ont proposé une méthode un peu plus générale ; ils considèrent un estimateur de la forme :

$$\hat{f}_n(\omega; \omega_1, \dots, \omega_n) = \frac{1}{n} \sum_{i=1}^n K_n(\omega - \omega_i), \quad \forall (\omega, \omega_1, \dots, \omega_n) \in \Omega^{n+1} \quad (2)$$

leurs résultats ont été complétés récemment par PICKANDS III [24] (1969).

La Méthode du noyau a certaines propriétés intéressantes (notamment la rapidité de convergence et le fait que l'estimateur est strict), mais



elle a, pour un praticien, certains inconvénients (complication et manque de robustesse des estimateurs optimaux).

## 2. La méthode des fonctions orthogonales.

La méthode des fonctions orthogonales, que nous étudierons dans ce chapitre, a été introduite par CENCOV [25] (1962)<sup>(1)</sup>. Le principe de cette méthode est le suivant :

Soit  $\mu$  une mesure sur  $(\Omega, \mathcal{A})$  et soit  $\mathcal{D}$  l'ensemble des mesures de probabilité sur  $(\Omega, \mathcal{A})$  admettant par rapport à  $\mu$ , une densité de Radon-Nikodym appartenant à  $L^2(\mu)$ .

On sait que, muni de la norme  $\int_{\Omega} |f|^2 d\mu$ <sup>1/2</sup>,  $L^2(\mu)$  est un espace de Hilbert ; soit alors  $(e_i)_{i \in I}$

une base orthonormale de  $L^2(\mu)$  et soit  $(e_{i_k, n})_{1 \leq k \leq q(n)}$

un sous-ensemble fini de  $(e_i)$  ; une version  $\hat{f}_n^*$  de l'estimateur  $\hat{f}_n$  de  $f = \frac{dP}{d\mu}$  sera défini par la relation :

$$\hat{f}_n^*(\omega; \omega_1, \dots, \omega_n) = \sum_{k=1}^{q(n)} \frac{1}{n} \left[ e_{i_k, n}^0(\omega_1) + \dots + e_{i_k, n}^0(\omega_n) \right] e_{i_k, n}^*(\omega),$$

$$\forall (\omega, \omega_1, \dots, \omega_n) \in \Omega^{n+1} \quad (3)$$

où les  $e_{i_k, n}^*$  ( $k = 1, \dots, q(n)$ ) sont des versions des

$e_{i_k, n}$  ( $k = 1, \dots, q(n)$ ) et les  $e_{i_k, n}^0$  des versions fixées des  $e_{i_k, n}$ .

(1) Et, indépendamment, par TIAGO DE OLIVEIRA [26].

On voit que cette méthode consiste à estimer la projection de  $\frac{dP}{d\mu}$  sur un sous-espace de  $L^2(\mu)$  de dimension finie : elle ramène l'estimation de  $\frac{dP}{d\mu}$  à celle d'un nombre fini de paramètres numériques.

Remarquons, d'autre part, que le principe de cette méthode peut s'appliquer à l'estimation de n'importe quel paramètre à valeurs dans un espace de Hilbert.

Depuis 1962 l'estimateur défini par (3) a été peu étudié ; cependant, deux cas particuliers importants ont été abordés : S. C. SCHWARTZ [27] (1967) a pris, comme système orthonormal dans  $L^2$ , les fonctions d'Hermite alors que KRONMAL et TARTER [28] (1968) s'intéressaient au cas des fonctions trigonométriques.

Dans la suite, nous généralisons leurs résultats et nous donnons, dans le cas général, des résultats qui, à notre connaissance, sont nouveaux (en particulier ceux de notre note aux CR du 10 mars 1969 [32] ).

## II - CONVERGENCE EN MOYENNE QUADRATIQUE INTÉGRÉE.

Les conditions de convergence du MISE (défini au § II du chapitre I) sont simples et connues<sup>(1)</sup>, nous les énoncerons sous la forme suivante:

### THÉORÈME 10 :

Si les séquences  $(e_{i_k, n})_{1 \leq k \leq q(n)}$  sont telles que la suite

$$\sum_{k=1}^{q(n)} \left[ \int_{\Omega} e_{i_k, n}(\omega) f(\omega) d\mu(\omega) \right]^2 \text{ converge vers}$$

(1) Car on peut considérer qu'elles sont contenues, implicitement, dans les théorèmes 1 a et 1 b de Cencov [25].

$\int_{\Omega} f(\omega)^2 d\mu(\omega)^{(1)}$ , une condition nécessaire et suffisante pour que  $\hat{f}_n$  converge en moyenne quadratique intégrée vers  $f$  est que

$$\lim_{n \rightarrow \infty} \int_{\Omega} \frac{1}{n} \left[ \sum_{k=1}^{q(n)} e_{i_{k,n}}^2(\omega) \right] f(\omega) d\mu(\omega) = 0 \quad (4)$$

DÉMONSTRATION

Pour simplifier les notations, nous poserons

$$a_i = \int_{\Omega} e_i(\omega) f(\omega) d\mu(\omega) \quad \forall i \in I$$

$$a_{\hat{i}_{k,m}} = \frac{1}{n} \sum_{j=1}^n e_{i_{k,n}}^o(\omega_j) \quad \forall k = 1, \dots, q(n); \quad \forall n \in \mathbb{N}.$$

et

$$\sum_{k=1}^{q(n)} a_{i_{k,n}} \cdot e_{i_{k,n}} = f_n \quad \forall n \in \mathbb{N}$$

alors puisque  $\hat{f}_n - f_n$  et  $f_n - f$  sont orthogonaux, le MISE peut s'écrire :

$$J_n = E \int_{\Omega} [ \hat{f}_n(\omega) - f_n(\omega) ]^2 d\mu(\omega) + \int_{\Omega} [ f_n(\omega) - f(\omega) ]^2 d\mu(\omega) \quad (5)$$

soit

$$J_n = \frac{1}{n} \sum_{k=1}^{q(n)} \left[ \int_{\Omega} e_{i_{k,n}}^2(\omega) f(\omega) d\mu(\omega) - a_{i_{k,n}}^2 \right] + \sum_{j \in I_n} a_j^2 \quad (6)$$

où l'on a posé  $I_n = I - \{ i_{1,n}, \dots, i_{q(n),n} \}$

ou encore :

(1) Dans la suite, nous appellerons cette condition "condition S".

$$J_n = \int_{\Omega} \frac{1}{n} \left[ \sum_{k=1}^{q(n)} e^2_{i_{k,n}}(\omega) |f(\omega)| - \frac{n+1}{n} \sum_{k=1}^{q(n)} a^2_{i_{k,n}} \right] + \sum_{j \in I} a_j^2 \quad (7)$$

Or l'hypothèse faite sur les séquences  $(e_{i_{k,n}})$  implique que

$$\lim_{n \rightarrow \infty} \frac{n+1}{n} \sum_{k=1}^{q(n)} a^2_{i_{k,n}} = \sum_{j \in I} a_j^2 \quad (8)$$

donc, d'après (7) et (8),  $J_n \rightarrow 0$  si et seulement si on a (4). CQFD.

La condition (4) dépend étroitement de  $f$ , cependant nous allons voir que, dans de nombreux cas particuliers importants, cette dépendance n'est qu'apparente :

COROLLAIRE 5 :

1°) - Si l'une des conditions suivantes est remplie :

a) Il existe une version des  $e_i$  uniformément bornée pour  $i \in I$   
et  $\omega \in \Omega$ .

b) Il existe  $g \in L^4(\mu)$  qui majore  $|e_i|$  pour tout  $i \in I$ .

c) Il existe une version de  $f$  uniformément bornée pour  $\omega \in \Omega$ .

Alors, si la condition S est vérifiée,

$$\lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0 \quad \text{entraîne} \quad \lim_{n \rightarrow \infty} J_n = 0.$$

2°) - Si il existe une version de  $f$  uniformément minorée par un nombre positif  $m$  pour  $\omega \in \Omega$  et si la condition S est vérifiée, alors :

$$\lim_{n \rightarrow \infty} J_n = 0 \quad \text{entraîne} \quad \lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0.$$

DÉMONSTRATION.

Les résultats proviennent des implications suivantes :

$$1^{\circ) \text{ a)} \Rightarrow \sup_{i \in I} \int_{\Omega} e_i^2(\omega) f(\omega) \, d\mu(\omega) = O(1)$$

$$\Rightarrow \int_{\Omega} \frac{1}{n} \left[ \sum_{k=1}^{q(n)} e_{i_{k,n}}^2(\omega) \right] f(\omega) \, d\mu(\omega) = O\left(\frac{q(n)}{n}\right)$$

$$b) \Rightarrow \sup_{i \in I} \int_{\Omega} e_i^2(\omega) f(\omega) \, d\mu(\omega) \leq \int_{\Omega} g^2(\omega) f(\omega) \, d\mu(\omega)$$

$$\leq \left[ \int_{\Omega} g^4(\omega) \, d\mu(\omega) \right]^{1/2} \cdot \left[ \int_{\Omega} f^2(\omega) \, d\mu(\omega) \right]^{1/2}$$

$$c) \Rightarrow \sup_{i \in I} \int_{\Omega} e_i^2(\omega) f(\omega) \, d\mu(\omega) = O(1)$$

$$2^{\circ) \text{ inf}_{\omega \in \Omega} f^*(\omega) \geq m \Rightarrow \int_{\Omega} \frac{1}{n} \left[ \sum_{k=1}^{q(n)} e_{i_{k,n}}^2(\omega) \right] f(\omega) \, d\mu(\omega) \geq m \cdot \frac{q(n)}{n}$$

COROLLAIRE 6 :

On suppose que  $\Omega$  est un intervalle compact de  $\mathbf{R}$ , que  $\mu$  est la mesure de Lebesgue sur  $\Omega$  et que les  $e_i$  ( $i \in \mathbf{N}$ ) sont les fonctions trigonométriques sur  $\Omega$ , alors, si

$$e_{i_{k,n}} = e_k, \quad \forall k \in \mathbf{N}, \quad \forall n \in \mathbf{N}$$

et si  $\lim_{n \rightarrow \infty} q(n) = +\infty$ ; pour que  $\lim_{n \rightarrow \infty} J_n = 0$ ,

il faut et il suffit que  $\lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0$ .

DÉMONSTRATION.

On peut, sans perte de généralité, supposer que  $\Omega = [0, 2\pi]$ ; on aura alors :

$$e_1^*(\omega) = \frac{1}{\sqrt{2\pi}}, \quad e_2^*(\omega) = \frac{\cos \omega}{\sqrt{\pi}}, \quad e_3^*(\omega) = \frac{\sin \omega}{\sqrt{\pi}},$$

$$e_4^*(\omega) = \frac{\cos 2\omega}{\sqrt{\pi}} \quad \text{etc ...}$$

Un calcul élémentaire montre que :

$$\frac{q(n) - 1}{2\pi} \leq \sum_{k=1}^{q(n)} e_k^{*2}(\omega) \leq \frac{q(n) + 1}{2\pi} \left[ \forall q(n) \in \mathbb{N}, \forall \omega \in \Omega \right] \quad (9)$$

Or, comme  $\lim_{n \rightarrow \infty} q(n) = +\infty$ , la condition S est vérifiée, donc (4) est une condition nécessaire et suffisante de convergence, et comme, d'après (9), (4) est équivalent à  $\lim_{n \rightarrow \infty} \frac{q(n)}{n} = 0$ , le corollaire est démontré.

REMARQUE :

Le corollaire 6 améliore la deuxième partie du théorème 3 de [28] où l'on montre que  $\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} = 0$  est une condition suffisante pour que

$$\lim_{n \rightarrow \infty} J_n = 0.$$

**Optimalité et vitesse de convergence.**

La relation (7) montre que  $J_n$  dépend fortement des coefficients  $a_i$ ; il n'est donc pas possible, en général, d'obtenir un estimateur uniformément optimal (pour la fonction de risque  $J_n$ ) dans la famille des estimateurs définis par (3); c'est ce que confirme le

LEMME 8 :

On suppose que la famille  $e_i$  est dénombrable, que  $e_{i_k, n} = e_k, \forall k \in \mathbb{N}, \forall n \in \mathbb{N}$  et qu'il existe deux constantes  $m$  et  $M$  telles que

$$0 < m \leq \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) \leq M, \forall i \in \mathbb{N} \quad (10)$$

Dans ces conditions et  $\forall n \in \mathbb{N}$ , il existe un entier  $q_0(n)$  qui minimise  $J_n$ ; de plus, si  $|a_i|$  est fonction décroissante de  $i$ ,  $q_0(n)$  est tel que,

au moins pour  $n$  assez grand :

$$|a_{q_0(n)}| \geq \frac{\sqrt{m}}{\sqrt{n+1}} \text{ et } |a_{q_0(n)+1}| \leq \frac{\sqrt{m}}{\sqrt{n+1}} \quad (11)$$

DÉMONSTRATION.

Désignons par  $\Delta_q J_n$  l'accroissement de  $J_n$  quand  $q(n)$  passe de  $q$  à  $q+1$  ; (7) entraîne alors :

$$\Delta_q J_n = \frac{1}{n} \left[ \int_{\Omega} e^{2q+1}(\omega) f(\omega) d\mu(\omega) - (n+1) a^2_{q+1} \right] \quad (12)$$

d'où, en tenant compte de (10) :

$$\frac{1}{n} [m(n+1) a^2_{q+1}] \leq \Delta_q J_n \leq \frac{1}{n} [M(n+1) a^2_{q+1}] \quad (13)$$

Donc,  $\forall n \in \mathbb{N}$ , il existe un entier  $q_1(n)$  tel que

$$q \geq q_1(n) \implies |a_{q+1}| < \frac{\sqrt{m}}{\sqrt{n+1}} \implies \Delta_q J_n > 0 \quad (14)$$

c'est-à-dire que, pour  $q$  assez grand,  $J_n$  est fonction croissante de  $q$  : la première partie du lemme est démontrée. Remarquons que l'on a aussi montré que  $q_0(n) \leq q_1(n)$ .

Supposons maintenant que  $|a_q|$  soit fonction décroissante de  $q$ , alors si  $q_1(n)$  est le plus petit entier tel que l'on ait (14), on en déduit

que  $|a_{q_0(n)}| \geq \frac{\sqrt{m}}{\sqrt{n+1}}$ , au moins pour  $n$  assez grand.

Soit  $q_2(n)$  le plus grand entier tel que  $|a_{q+1}| > \frac{\sqrt{m}}{\sqrt{n+1}}$  ( $q_2(n)$  existe au moins pour  $n$  assez grand), d'après (13), on a l'implication :

$$q \leq q_2(n) \implies \Delta_q J_n < 0$$

On en déduit  $q_0(n) > q_2(n)+1$ ; d'où, d'après l'hypothèse faite sur  $q_2(n)$ ,

$|a_{q_0(n)+1}| \leq \frac{\sqrt{M}}{\sqrt{n+1}}$ , ce qui achève la démonstration.

Le lemme 8 montre qu'il faut connaître l'ordre de grandeur des coefficients  $a_q$  pour connaître celui de  $J_n$ .

On peut, cependant, obtenir des résultats partiels : par exemple si

$|a_k| = O\left(\frac{1}{k^r}\right)$ ,  $r > \frac{1}{2}$ , et si l'on prend  $q(n) = O(n^{1/2r})$  on voit

facilement que  $J_n = O\left(\frac{1}{n^{2r-1/2r}}\right)$ ; si, de plus,  $|a_k|$  est de l'ordre

de  $\frac{1}{k^r}$ , on obtient ainsi un estimateur asymptotiquement optimal au sens suivant :  $J_n$  est asymptotiquement du même ordre que le MISE associé à l'estimateur optimal en  $f$ .

Remarquons enfin que l'on peut évaluer empiriquement  $q_0(n)$  en remplaçant  $J_n$  par une estimation  $\hat{J}_n$  et en étudiant le signe de  $\Delta_q \hat{J}_n$ . Cette méthode a été employée par KRONMAL et TARTER [28].

### III. CONVERGENCE UNI FORME EN MOYENNE QUADRATIQUE ET CONVERGENCE UNIFORME PRESQUE SURE.

#### 1. Introduction.

Dans les problèmes d'estimation de la densité, on cherche souvent à estimer une version spécifiée de la densité; il s'agira, en général, d'une version *continue* de la densité. On peut alors étudier des modes de convergence plus précis que celui du paragraphe précédent.

Rappelons d'abord à quelle condition une version continue de la densité est unique :

LEMME 9 :

Soit  $\Omega$  un espace topologique muni de sa tribu Borélienne, soit  $\mu$  une mesure sur  $\Omega$  qui charge tout ouvert non vide et soit  $P$  une mesure de probabilité sur  $\Omega$  admettant, par rapport à  $\mu$ , une densité de Radon-Nikodym  $\frac{dP}{d\mu}$ . Alors, s'il existe une version continue de  $\frac{dP}{d\mu}$ , cette version est unique.



DÉMONSTRATION.

Soient  $f$  et  $g$  deux versions continues de  $\frac{dP}{d\mu}$  et soit  $\omega_0 \in \Omega$  tel que  $f(\omega_0) - g(\omega_0) > 0$ , alors, il existe un voisinage ouvert  $U$  de  $\omega_0$  tel que l'on ait simultanément :

$$\int_U \frac{dP}{d\mu}(\omega) d\mu(\omega) = \int_U f(\omega) d\mu(\omega) = \int_U g(\omega) d\mu(\omega)$$

et

$$\int_U [f(\omega) - g(\omega)] d\mu(\omega) > 0$$

ce qui est évidemment contradictoire. QQFD.

EXEMPLE :

Les conditions du lemme 8 sont vérifiées si  $\Omega$  est un groupe topologique localement compact et si  $\mu$  est une mesure de Haar sur  $\Omega$ .

REMARQUES :

1.- Si  $\mu$  s'annule sur un ouvert non vide de  $\Omega$ , s'il existe une version continue de  $\frac{dP}{d\mu}$  et si  $\Omega$  est normal, il existe, en général, une infinité de versions continues de  $\frac{dP}{d\mu}$ .

2.- Dans la suite, si les conditions du lemme 9 sont vérifiées, nous disons "  $P$  admet une densité continue " au lieu de " il existe une version continue de  $\frac{dP}{d\mu}$  " ; de même, nous dirons "  $e_1$  est continue " au lieu de " il existe une version continue de  $e_1$  " (1).

**2. Convergence uniforme en moyenne quadratique.**

THÉORÈME 11 :

Sous les conditions suivantes :

(1) Car le lemme 9 s'applique à  $e_1^2$  donc à  $e_1$ .

1°)  $f$  est continue

2°)  $e_i$  est continue et bornée,  $\forall i \in I$

3°) La famille  $\sum_{i \in I} a_i e_i$  est uniformément sommable de somme  $f$ .

4°) Les séquences  $(e_{i_k, n})_{1 \leq k \leq q(n)}$  sont telles que la suite

$$f_n = \sum_{k=1}^{q(n)} a_{i_k, n} e_{i_k, n} \text{ converge uniformément vers } f \quad (1)$$

$$\text{Si } \lim_{n \rightarrow \infty} \left[ \sup_{1 \leq k \leq q(n)} \sup_{\omega \in \Omega} |e_{i_k, n}(\omega)| \right]^4 \cdot \frac{q(n)^2}{n} = 0 \quad (15)$$

$$\text{Alors } Q_n = \sup_{\omega \in \Omega} E [ |\hat{f}_n(\omega) - f(\omega)|^2 ] \rightarrow 0 \text{ quand } n \rightarrow +\infty \quad (16)$$

DÉMONSTRATION.

$$\text{Posons pour simplifier } \sup_{1 \leq k \leq q(n)} \sup_{\omega \in \Omega} |e_{i_k, n}(\omega)| = M_n$$

D'après la relation

$$E [ |\hat{f}_n(\omega) - f(\omega)|^2 ] = E [ |\hat{f}_n(\omega) - f_n(\omega)|^2 ] + |f_n(\omega) - f(\omega)|^2$$

et la condition S', il suffit de montrer que

$$\lim_{n \rightarrow \infty} \sup_{\omega \in \Omega} E [ |\hat{f}_n(\omega) - f_n(\omega)|^2 ] = 0$$

$$\text{Or } \hat{f}_n(\omega) - f_n(\omega) = \sum_{k=1}^{q(n)} (\hat{a}_{i_k, n} - a_{i_k, n}) e_{i_k, n}(\omega)$$

$$\text{D'où } E [ |\hat{f}_n(\omega) - f_n(\omega)|^2 ] = \sum_{k=1}^{q(n)} \text{Var} (\hat{a}_{i_k, n}) e_{i_k, n}^2(\omega)$$

$$+ \sum_{k \neq k'} \text{Cov} (\hat{a}_{i_k, n}; \hat{a}_{i_{k'}, n}) e_{i_k, n}(\omega) e_{i_{k'}, n}(\omega)$$

On en tire

$$E [ |\hat{f}_n(\omega) - f_n(\omega)|^2 ] \leq M_n^2 \left[ \sum_{k=1}^{q(n)} \text{Var} (\hat{a}_{i_k, n}) + \sum_{k \neq k'} |\text{Cov} (\hat{a}_{i_k, n}; \hat{a}_{i_{k'}, n})| \right]$$

(1) Dans la suite, nous appellerons cette condition "condition S'".

d'où :

$$E [ |\hat{f}_n(\omega) - f_n(\omega)|^2 ] \leq M_n^2 \left[ \sum_{k=1}^{q(n)} \text{Var}(\hat{a}_{i_k, n}) + \sum_{k \neq k'} \sqrt{\text{Var}(\hat{a}_{i_k, n}) \cdot \text{Var}(\hat{a}_{i_{k'}, n})} \right]$$

soit, en posant  $\sigma^2(e_i) = \text{Var} e_i$ ,  $i \in I$  :

$$E [ |\hat{f}_n(\omega) - f_n(\omega)|^2 ] \leq M_n^2 \cdot \left[ \sum_{k=1}^{q(n)} \frac{\sigma(e_{i_k, n})}{\sqrt{n}} \right]^2 \leq M_n^4 \cdot \frac{q(n)^2}{n}$$

et la condition (10) permet de conclure.

REMARQUES :

1.- Si les  $e_i$  sont bornés, uniformément pour  $i \in I$ , la condition (15) s'écrit simplement

$$\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} = 0 \tag{15 bis}$$

condition qui a été obtenue pour les fonctions d'Hermite dans [27] et pour les fonctions trigonométriques dans [28].

2.- Sous les mêmes hypothèses, et d'une manière tout à fait analogue, on peut montrer que (15) entraîne

$$\lim_{n \rightarrow \infty} E [ (\sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)|)^2 ] = 0 \tag{17}$$

et l'on obtient une convergence plus forte que (16). Si  $f$  est bornée, (17) peut s'écrire

$$\lim_{n \rightarrow \infty} R(\hat{f}_n, f) = 0$$

$R$  étant la fonction de risque définie au chapitre I quand  $\Theta$  est l'espace des fonctions continues et bornées sur  $\Omega$ , muni de la norme de la convergence uniforme.

Vitesse de convergence :

La vitesse de convergence de  $Q_n$  vers 0 dépend de celle de la famille sommable  $\sum_{i \in I} a_i e_i$  vers  $f$ .

Pour mettre en évidence cette dépendance, on peut donner, par exemple, le résultat suivant <sup>(1)</sup> :

COROLLAIRE 7 :

On suppose que la famille  $(e_i)$  est dénombrable, que les  $e_i$  sont bornés uniformément pour  $i \in \mathbb{N}$ , et que  $e_{i_k} = e_k$ ,  $\forall k \in \mathbb{N}$ ,  $\forall n \in \mathbb{N}$ .

Alors, si  $|a_k| \ll \frac{\alpha}{k^r}$  avec  $r > 1$  et  $\alpha > 0$ ,  $\forall k \in \mathbb{N}$

et si  $q(n) = O(n^{1/2r})$ , on a  $Q_n = O\left(\frac{1}{n^{\frac{r-1}{r}}}\right)$

DÉMONSTRATION.

C'est une conséquence immédiate de l'inégalité

$$Q_n \leq M^4 \left[ \frac{q(n)^2}{n} + \left( \sum_{k > q(n)} |a_k| \right)^2 \right]$$

où  $M = \text{Max} \left( \sup_{\omega \in \Omega, i \in \mathbb{N}} |e_i(\omega)|, 1 \right)$ .

### 3. Convergence uniforme presque sûre.

Nous allons maintenant donner des conditions pour que

$\hat{d}_n = \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)|$  tende presque sûrement vers 0 quand  $n$  tend vers l'infini.

Bien que la définition 5 du chapitre I ne fasse pas intervenir la mesurabilité de l'estimateur, il semble intéressant d'indiquer une condition suffisante pour que  $\hat{d}_n$  soit une variable aléatoire.

(1) Extension de la deuxième partie du théorème 2 de S.C. SCHWARTZ [27].

LEMME 10 :

Soit  $\Omega$  un espace mesurable, soit  $X$  un espace topologique localement compact à base dénombrable, et soit  $T(x, \omega)$  une application de  $X \times \Omega$  dans  $\mathbf{R}^+$ . Alors, si  $\forall \omega_0 \in \Omega$ ,  $T(x, \omega_0)$  est continue et  $\forall x_0 \in X$ ,

$T(x_0, \omega)$  est mesurable, l'application  $h(\omega) = \sup_{x \in X} T(x, \omega)$  est mesurable.

DÉMONSTRATION.

Soit  $K$  un compact de  $X$  et soit  $d$  une distance sur  $X$  compatible avec sa topologie.

$\forall \omega \in \Omega$ ,  $T(x, \omega)$  est uniformément continue sur  $K$ , il existe donc  $\alpha_n(\omega) > 0$  tel que

$$d(x, x') < \alpha_n(\omega) \implies |T(x, \omega) - T(x', \omega)| < \frac{1}{n} \quad \forall x, x' \in K; \quad \forall \omega \in \Omega.$$

Soit  $\{B(x_i, \alpha_n(\omega))\}$ ,  $i = 1, 2, \dots, \lambda_n$  un recouvrement de  $K$  par un nombre fini de Boules ouvertes de rayon  $\alpha_n(\omega)$ .

L'application  $h_{\lambda_n}(\omega) = \sup_{1 \leq i \leq \lambda_n} T(x_i, \omega)$  est mesurable.

Posons  $h_K(\omega) = \sup_{x \in K} T(x, \omega)$ ; on aura

$$h_{\lambda_n}(\omega) \leq h_K(\omega) \leq h_{\lambda_n}(\omega) + \frac{1}{n} \quad \forall \omega \in \Omega$$

donc  $\lim_{n \rightarrow \infty} h_{\lambda_n}(\omega) = h_K(\omega)$ ,  $\forall \omega \in \Omega$  ce qui entraîne que  $h_K$  est mesurable.

Enfin, soit  $(K_p)_{p \in \mathbf{N}}$  une suite croissante de compacts qui recouvre  $X$ ; on peut écrire :

$$h(\omega) = \sup_{p \in \mathbf{N}} h_{K_p}(\omega), \quad \forall \omega \in \Omega$$

donc  $h(\omega)$  est mesurable.

CONSÉQUENCE :

Si  $\Omega$  est un espace topologique localement compact à base dénombrable, si  $f$  est continue et si  $e_i$  est continue,  $\forall i \in I$ ,  $d_n$  est mesurable.

Passons maintenant aux théorèmes de convergence. Nous supposons que la famille  $e_i$  est dénombrable, que

$$e_{i_k, n} = e_k, \quad \forall k \in \mathbb{N}, \quad \forall n \in \mathbb{N}, \quad \text{que } e_i \text{ est continue,}$$

$\forall i \in \mathbb{N}$ , et que  $f$  est continue :

THÉORÈME 12 :

Sous les conditions suivantes :

1°)  $e_i$  est bornée uniformément pour  $i \in \mathbb{N}$ .

2°) La série  $\sum_{i \in \mathbb{N}} a_i e_i$  converge uniformément vers  $f$ .

3°) Il existe une constante positive  $m$  telle que

$$\int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) \geq m, \quad \forall i \in \mathbb{N}$$

4°)  $\lim_{n \rightarrow \infty} q(n) = +\infty$

5°)  $\sum_{n=1}^{+\infty} q(n) \exp \left[ -\lambda \frac{n}{q(n)^2} \right] < +\infty, \quad \forall \lambda > 0$

On a  $\lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} | \hat{f}_n(\omega) - f(\omega) | = 0$

DÉMONSTRATION.

Comme  $\lim_{n \rightarrow \infty} q(n) = +\infty$ , la condition S' est vérifiée, or

$$\sup_{\omega \in \Omega} | \hat{f}_n(\omega) - f(\omega) | \leq \sup_{\omega \in \Omega} | \hat{f}_n(\omega) - f_n(\omega) | + \sup_{\omega \in \Omega} | f_n(\omega) - f(\omega) | \quad (18)$$

il suffit de montrer que

$$\hat{\delta}_n = \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f_n(\omega)| \xrightarrow{\text{p.s.}} 0 \text{ quand } n \rightarrow \infty.$$

Soit  $X$  une variable aléatoire à valeurs dans  $\Omega$  et dont la loi a pour densité  $f$  ; on a la majoration :

$$\sigma_i^2 = \sigma^2 [e_i(X)] \leq \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) \leq M^2 \cdot \int_{\Omega} f(\omega) d\mu(\omega) = M^2 \quad (19)$$

$$\forall i \in \mathbf{N}, \text{ où l'on a posé } M = \sup_{i \in \mathbf{N}, \omega \in \Omega} |e_i(\omega)|.$$

D'autre part, il existe  $i_0 \in \mathbf{N}$  tel que  $i > i_0$  entraîne

$$|a_i|^2 \leq \frac{m}{2}. \text{ D'où compte tenu de } 3^\circ) :$$

$$\sigma^2 [e_i(X)] = \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) - a_i^2 \geq \frac{m}{2}, \quad \forall i > i_0 \quad (20)$$

Pour  $n$  assez grand, on aura  $q(n) > i_0$ , donc

$$\begin{aligned} \hat{\delta}_n = \sup_{\omega \in \Omega} \left| \sum_{i=1}^{q(n)} (\hat{a}_{i_n} - a_i) e_i(\omega) \right| &\leq M \sum_{i=1}^{i_0} |\hat{a}_{i_n} - a_i| \\ &+ M \sum_{i=i_0+1}^{q(n)} |\hat{a}_{i_n} - a_i| \end{aligned} \quad (21)$$

$$\text{Comme } \hat{a}_{i_n} = \frac{e_i(\omega_1) + \dots + e_i(\omega_n)}{n} \text{ et } a_i = E(\hat{a}_{i_n}) \text{ la loi}$$

forte des grands nombres montre que  $\hat{a}_{i_n} \xrightarrow{\text{p.s.}} a_i \quad (1)$  quand  $n \rightarrow \infty$ ,

$\forall i \in \mathbf{N}$ , donc :

$$\lim_{n \rightarrow \infty} \text{p.s.} \left[ M \sum_{i=1}^{i_0} |\hat{a}_{i_n} - a_i| \right] = 0$$

(1) Et même presque complètement sûrement.

D'après (21), il suffit de montrer que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sum_{i=1}^{q(n)} |\hat{a}_{in} - a_i| = 0 \quad (22)$$

Pour cela, nous allons appliquer une inégalité de Tchebychev renforcée (1). Il vient :

$$\begin{aligned} \text{Prob} [ |e_i(\omega_1) + \dots + e_i(\omega_n) - n a_i| \geq \mu \sigma_i \sqrt{n} ] \\ \leq 2 \exp \left[ - \frac{\mu^2}{\mu M \left( 2 \left( 1 + \frac{\mu}{2 \sigma_i \sqrt{n}} \right)^2 \right)} \right] \end{aligned}$$

quel que soit  $\mu$  tel que  $0 < \mu \leq \frac{\sigma_i \sqrt{n}}{M}$ .

Donnons nous  $\varepsilon$  tel que  $0 < \varepsilon \leq \frac{\sqrt{m}}{\sqrt{2}}$  et choisissons  $\mu = \frac{\varepsilon \sqrt{n}}{M}$  ;

on a bien  $0 < \mu \leq \frac{\sigma_i \sqrt{n} \sqrt{2}}{M}$  puisque

$$0 < \frac{\varepsilon \sqrt{n}}{M} \leq \frac{\sqrt{m}}{\sqrt{2}} \cdot \frac{\sqrt{n}}{M} \leq \sigma_i \frac{\sqrt{n}}{M}$$

Alors, compte tenu de (19) et (20), on peut écrire,  $\forall i > i_0$  :

$$\begin{aligned} \text{Prob} [ |\hat{a}_{in} - a_i| \geq \varepsilon ] &\leq \text{Prob} [ |\hat{a}_{in} - a_i| \geq \varepsilon \frac{\sigma_i}{M} ] \\ &\leq 2 \exp \left[ - \frac{n \varepsilon^2}{2M^2 \left( 1 + \frac{\varepsilon}{2 \sigma_i} \right)^2} \right] \\ &\leq 2 \exp \left[ - \frac{2n \varepsilon^2}{9M^2} \right] \end{aligned}$$

inégalité valable pour tout  $\varepsilon \leq \frac{\sqrt{m}}{\sqrt{2}}$ .

(1) cf. RENEYI [30] théorème 1 p. 365, cette inégalité est due à Bernstein.



On a donc

$$\begin{aligned} \text{Prob} \left[ \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| > \varepsilon \right] &\leq \sum_{i=i_0+1}^{q(n)} \text{Prob} \left[ |\hat{a}_{in} - a_i| > \frac{\varepsilon}{q(n) - i_0} \right] \\ &\leq (q(n) - i_0) 2 \exp \left[ -\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{[q(n) - i_0]^2} \right] \\ &\leq 2q(n) \exp \left[ -\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{q(n)^2} \right] \end{aligned}$$

d'où (22), d'après la condition 5°, ce qui achève la démonstration.

**COROLLAIRE 8 :**

Si les conditions 1°, 2°, 3° et 4° du théorème 12 sont vérifiées et si  $\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} \text{Log } n = 0$  on a encore :

$$\lim_{n \rightarrow \infty} \text{p.s.} \quad \hat{d}_n = 0$$

**DÉMONSTRATION.**

Il suffit de montrer que

$$\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} \text{Log } n = 0 \implies \sum_{n=1}^{+\infty} q(n) \exp \left[ -\lambda \frac{n}{q(n)^2} \right] < +\infty, \quad \forall \lambda > 0.$$

Pour cela posons  $\alpha(n) = \frac{q(n)^2}{n} \text{Log } n$ , il vient

$$\begin{aligned} q(n) \exp \left[ -\lambda \frac{n}{q(n)^2} \right] &= \frac{\alpha(n)^{1/2} n^{1/2}}{(\text{Log } n)^{1/2}} \exp \left[ -\lambda \frac{\text{Log } n}{\alpha(n)} \right] \\ &= \left( \frac{\alpha(n)}{\text{Log } n} \right)^{1/2} \cdot \frac{1}{\frac{\lambda}{\alpha(n)} - 1/2} \end{aligned}$$

Donc, pour  $n$  assez grand,  $q(n) \exp \left[ -\lambda \frac{n}{q(n)^2} \right] < \frac{1}{n^2}$  C.Q.F.D.

REMARQUES :

1. On peut comparer la condition  $\lim_{n \rightarrow \infty} \frac{q(n)^2}{n} \text{Log } n = 0$  à la condition

15 bis. Si  $q(n) = [An^\alpha]$ ,  $A > 0$ ,  $\alpha > 0$ , les deux conditions sont équivalentes.

2. Si  $\Omega$  est compact et si  $f$  est continue et positive sur  $\Omega$ , la condition 3°) du théorème 12 est vérifiée. En effet, on a alors :

$$\int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) \geq \inf_{\omega \in \Omega} f(\omega) \cdot \int_{\Omega} e_i^2(\omega) d\mu(\omega) > 0$$

Si les  $e_i$  sont les fonctions trigonométriques, la condition est vérifiée  $\forall f \in L^2(\Omega)$ . En effet, si  $i$  est pair ( $i = 2k$ ), on a, en prenant par exemple  $\Omega = [0, 2\pi]$  :

$$\begin{aligned} \int_{\Omega} e_i^2(\omega) f(\omega) d\mu(\omega) &= \frac{1}{\pi} \int_0^{2\pi} \text{Cos}^2 k\omega f(\omega) d\omega \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega) d\omega + \frac{1}{2\pi} \int_0^{2\pi} \text{Cos } 2k\omega \cdot f(\omega) d\omega \\ &\geq \frac{1}{4\pi} \text{ pour } k \text{ assez grand.} \end{aligned}$$

Le calcul est analogue pour  $i$  impair.

Nous ne savons pas si la condition 3°) est vérifiée pour n'importe quel système orthonormal,  $\forall \Omega$ ,  $\forall \mu$ ,  $\forall f \in L^2(\mu)$ .

Nous allons maintenant donner un théorème de convergence uniforme presque sûre, valable pour  $\Omega = \mathbf{R}$ , et qui ne fait pas intervenir la condition 3°). La méthode de démonstration est analogue à celle de NADARAYA [18].

THÉORÈME 13 :

On suppose que :

1°)  $\Omega$  est un intervalle de  $\mathbf{R}$ , fini ou infini et  $\mu$  est la mesure de Lebesgue sur  $\mathbf{R}$ .

2°)  $e_i$  est la variation bornée,  $\forall i \in \mathbf{N}$ .

3°) La série  $\sum_{i \in \mathbf{N}} a_i e_i$  converge uniformément vers  $f$ .

4°)  $\lim_{n \rightarrow \infty} q(n) = +\infty$ .

Alors, pour que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)| = 0$$

il suffit que

$$\sum_{n=1}^{+\infty} \exp \left[ -\alpha \frac{n}{M_q^2(n) \cdot V_q^2(n)} \right] < +\infty, \quad \forall \alpha > 0 \quad (23)$$

avec  $M_q(n) = \sup_{1 \leq i \leq q(n)} \sup_{\omega \in \Omega} |e_i(\omega)|$  <sup>(1)</sup>

et  $V_q(n) = \sum_{i=1}^{q(n)} \int |de_i|$ .

DÉMONSTRATION.

Les conditions 3°) et 4°) entraînent la condition S' donc, d'après (18), il suffit de montrer que

$$\hat{\delta}_n = \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f_n(\omega)| \xrightarrow{\text{p.s.}} 0 \quad \text{quand } n \rightarrow \infty.$$

Soit  $F$  la fonction de répartition associée à  $f$  et soit  $F_n$  la fonction de répartition empirique. On peut écrire :

$$a_i = \int_{\Omega} e_i(\omega) dF(\omega) \quad \text{et} \quad \hat{a}_{in} = \int_{\Omega} e_i(\omega) dF_n(\omega), \quad \forall i \in \mathbf{N}$$

Donc

$$\begin{aligned} \hat{f}_n(\omega) - f_n(\omega) &= \sum_{i=1}^{q(n)} (\hat{a}_{in} - a_i) e_i(\omega) \\ &= \int_{\Omega} \left[ \sum_{i=1}^{q(n)} e_i(t) e_i(\omega) \right] \cdot [dF_n(t) - dF(t)], \quad \forall \omega \in \Omega \end{aligned}$$

(1)  $e_i$  étant à variation bornée,  $\forall i \in \mathbf{N}$ , on a  $M_{q(n)} < +\infty$ .

comme les  $e_i$  sont à variation bornée, on peut intégrer par parties et il vient :

$$\hat{f}_n(\omega) - f_n(\omega) = - \int_{\Omega} [F_n(t) - F(t)] \cdot \left[ \sum_{i=1}^{q(n)} e_i(\omega) de_i(t) \right], \quad \forall \omega \in \Omega$$

d'où

$$\hat{\delta}_n = \sup_{\omega \in \Omega} | \hat{f}_n(\omega) - f_n(\omega) | \leq D_n \cdot M_{q(n)} \cdot V_{q(n)}$$

$$\text{avec } D_n = \sup_{t \in \Omega} | F_n(t) - F(t) |$$

alors, d'après le lemme 6 (Chapitre I) :

$$\text{Prob} [ | \hat{\delta}_n | > \varepsilon ] \leq \text{Prob} [ D_n \cdot M_{q(n)} \cdot V_{q(n)} > \varepsilon ]$$

$$\leq C \exp \left[ - 2 \frac{n \varepsilon^2}{M_{q^2(n)} \cdot V_{q^2(n)}} \right], \quad \forall \varepsilon > 0 \quad (23 \text{ bis})$$

et (23) permet de conclure.

#### EXEMPLES :

1. - Si  $\Omega = [0, 2\pi]$  et si les  $e_i$  sont les fonctions trigonométriques, le théorème 12 s'applique pour  $f$  continue à variation bornée et telle que  $f(0) = f(2\pi)$  et  $q(n) = O(n^r)$ ,  $r < \frac{1}{2}$ .

2. - Nous verrons au Chapitre III l'application du théorème 13 dans le cas où les  $e_i$  sont les fonctions d'Hermite.

Donnons enfin, sous la forme d'une condition nécessaire et suffisante, un énoncé qui contient une réciproque du théorème 13 :

On suppose que la condition (23) et les hypothèses 1<sup>o</sup>), 2<sup>o</sup>) et 4<sup>o</sup>) du théorème 13 sont vérifiées et on désigne par  $\mathcal{C}$  l'ensemble des densités définies sur  $(\Omega, \mu)$  et admettant, par rapport aux  $e_i$ , une série de Fourier uniformément convergente. On a alors le résultat suivant :

**THÉORÈME 13 BIS :**

Pour que :

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)| = 0$$

il faut et il suffit que  $f$  appartienne à  $\mathcal{C}$ .

**DÉMONSTRATION.**

La condition suffisante n'est autre que le théorème 13.

La condition nécessaire est une conséquence immédiate de l'inégalité :

$$\sup_{\omega \in \Omega} |f_n(\omega) - f(\omega)| \leq \hat{\delta}_n + \sup_{\omega \in \Omega} |\hat{f}_n(\omega) - f(\omega)|$$

et du fait que, d'après la démonstration du théorème 13,

$$\lim_{n \rightarrow \infty} \text{p.s.} \hat{\delta}_n = 0 \text{ quelle que soit } f.$$

**IV - LOI LIMITE DE L'ESTIMATEUR**

Nous supposons, comme au paragraphe précédent, que l'on estime une version de la densité et que  $\hat{f}_n$  est défini à l'aide d'une version des  $e_i$ . Nous cherchons la loi limite de  $f_n(\omega_0)$ ,  $\omega_0$  étant un point donné de  $\Omega$ .

**THÉORÈME 14 :**

Si

$$1^\circ) M = \sup_{i \in I} \sup_{\omega \in \Omega} |e_i(\omega)| < +\infty$$

2°) il existe  $A > 0$  tel que, pour  $n$  assez grand :

$$q(n)^{-1} \cdot \int_{\Omega} \left[ \sum_{k=1}^{q(n)} e_{i_k, n}(\omega) e_{i_k, n}(\omega_0) \right]^2 f(\omega) d\mu(\omega) \geq A$$

3°)  $f_n(\omega_0) \rightarrow f(\omega_0)$  quand  $n \rightarrow +\infty$

4°) Il existe  $\alpha > 0$  tel que  $\lim_{n \rightarrow \infty} \frac{q(n)^{1+\alpha}}{n^\alpha} = 0$

5°)  $\lim_{n \rightarrow \infty} q(n) = +\infty$

Alors  $\frac{\hat{f}_n(\omega_0) - f_n(\omega_0)}{\sigma[\hat{f}_n(\omega_0)]}$  converge en loi vers une variable aléatoire

dont la loi est normale de moyenne nulle et d'écart type 1.

DÉMONSTRATION.

On peut écrire  $\hat{f}_n(\omega_0)$  sous la forme d'une somme de variables aléatoires indépendantes et équidistribuées, en effet :

$$\begin{aligned} \hat{f}_n(\omega_0) &= \sum_{k=1}^{q(n)} \frac{1}{n} [e_{i_{k,n}}(\omega_1) + \dots + e_{i_{k,n}}(\omega_n)] e_{i_{k,n}}(\omega_0) \\ &= \sum_{k=1}^n y_{kn} \end{aligned}$$

avec  $y_{kn} = \frac{1}{n} \sum_{j=1}^{q(n)} e_{i_{j,n}}(\omega_k) e_{i_{j,n}}(\omega_0)$  (24)

Donc  $\frac{\hat{f}_n(\omega_0) - f_n(\omega_0)}{\sigma[\hat{f}_n(\omega_0)]} = \sum_{k=1}^n z_{kn}$

avec  $z_{kn} = \frac{y_{kn} - E(y_{kn})}{\sqrt{n} \sigma[y_{kn]}}$   $k = 1, \dots, n$ .

Nous allons montrer que l'on peut appliquer le théorème central limite (cf LOEVE [31] p. 316) aux variables  $z_{kn}$ . Pour cela, il suffit de montrer que :

a)  $P_n(\epsilon) = \sum_{k=1}^n \text{Prob}[|z_{kn}| > \epsilon] \rightarrow 0$  quand  $n \rightarrow +\infty$ ,  $\forall \epsilon > 0$

b) Il existe  $\tau > 0$  tel que, quand  $n \rightarrow +\infty$

$$\sum_{k=1}^n \sigma_{kn}^2(z_{kn}(\tau)) \rightarrow 1 \text{ et } \sum_{k=1}^n E[z_{kn}(\tau)] \rightarrow 0$$

$z_{kn}(\tau)$  désignant la variable aléatoire  $z_{kn}$  tronquée à  $\tau$ .

DÉMONSTRATION DE a).

On peut écrire :

$$P_n(\varepsilon) = n \text{ Prob}(|z_{1n}| > \varepsilon) \leq \frac{n}{\varepsilon^{2+2\alpha}} \cdot \frac{E[|y_{\ell n} - E(y_{\ell n})|^{2+2\alpha}]}{n^{1+\alpha} (\sigma[y_{\ell n}])^{2+2\alpha}} \quad (25)$$

Or, d'après (24) et la condition 1°)

$$|y_{\ell n}| \leq M^2 \frac{q(n)}{n} \quad (26)$$

et d'après la condition 3°), il existe  $\delta > 0$  tel que, pour  $n$  assez grand :

$$|E(y_{\ell n})| = \frac{f_n(\omega_0)}{n} \leq \frac{f(\omega_0) + \delta}{n} \quad (27)$$

Donc, d'après 5°), on aura, pour  $n$  assez grand :

$$|y_{\ell n} - E(y_{\ell n})| \leq 2M^2 \frac{q(n)}{n} \quad (28)$$

D'autre part, en tenant compte de 2°), on obtient

$$\begin{aligned} E(y_{\ell n}^2) &= \frac{1}{n^2} \int_{\Omega} \left[ \sum_{k=1}^{q(n)} e_{i_{k,n}}(\omega) e_{i_{k,n}}(\omega_0) \right]^2 f(\omega) d\mu(\omega) \\ &\geq A \frac{q(n)}{n^2} \text{ pour } n \text{ assez grand} \end{aligned} \quad (29)$$

(27) et (28) entraînent, pour  $n$  assez grand :

$$\sigma^2(y_{\ell n}) = E(y_{\ell n}^2) - [E(y_{\ell n})]^2 \geq \frac{A}{2} \cdot \frac{q(n)}{n^2} \quad (30)$$

En reportant dans (25), on trouve :

$$P_n(\varepsilon) \leq \frac{1}{\varepsilon^{2+2\alpha}} \cdot \frac{1}{n^\alpha} (2M^2)^{2+2\alpha} \cdot \frac{q(n)^{2+2\alpha}}{n^{2+2\alpha}} \cdot \left(\frac{2}{A}\right)^{1+\alpha} \cdot \frac{n^{2+2\alpha}}{q(n)^{1+\alpha}}$$

$$\text{soit } P_n(\varepsilon) \leq K(\varepsilon) \frac{q(n)^{1+\alpha}}{n^\alpha} \quad (31)$$

avec  $K(\varepsilon) = \left( \frac{2\sqrt{2}M^2}{\varepsilon\sqrt{A}} \right)^{2+2\alpha}$

d'où a) d'après 4°).

DÉMONSTRATION DE b).

D'après (28) et (30), on peut écrire

$$|z_{l_n}| \leq \frac{|y_{l_n} - E(y_{l_n})|}{\sqrt{n} \sigma(y_{l_n})} \leq \frac{1}{\sqrt{n}} \cdot 2M^2 \frac{q(n)}{n} \cdot \sqrt{\frac{2}{A}} \frac{n}{\sqrt{q(n)}}$$

soit  $|z_{l_n}| \leq \frac{\hat{2}\sqrt{2}M^2}{\sqrt{A}} \cdot \frac{\sqrt{q(n)}}{n}$

Or 4°) entraîne  $\frac{\sqrt{q(n)}}{n} \rightarrow 0$  puisque

$$\left( \frac{\sqrt{q(n)}}{n} \right)^{2+2\alpha} = \frac{1}{n} \cdot \frac{q(n)^{1+\alpha}}{n^\alpha}$$

donc,  $\forall \tau > 0, z_{l_n}(\tau) = z_{l_n}$  pour n assez grand,

et comme  $\sigma(z_{l_n}) = \frac{1}{\sqrt{n}}$  et  $E(z_{l_n}) = 0$  b) est démontré.

CONCLUSION.

Le théorème limite central s'applique et

$$\frac{\hat{f}_n(\omega_0) - f_n(\omega_0)}{\sigma[\hat{f}_n(\omega_0)]} \mathcal{L} \mathcal{N}(0,1) \quad \text{C.Q.F.D.}$$

EXEMPLE :

Si les  $e_i$  sont les fonctions trigonométriques sur  $[0, 2\pi]$ , si f est à variation bornée, telle que  $\inf_{0 \leq \omega \leq 2\pi} f(\omega) > 0$  et continue au point  $\omega_0$ ,



enfin si  $q(n) = O(n^{-\beta})$ ,  $0 < \beta < 1$ , le théorème 14 s'applique  
 $\forall \omega_0 \in ]0, 2\pi[$ .

## V - APPLICATIONS

### 1. Estimation du mode .

COROLLAIRE 9 :

Pour  $\Omega = \mathbf{R}^p$  et  $\mu = \lambda^p$  (mesure de Lebesgue sur  $\mathbf{R}^p$ ), si  $e_i$  est continue et telle que  $\lim_{\|\omega\| \rightarrow \infty} e_i(\omega) = 0$ ,  $\forall i \in \mathbf{N}$ ; si  $f$  est uniformément

continue et unimodale et si  $\hat{m}_n$  est une solution de l'équation

$$\hat{f}_n(\hat{m}_n) = \sup_{\omega \in \mathbf{R}^p} \hat{f}_n(\omega),$$

alors  $\hat{m}_n$  est un estimateur presque sûrement convergent du mode  $m$  de  $f$  quand les hypothèses du théorème 12<sup>(1)</sup> sont vérifiées.

### DÉMONSTRATION.

La démonstration est tout à fait analogue à celle du théorème 3A de Parzen [29] et du théorème 2 de Nadaraya [18]. On a, en effet, les propriétés suivantes :

$$1^\circ) \forall \varepsilon > 0, \quad \alpha > 0 \text{ tel que } \|m - \omega\| \geq \varepsilon \implies |f(m) - f(\omega)| \geq \alpha \quad (2)$$

$$2^\circ) |f(\hat{m}_n) - f(m)| \leq 2 \sup_{\omega \in \mathbf{R}^p} |\hat{f}_n(\omega) - f(\omega)| \quad (2)$$

donc,  $\forall \varepsilon > 0$

$$\begin{aligned} \text{Prob} [ \|\hat{m}_n - m\| \geq \varepsilon ] &\leq \text{Prob} [ |f(\hat{m}_n) - f(m)| \geq \alpha ] \\ &\leq \text{Prob} [ \sup_{\omega \in \mathbf{R}^p} |\hat{f}_n(\omega) - f(\omega)| \geq \frac{\alpha}{2} ] \end{aligned}$$

(1) Resp. du théorème 13 si  $p = 1$ .

(2) cf PARZEN p. 1072.

d'où le résultat puisque

$$\sum_{n=1}^{\infty} \text{Prob} \left[ \sup_{\omega \in \mathbf{R}} \left| \hat{f}_n(\omega) - f(\omega) \right| \geq \frac{\alpha}{2} \right] < +\infty .$$

**2. Estimation de la fonction de répartition.**

$\Omega$  étant un intervalle compact de  $\mathbf{R}$  muni de la mesure de Lebesgue, on peut définir un estimateur de la fonction de répartition  $F$ , analogue à  $\hat{f}_n$ , en posant :

$$\hat{F}_n(\omega; \omega_1, \dots, \omega_n) = \sum_{j=1}^{q(n)} \left[ \int_{\Omega} e_j(t) F_n^*(t; \omega_1, \dots, \omega_n) dt \right] e_j(\omega) \quad (32)$$

$\forall n \in \mathbf{N}$ ,  $\forall (\omega; \omega_1, \dots, \omega_n) \in \Omega^{n+1}$ , où  $F_n^*$  désigne la fonction de répartition empirique et où les  $e_j$  forment une base orthonormale de  $\mathcal{L}^2(\Omega)$  <sup>(1)</sup>.

COROLLAIRE 10 :

1°) Si  $\lim_{n \rightarrow \infty} q(n) = +\infty$ ,  $\hat{F}_n$  converge vers  $F$  en moyenne quadratique intégrée.

2°) Si les  $e_j$  sont bornés uniformément pour  $j \in \mathbf{N}$ ,

si  $\sum_{j \in \mathbf{N}} \left[ \int e_j F \right] e_j$  converge uniformément vers  $F$  et

$$\text{si } \sum_{n=1}^{+\infty} \exp \left[ -\gamma \frac{q(n)}{n} \right] < +\infty, \quad \forall \gamma > 0,$$

alors  $\lim_{n \rightarrow \infty} \text{p.s.} \sup_{\omega \in \Omega} \left| \hat{F}_n(\omega) - F(\omega) \right| = 0$

DÉMONSTRATION.

1°) Posons  $\Omega = [a, b]$  ( $a < b$ ),  $\hat{A}_{jn} = \int_a^b e_j F_n^*$  et

(1) Le cas des fonctions trigonométriques a été abordé dans [28].

$A_j = \int_a^b e_j F$ . Alors :

$$J_n^* = E \int_a^b (F_n^* - F)^2 = \sum_{j=1}^{+\infty} \text{Var } \hat{A}_{jn}$$

et 
$$\hat{J}_n = E \int_a^b (\hat{F}_n - F)^2 = \sum_{j=1}^{q(n)} \text{Var } \hat{A}_{jn} + \sum_{j>q(n)} A_j^2$$

D'où, en remarquant que  $E(\hat{A}_{jn}) = A_{jn}$  :

$$\begin{aligned} J_n &= J_n^* + \sum_{j>q(n)} [A_j^2 - \text{Var } \hat{A}_{jn}] \\ &= 2 \sum_{j>q(n)} A_j^2 - \sum_{j>q(n)} E(\hat{A}_{jn}^2) + J_n^* \end{aligned} \tag{33}$$

Or,  $J_n^* = \frac{1}{n} \int_a^b F(1-F)$

$$\sum_{j=1}^{\infty} A_j^2 = \int_a^b F^2$$

et  $\sum_{j=1}^{\infty} E(\hat{A}_{jn}^2) = \int_a^b F_n^{*2} < b-a$

D'où le résultat puisque  $q(n) \rightarrow \infty$ .

2°) La démonstration est analogue à celle du théorème 13 :

On pose  $E(\hat{F}_n) = F_n$  et l'on écrit :

$$\begin{aligned} \hat{F}_n(\omega) - F_n(\omega) &= \sum_{j=1}^{q(n)} (\hat{A}_{jn} - A_{jn}) e_j(\omega) \\ &= \int \left[ \sum_{j=1}^{q(n)} e_j(t) e_j(\omega) \right] \cdot [F_n^*(t) - F(t)] dt \end{aligned}$$

$$\forall \omega \in [a, b]$$

d'où, d'après l'inégalité de Schwarz :

$$\sup_{\omega \in [a,b]} |\hat{F}_n(\omega) - F_n(\omega)| \leq \left[ \int_a^b \left( \sum_{j=1}^{q(n)} e_j(t) e_j(\omega) \right)^2 dt \right]^{1/2}$$

$$\left[ \int_a^b [F_n^*(t) - F(t)]^2 dt \right]^{1/2}$$

$$\leq M q(n)^{1/2} (b-a)^{1/2} \sup_{\omega \in \Omega} |F_n^*(\omega) - F(\omega)|$$

avec  $M = \sup_{\omega \in [a,b]} \sup_{j \in \mathbb{N}} |e_j(\omega)|$ .

et l'on conclut en utilisant le lemme 6 (Chapitre I) et l'inégalité

$$\sup_{\omega \in [a,b]} |\hat{F}_n(\omega) - F(\omega)| \leq \sup_{\omega \in [a,b]} |\hat{F}_n(\omega) - F_n(\omega)|$$

$$+ \sup_{\omega \in [a,b]} |F_n(\omega) - F(\omega)|$$

REMARQUE :

Si les  $e_j$  sont les fonctions trigonométriques sur  $[a,b]$ ,  $\hat{F}_n$  est meilleur (au sens du MISE) que  $F_n^*$  pour un choix convenable de  $q(n)$  (cf [28] théorème 5, p.950). Ceci n'est pas en contradiction avec le théorème 9 (Chapitre I) car  $\hat{F}_n$  n'est pas, en général, un estimateur sans biais de  $F$ .

## III

## ESTIMATION DES DÉRIVÉES DE LA DENSITÉ

## I - GÉNÉRALITÉS

## 1. Introduction.

Le premier article sur l'estimation des dérivées de la densité semble être celui de P.K. BHATTACHARYA [33] (1967), ses résultats ont été complétés par SCHUSTER [14] (1969). Ces deux auteurs estiment la  $k^{\text{ème}}$  dérivée de la densité à l'aide de la  $k^{\text{ème}}$  dérivée de l'estimateur de la densité défini par la méthode du noyau. SCHUSTER a obtenu des conditions de convergence uniforme presque sûre.

Dans ce chapitre, nous étudierons des méthodes liées à celle des fonctions orthogonales. Les démonstrations étant analogues à celles du chapitre précédent, nous les exposerons brièvement; nous appliquerons ensuite les résultats obtenus à certains cas particuliers, notamment au cas des fonctions d'Hermite <sup>(1)</sup>.

## 2. Méthodes d'estimation.

Dans toute la suite du chapitre, nous supposerons que  $\Omega$  est un intervalle fini ou infini de  $\mathbf{R}$  et que  $\mu$  est la mesure de Lebesgue sur  $\Omega$ .

Remarquons que, dans le problème qui nous intéresse ici, on peut envisager deux points de vue :

- 1er point de vue: sachant que  $f$  est  $k$  fois dérivable, on cherche à estimer directement  $f^{(k)}$  ( $k^{\text{ème}}$  dérivée de  $f$ ).

- 2ème point de vue: sachant que  $f$  est  $k$  fois dérivable, on cherche à obtenir un estimateur de  $f$  qui converge suivant une topologie plus fine que celle de la convergence uniforme (en particulier la topologie de la convergence uniforme des dérivées jusqu'à l'ordre  $k$ ).

Il est clair que l'estimateur choisi pourra dépendre du point de vue adopté et du mode de convergence utilisé. Aussi allons-nous définir deux méthodes pour l'estimation de  $f^{(k)}$ .

(1) La plupart des résultats qui suivent sont dans notre note aux C.R. [34] du 24 novembre 1969, cf aussi [48].

lère méthode :

Posons  $\Omega = [a, b]$ ,  $a < b$ ,  $a \in \bar{\mathbf{R}}$ ,  $b \in \bar{\mathbf{R}}$  et désignons par  $\mathcal{L}^2(\Omega, k)$  l'espace des fonctions réelles définies sur  $\Omega$ ,  $k$  fois dérivables et de carré intégrable ainsi que leurs  $k$  dérivées.

Soit  $(e_i)_{i \in \mathbf{N}}$  une base orthonormale de  $\mathcal{L}^2(\Omega)$  telle que  $e_i \in \mathcal{L}^2(\Omega, k)$  et telle que

$$\lim_{x \rightarrow a(+)} e_i^{(h)}(x) f^{(k-h)}(x) = \lim_{x \rightarrow b(-)} e_i^{(h)}(x) f^{(k-h)}(x) \quad (1)$$

$$\forall i \in \mathbf{N}, \quad \forall h = 0, 1, \dots, k.$$

On peut alors estimer  $f^{(k)}$  par la méthode des fonctions orthogonales, ce qui revient à estimer les coefficients

$$a_{ik} = \int_{\Omega} f^{(k)}(x) e_i(x) dx, \quad i \in \mathbf{N}$$

Or, en intégrant  $k$  fois par parties et en tenant compte de (1), il vient

$$a_{ik} = (-1)^k \int_{\Omega} e_i^{(k)}(x) f(x) dx, \quad \forall i \in \mathbf{N}$$

Il semble donc naturel de choisir comme estimateur de  $a_{ik}$  :

$$\hat{a}_{ik,n} = (-1)^k \int_{\Omega} e_i^{(k)}(x) dF_n(x), \quad \forall i \in \mathbf{N} \quad (2)$$

où  $F_n$  désigne la fonction de répartition empirique associée à un échantillon de taille  $n$ .

L'estimateur sera donc de la forme :

$$\hat{f}_{n,k}(x) = \sum_{i=1}^{q(n)} \hat{a}_{ik,n} e_i(x), \quad \forall x \in \Omega, \quad \forall n \in \mathbf{N} \quad (3)$$

avec  $\lim_{n \rightarrow \infty} q(n) = +\infty$ .

2ème méthode :

Soit  $\mathcal{E}_k(\Omega)$  l'espace des fonctions réelles, définies et bornées sur  $\Omega$  et possédant  $k$  dérivées continues et bornées.

Soit  $(e_i)_{i \in \mathbb{N}}$  une base orthonormale de  $\mathcal{L}^2(\Omega)$  telle que  $e_i \in \mathcal{E}_k(\Omega)$ ,  $\forall i \in \mathbb{N}$  et soit  $f$  une densité appartenant à  $\mathcal{E}_k(\Omega)$ .

$\hat{f}_n$  désignant l'estimateur de  $f$  étudié au chapitre précédent, sa dérivée  $k^{\text{ème}}$ , soit  $\hat{f}_n^{(k)}$ , est un estimateur de  $f^{(k)}$ . L'estimateur est donc de la forme :

$$\hat{f}_n^{(k)}(x) = \sum_{i=1}^{q(n)} \hat{a}_{in} \cdot e_i^{(k)}(x), \quad \forall x \in \Omega, \quad \forall n \in \mathbb{N} \quad (4)$$

$$\text{avec } \hat{a}_{in} = \int_{\Omega} e_i(x) dF_n(x), \quad \forall i \in \mathbb{N}, \quad \forall n \in \mathbb{N} \quad (5)$$

et  $\lim_{n \rightarrow \infty} q(n) = +\infty$ .

## II - THÉORÈMES ASYMPTOTIQUES

### 1. Convergence de $\hat{f}_{n,k}$ .

THÉORÈME 15 :

1°) Pour que  $\hat{f}_{n,k}$  converge vers  $f^{(k)}$  en moyenne quadratique intégrée; il faut et il suffit que

$$\lim_{n \rightarrow \infty} \int_{\Omega} \left[ \frac{1}{n} \sum_{i=1}^{q(n)} e_i^{(k)2}(x) \right] f(x) dx = 0 \quad (6)$$

2°) Si  $e_i$  est bornée et si  $e_i^{(k)}$  est à variation bornée,  $\forall i \in \mathbb{N}$ ; si, de plus la série  $\sum_{i \in \mathbb{N}} a_{ik} e_i$  converge uniformément vers  $f^{(k)}$ , alors, pour que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in \Omega} |\hat{f}_{n,k}(x) - f^{(k)}(x)| = 0$$

il suffit que

$$\sum_{n=1}^{+\infty} \exp \left[ -\alpha \frac{n}{M_{q(n)}^2 \cdot V_{q(n),k}^2} \right] < +\infty, \quad \forall \alpha > 0 \quad (7)$$

avec  $M_{q(n)} = \sup_{1 \leq i \leq q(n)} \cdot \sup_{x \in \Omega} |e_i(x)|$

et  $V_{q(n),k} = \sum_{i=1}^{q(n)} \int_{\Omega} |d e_i^{(k)}|$

DÉMONSTRATION.

La démonstration de la 1ère partie est tout à fait analogue à celle du théorème 10 ; en effet, le MISE s'écrit

$$J_n = \int_{\Omega} \frac{1}{n} \left[ \sum_{i=1}^{q(n)} e_i^{(k)^2}(x) \right] \cdot f(x) dx - \frac{n+1}{n} \sum_{i=1}^{q(n)} a_{ik}^2 + \sum_{i \in N} a_{ik}^2$$

ce qui permet de conclure facilement.

De même, la démonstration de la 2ème partie s'effectue comme celle du théorème 13 : une intégration par parties permet d'écrire l'inégalité :

$$\sup_{x \in \Omega} | \hat{f}_{n,k}(x) - E \hat{f}_{n,k}(x) | \leq D_n \cdot M_{q(n)} \cdot V_{q(n),k}$$

et l'on conclut à l'aide du lemme 6 .

2. Convergence de  $\hat{f}_n^{(k)}$  .

Munissons  $\mathcal{E}_k(\Omega)$  de la norme suivante :

$$\| \varphi \|_k = \sup_{0 \leq \ell \leq k} \cdot \sup_{x \in \Omega} | \varphi^{(\ell)}(x) |, \quad \forall \varphi \in \mathcal{E}_k(\Omega)$$

et désignons par  $C_k$  l'ensemble des fonctions  $\varphi \in \mathcal{E}_k(\Omega)$  telles que

$$\lim_{n \rightarrow \infty} \left\| \sum_{i=1}^n \left[ \int_{\Omega} e_i \varphi \right] e_i - \varphi \right\|_k = 0 .$$

On a alors les résultats suivants :



## THÉORÈME 16 :

On suppose qu'il existe une constante positive  $m$  telle que

$$\int_{\Omega} e_i^2(x) f(x) dx \geq m, \quad \forall i \in \mathbb{N}$$

que  $M = \sup_{i \in \mathbb{N}} \sup_{x \in \Omega} |e_i(x)| < +\infty$

et que

$$\sum_{n=1}^{+\infty} q(n) \exp \left[ -\lambda \frac{n}{M_{q(n),k}^2 \cdot q(n)^2} \right] < +\infty, \quad \forall \lambda > 0 \quad (8)$$

où  $M_{q(n),k} = \sup_{1 \leq i \leq q(n)} \|e_i\|_k$ .

Alors pour que

$$\lim_{n \rightarrow \infty} \text{p.s.} \quad \|\hat{f}_n - f\|_k = 0$$

il faut et il suffit que  $f$  appartienne à  $C_k$ .

## DÉMONSTRATION.

Nous allons procéder de la même façon que pour le théorème 12.

Montrons d'abord que :

$$\|\hat{f}_n - E \hat{f}_n\|_k \xrightarrow{\text{p.s.}} 0, \quad \forall f \in \mathcal{G}_k(\Omega)$$

Pour cela, il suffit de remarquer que les relations obtenues dans la démonstration du théorème 12 restent valables, c'est-à-dire qu'il existe un entier  $i_0$  tel que :

$$\text{Prob} \left[ \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| > \varepsilon \right] \leq 2q(n) \exp \left[ -\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{q(n)^2} \right] \quad (9)$$

quel que soit  $\varepsilon \in ]0, \frac{\sqrt{m}}{\sqrt{2}}[$ .

Or,

$$\|\hat{f}_n - E\hat{f}_n\|_k \leq \left[ \sup_{1 \leq i \leq i_0} \|e_i\|_k \right] \cdot \sum_{i=1}^{i_0} |\hat{a}_{in} - a_i| + M_{q(n),k} \cdot \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i|$$

et comme  $\sum_{i=1}^{i_0} |\hat{a}_{in} - a_i| \xrightarrow{\text{p.s.}} 0$ , il suffit de montrer que

$$M_{q(n),k} \cdot \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| \xrightarrow{\text{p.s.}} 0. \text{ Mais, compte tenu de (9), on peut}$$

écrire :

$$\text{Prob} \left[ \sum_{i=i_0+1}^{q(n)} |\hat{a}_{in} - a_i| > \frac{\varepsilon}{M_{q(n),k}} \right] \leq 2q(n) \exp \left[ -\frac{2\varepsilon^2}{9M^2} \cdot \frac{n}{M_{q(n),k}^2 \cdot q(n)^2} \right]$$

$$\forall \varepsilon \in ]0, \frac{M_{q(n),k} \cdot \sqrt{m}}{2} [ \text{ donc, a fortiori,}$$

$$\forall \varepsilon \in ]0, \frac{\|e_1\|_k \cdot \sqrt{m}}{\sqrt{2}} [ .$$

La relation (8) permet alors de conclure.

La démonstration du théorème s'achève facilement en utilisant les inégalités :

$$\|\hat{f}_n - f\|_k \leq \|\hat{f}_n - E\hat{f}_n\|_k + \|E\hat{f}_n - f\|_k$$

$$\text{et } \|E\hat{f}_n - f\|_k \leq \|E\hat{f}_n - \hat{f}_n\|_k + \|\hat{f}_n - f\|_k$$

THÉORÈME 17 :

Si  $e_i$  est à variation bornée,  $\forall i \in \mathbb{N}$ ; si

$$\mu_{q(n),1} = \sup_{1 \leq i \leq q(n)} \sup_{x \in \Omega} |e_i^{(\ell)}(x)| < +\infty, \quad \forall n \in \mathbb{N}$$

$$\forall \ell = 0, 1, \dots, k$$

et si

$$\sum_{n+1}^{+\infty} \exp \left[ -\alpha \frac{n}{\mu_{q(n), \ell}^2 \cdot V_{q(n)}^2} \right] < +\infty, \quad \forall \alpha > 0$$

$$\forall \ell = 0, 1, \dots, k$$

où

$$V_{q(n)} = \sum_{i=1}^{q(n)} \int_{\Omega} |d e_i(x)|$$

Alors, pour que  $\lim_{n \rightarrow \infty} \text{p.s.} \|\hat{f}_n - f\|_k = 0$

Il faut et il suffit que  $f$  appartienne à  $C_k$ .

DÉMONSTRATION.

Ce théorème est une généralisation du théorème 13 bis et la méthode de démonstration est analogue :

Il est clair que tout revient à montrer que

$$\|\hat{f}_n - E \hat{f}_n\|_k \xrightarrow{\text{p.s.}} 0, \quad \text{donc que} \quad \sup_{x \in \Omega} |\hat{f}_n^{(\ell)}(x) - E \hat{f}_n^{(\ell)}(x)| \xrightarrow{\text{p.s.}} 0$$

pour  $\ell = 0, 1, \dots, k$ .

Or, on a :

$$\sup_{x \in \Omega} |\hat{f}_n^{(\ell)}(x) - E \hat{f}_n^{(\ell)}(x)| < D_n \cdot \mu_{q(n), \ell} \cdot V_{q(n)}$$

$$\forall \ell = 0, 1, \dots, k$$

ce qui permet d'achever facilement en utilisant de nouveau le lemme 6.

3. Loi limite de  $\hat{f}_n^{(k)}$ .

THÉORÈME 18 :

Si

$$1^{\circ}) \mu_{q(n), k} = \bigcirc (q(n)^k)$$

2°) Il existe  $A > 0$  tel que, pour  $n$  assez grand :

$$q(n)^{-2k-1} \int_{\Omega} \left[ \sum_{i=1}^{q(n)} e_i(t) e_i^{(k)}(x) \right]^2 \cdot f(t) dt \geq A$$

3°)  $E \hat{f}_n^{(k)}(x) \rightarrow f^{(k)}(x)$  quand  $n \rightarrow +\infty$

4°) Il existe  $\alpha > 0$  tel que  $\lim_{n \rightarrow \infty} \frac{q(n)^{1+\alpha}}{n^\alpha} = 0$

Alors

$$\frac{\hat{f}_n^{(k)}(x) - E \hat{f}_n^{(k)}(x)}{\sigma[\hat{f}_n^{(k)}(x)]} \underset{\mathcal{L}}{\mathcal{N}}(0,1)$$

DÉMONSTRATION.

Ce théorème, qui est une généralisation du théorème 14, se démontre de la même façon que ce dernier :

On écrit  $\hat{f}_n^{(k)}(x)$  sous la forme

$$\hat{f}_n^{(k)}(x) = \sum_{j=1}^n y_{jn}^{(k)} \text{ avec } y_{jn}^{(k)} = \frac{1}{n} \sum_{i=1}^{q(n)} e_i^{(k)}(x) e_i(x_j),$$

$j = 1, \dots, n$ .

Les hypothèses faites entraînent :

$$|y_{1n}^{(k)} - E(y_{1n}^{(k)})| = O\left(\frac{q(n)^{k+1}}{n}\right)$$

et

$$\frac{1}{\sigma^2(y_{1n}^{(k)})} = O\left(\frac{n^2}{q(n)^{2k+1}}\right)$$

l'inégalité

$$\text{Prob} \left[ \frac{|y_{1n} - E(y_{1n}^{(k)})|}{\sqrt{n} \sigma(y_{1n}^{(k)})} \geq \varepsilon \right] \leq \frac{E[|y_{1n}^{(k)} - E(y_{1n}^{(k)})|^{2+2\alpha}]}{\varepsilon^{2+2\alpha} n^{1+\alpha} \sigma(y_{1n}^{(k)})^{2+2\alpha}}$$

et le théorème limite central permet alors de conclure.

REMARQUE :

La condition imposée à  $q(n)$  est indépendante de  $k \in \mathbb{N}$ . Pour assurer cette condition, on peut choisir, par exemple,  $q(n) = n^\gamma$ ,  $0 < \gamma < 1$ ; la condition est alors vérifiée pour

$$\alpha > \frac{\gamma}{1 - \gamma} .$$

III - ETUDE DE CAS PARTICULIERS

1. Fonctions trigonométriques.

On suppose que  $\Omega$  est un intervalle fini  $[a, b]$  et que les  $e_i$  sont les fonctions trigonométriques sur  $[a, b]$ .

En appliquant les théorèmes du § II, on obtient les résultats suivants :

COROLLAIRE 11 :

Si les  $e_i$  sont les fonctions trigonométriques sur  $[a, b]$

si  $f \in \mathcal{L}^2([a, b], k)$  et si

$$f^{(\ell)}(a) = f^{(\ell)}(b) \quad \forall \ell = 0, 1, \dots, k \quad (10)$$

Alors,

1°)  $\hat{f}_{n,k}$  tend vers  $f^{(k)}$  en moyenne quadratique intégrée, si et seulement

$$\text{si } \lim_{n \rightarrow \infty} \frac{q(n)^{2k+1}}{n} = 0 .$$

2°) Si  $f^{(k)}$  est continue et à variation bornée, et si

$$\sum_{n=1}^{\infty} \exp \left[ -\alpha \frac{n}{q(n)^{2k+2}} \right] < +\infty, \quad \forall \alpha > 0$$

alors  $\sup_{a \leq x \leq b} |\hat{f}_{n,k}(x) - f^{(k)}(x)| \rightarrow 0$  p.s.

quand  $n \rightarrow \infty$ .

3°) Si  $f \in \tilde{\mathcal{G}}_{k+2}([a, b])$ , si  $f^{(k+1)}(a) = f^{(k+1)}(b)$  ;

$i = 1, 2$  si  $\inf_{a \leq x \leq b} f(x) > 0$  et s'il existe  $\alpha > 0$  tel que

$$\lim_{n \rightarrow \infty} \frac{q(n)^{1+\alpha}}{n^\alpha} = 0 \quad \text{alors} \quad \frac{\hat{f}_n^{(k)}(x) - E f_n^{(k)}(x)}{\sigma[\hat{f}_n^{(k)}(x)]} \mathcal{L} \mathcal{N}(0, 1)$$

4°) Si  $f \in \tilde{\mathcal{G}}_{k+2}([a, b])$ , si  $f^{(k+1)}(a) = f^{(k+1)}(b)$

$i = 1, 2$  et si

$$\sum_{n=1}^{+\infty} \exp \left[ -\beta \frac{n}{q(n)^{2k+2}} \right] < +\infty, \quad \forall \beta > 0 \quad (11)$$

alors  $\lim_{n \rightarrow \infty} \text{p.s.} \|\hat{f}_n - f\|_k = 0$ .

DÉMONSTRATION.

Des calculs trigonométriques élémentaires montrent que :

a)  $\sum_{i=1}^{q(n)} e_i^{(k)^2}(x)$  est de l'ordre de  $q(n)^{2k+1}$ , uniformément pour  $x \in [a, b]$ .

b)  $V_{q(n), \ell}$  est de l'ordre de  $q(n)^{\ell+1}$ ,  $\forall \ell = 0, 1, \dots, k$ .

c)  $M_{q(n), k}$  est de l'ordre de  $q(n)^k$ .

d)  $\mu_{q(n), \ell}$  est de l'ordre de  $q(n)^\ell$ ,  $\forall \ell = 0, 1, \dots, k$ .

e)  $\int_a^b \left[ \sum_{i=1}^{q(n)} e_i(t) e_i^{(k)}(x) \right]^2 f(t) dt$  est au moins de l'ordre de

$$q(n)^{2k+1} \quad \text{si} \quad \inf_{a \leq x \leq b} f(x) > 0.$$

Dans ces conditions :

DÉMONSTRATION DU 1°) :

C'est une conséquence immédiate de a) et de la première partie du théorème 15 .

DÉMONSTRATION DU 2°) :

Comme  $f^{(k)}$  est continue à variation bornée,  $E \hat{f}_{n,k}(x) \rightarrow f^{(k)}(x)$  uniformément sur  $[a,b]$  et l'on peut, en tenant compte de b), appliquer la deuxième partie du théorème 15 .

DÉMONSTRATIONS DU 3°) ET DU 4°) :

Des hypothèses faites sur  $f$ , on déduit que les coefficients de Fourier de  $f$  sont, au plus, de l'ordre de  $\frac{1}{n^{k+2}}$  ce qui implique la convergence uniforme de  $E \hat{f}_n^{(\ell)}$  vers  $f^{(\ell)}$  pour  $\ell = 0, 1, \dots, k$ . Alors de d), e) et du théorème 18, on tire 3°). Enfin de b), d) et du théorème 17, on tire 4°).

REMARQUES :

1 - Les conditions (10) présentent des inconvénients pour les applications. On pourrait éviter ces inconvénients en estimant  $f$  par son développement en série par rapport à un système orthonormal sur  $[a,b]$  tel que  $e_i(a) = e_i(b) = 0, \forall i \in \mathbb{N}$ .

2 - On pourrait, pour obtenir une convergence uniforme presque sûre de  $\|\hat{f}_n - f\|_k$ , utiliser le théorème 16 ; mais la condition imposée à  $q(n)$  serait plus forte que (11). En effet, d'après c), on aurait la condition :

$$\sum_{n=1}^{\infty} q(n) \exp \left[ -\beta \frac{n}{q(n)^{2k+2}} \right] < +\infty, \forall \beta > 0.$$

COROLLAIRE 12 :

Si  $f$  est indéfiniment dérivable et telle que

$f^{(k)}(a) = f^{(k)}(b), \forall k \in \mathbb{N}$  ; si les  $e_i$  sont les fonctions trigonomé-

triques sur  $[a, b]$  et si  $q(n) = (\text{Log } n)$ , alors

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{a \leq x \leq b} |\hat{f}_n^{(k)}(x) - f^{(k)}(x)| = 0, \quad \forall k \in \mathbf{N} \quad (12)$$

**DÉMONSTRATION.**

Il est clair que l'on peut appliquer le 4°) du corollaire 11,  $\forall k \in \mathbf{N}$ , d'où le résultat.

**REMARQUE :**

On peut donner le même énoncé pour  $\hat{f}_{n,k}$ .

**2. Fonctions d'Hermite.**

Sur  $L^2(\mathbf{R})$  le système orthonormal le plus utilisé est construit avec les fonctions d'Hermite  $h_j$  ( $j = 0, 1, \dots$ ). Elles sont définies de la façon suivante :

$$h_j(x) = (2^j j! \pi^{1/2})^{-1/2} \cdot H_j(x) e^{-x^2/2} \quad j = 0, 1, \dots; \quad x \in \mathbf{R}$$

avec  $H_j(x) = e^{x^2} \frac{d^j}{dx^j} (e^{-x^2}) \quad j = 0, 1, \dots; \quad x \in \mathbf{R}$

Les  $H_j$  sont les polynômes d'Hermite .

L'application des résultats du § II nécessite la démonstration de quelques lemmes :

**LEMME 11 (S.C. SCHWARTZ)**

Si  $e^{x^2} \frac{d^k}{dx^k} [e^{-x^2/2} f(x)]$  est de carré intégrable

alors  $|a_j| = \left| \int_{\mathbf{R}} f(x) h_j(x) dx \right| = O(j^{-k/2})$

**DÉMONSTRATION.**

Voir S.C. SCHWARTZ [27] p. 1262 .



LEMME 12 :

Pour  $j \geq 1$  et  $k \geq 0$ , on a l'inégalité

$$\sup_{x \in \mathbf{R}} |h_j^{(k)}(x)| \leq c [2(j+k)]^{k/2} \quad (13)$$

où  $c$  est une constante indépendante de  $j$  et  $k$ .

DÉMONSTRATION.

Nous allons procéder par récurrence sur  $k$ . On sait (cf. SANSONE [35] p. 324) qu'il existe une constante  $c$  telle que

$$\sup_{j \in \mathbf{N}} \sup_{x \in \mathbf{R}} |h_j(x)| \leq c \quad (14)$$

D'autre part, les relations suivantes entre polynômes d'Hermite sont bien connues :

$$H_{j+1} + 2x H_j + 2j H_{j-1} = 0, \quad \forall j \geq 1$$

$$H'_j = -2j H_{j-1}, \quad \forall j \geq 1$$

$$\text{Or } h'_j(x) = e^{-x^2/2} (2^j j! \pi^{1/2})^{-1/2} [H'_j - x H_j].$$

$$\text{donc } h'_j(x) = e^{-x^2/2} (2^j j! \pi^{1/2})^{-1/2} [-2j H_{j-1} + \frac{H_{j+1}}{2} + j H_{j-1}]$$

on en tire facilement la relation :

$$h'_j(x) = h_{j+1}(x) \frac{\sqrt{j+1}}{\sqrt{2}} - h_{j-1}(x) \frac{\sqrt{j}}{\sqrt{2}} \quad (15)$$

$$\forall j \geq 1, \quad \forall x \in \mathbf{R}$$

Compte tenu de (14), il vient

$$|h'_j(x)| \leq c \left[ \frac{\sqrt{j+1}}{\sqrt{2}} + \frac{\sqrt{j}}{\sqrt{2}} \right] < c [2(j+q)]^{1/2} \quad (16)$$

$$\forall x \in \mathbf{R} \quad \forall j \geq 1.$$

Or, de (15), on tire

$$h_j^{(k)}(x) = h_{j+1}^{(k-1)}(x) \frac{\sqrt{j+1}}{\sqrt{2}} - h_{j-1}^{(k-1)}(x) \cdot \frac{\sqrt{j}}{\sqrt{2}} \quad (15 \text{ bis})$$

$$\forall x \in \mathbf{R}, \quad \forall j \geq 1, \quad \forall k \geq 1$$

donc, si l'on suppose que

$$\sup_{x \in \mathbf{R}} |h_j^{(k-1)}(x)| \leq c [2(j+k-1)]^{\frac{k-1}{2}}, \quad \forall j \geq 1$$

on en tire (13) et, compte tenu de (14) et (16), on en déduit que (13) est valable  $\forall j \geq 1$  et  $\forall k \geq 0$ . C.Q.F.D.

LEMME 13 :

$$\forall \alpha > \frac{1}{2}, \quad \lim_{j \rightarrow \infty} \int_{|x| \geq j} |x|^{-\alpha} |h_j(x)| dx = 0$$

DÉMONSTRATION.

On sait <sup>(1)</sup> que le  $j^{\text{ème}}$  polynôme d'Hermite peut s'écrire :

$$H_j(x) = \sum_{k=0}^{[j/2]} (-1)^k \frac{i(j-1) \dots (j-2k+1)}{k!} (2x)^{j-2k}$$

d'où

$$|H_j(x)| \leq \left( \sum_{k=0}^{[j/2]} \frac{j^k}{k!} \right) (2|x|)^j < (2e|x|)^j$$

donc

$$|h_j(x)| \leq (2^j j! \pi^{1/2})^{-1/2} (2e|x|)^j e^{-\frac{x^2}{2}}$$

d'où, pour  $j$  assez grand et compte tenu de la formule de Stirling :

(1) cf SANSONE [35] p. 306.

$$|h_j(x)| \leq K \frac{2^{j/2} e^j}{j^{j/2} e^{-j/2} j^{1/4}} |x|^j e^{-\frac{x^2}{2}}$$

où  $K$  est une constante indépendante de  $j$ .

Alors, pour  $j$  assez grand, il vient :

$$|h_j(x)| \leq |x|^j e^{-\frac{x^2}{2}} \tag{17}$$

mais, pour  $|x| \geq j^\alpha$ ,  $\alpha > \frac{1}{2}$ , on a

$$j \text{ Log } |x| \leq |x| \frac{1}{\alpha} \text{ Log } |x|$$

Ce qui entraîne, pour  $j$  assez grand,

$$j \text{ Log } |x| \leq \frac{x^2}{4} \text{ soit } |x|^j \leq e^{\frac{x^2}{4}}$$

et, en reportant dans (17), on trouve :

$$|h_j(x)| < e^{-\frac{x^2}{4}} \tag{18}$$

d'où l'on déduit immédiatement le résultat cherché.

LEMME 14 :

$\forall \alpha > \frac{1}{2}$ , il existe une constante positive  $K$  telle que, pour  $j$  assez grand

$$\int_{\mathbf{R}} |h_j(x)| \cdot dx \leq K j^{\alpha/2} \tag{19}$$

DÉMONSTRATION.

D'après l'inégalité de Schwarz, on a

$$\int_{|x| \leq j^\alpha} |h_j(x)| dx \leq \left[ \int_{|x| \leq j^\alpha} h_j^2(x) dx \right]^{1/2} \cdot (2j^\alpha)^{1/2} < (2j^\alpha)^{1/2}$$

Donc, d'après le lemme 13 et l'égalité

$$\int_{\mathbf{R}} |h_j(x)| dx = \int_{|x| \leq j^\alpha} |h_j(x)| dx + \int_{|x| > j^\alpha} |h_j(x)| dx$$

on a, pour  $j$  assez grand :

$$\int_{\mathbf{R}} |h_j(x)| dx < 2 \cdot j^{\alpha/2} \quad (20)$$

C.Q.F.D.

LEMME 15 :

$\forall \alpha > \frac{1}{2}$  on a, pour  $j$  assez grand

$$\frac{1}{2c} \cdot j^{1/2} < \int_{\mathbf{R}} |h'_j(x)| dx < 2 \cdot j^{1/2} + \alpha/2 \quad (21)$$

DÉMONSTRATION.

D'après (15), on peut écrire

$$\int_{\mathbf{R}} |h'_j(x)| dx \leq \frac{\sqrt{j+1}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j+1}(x)| dx + \frac{\sqrt{j}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j-1}(x)| dx$$

et en tenant compte de (20), on obtient facilement la deuxième inégalité de (21).

Toujours d'après (15), on peut écrire :

$$h'_j(x) = \frac{j+1}{2} h_{j+1}^2(x) + \frac{j}{2} h_{j-1}^2(x) - \sqrt{j(j+1)} h_{j+1}(x) h_{j-1}(x)$$

ce qui donne, en intégrant :

$$\int_{\mathbf{R}} h_j'^2(x) dx = j + \frac{1}{2}$$

$$\text{d'où : } j + \frac{1}{2} = \int_{\mathbf{R}} h_j'^2(x) dx < \sup_{x \in \mathbf{R}} |h'_j(x)| \cdot \int_{\mathbf{R}} |h'_j(x)| dx$$

ce qui entraîne, d'après (13) :

$$\int_{\mathbf{R}} |h'_j(x)| \, dx > \frac{j + \frac{1}{2}}{c [2(j+1)]^{1/2}}$$

et la première inégalité de (21) s'en déduit immédiatement.

LEMME 16 :

$\forall \alpha > \frac{1}{2}$  et  $\forall k \geq 1$ , on a

$$\int_{\mathbf{R}} |h_j^{(k)}(x)| \, dx = O\left(j^{\frac{k+\alpha}{2}}\right) \quad (22)$$

DÉMONSTRATION.

Nous allons procéder par récurrence.

Le lemme 15 montre que la propriété est vraie pour 1, supposons-la vraie pour  $k-1$ .

De (15 bis), on tire :

$$\int_{\mathbf{R}} |h_j^{(k)}(x)| \, dx \leq \frac{\sqrt{j+1}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j+1}^{(k-1)}(x)| \, dx + \frac{\sqrt{j}}{\sqrt{2}} \int_{\mathbf{R}} |h_{j-1}^{(k-1)}(x)| \, dx \quad (23)$$

Or, d'après l'hypothèse de récurrence :

$$\int_{\mathbf{R}} |h_{j+1}^{(k-1)}(x)| \, dx = O\left(j^{\frac{k-1+\alpha}{2}}\right)$$

et

$$\int_{\mathbf{R}} |h_{j-1}^{(k-1)}(x)| \, dx = O\left(j^{\frac{k-1+\alpha}{2}}\right)$$

en reportant dans (23), on en tire (22). C.Q.F.D.

Nous pouvons maintenant appliquer les théorèmes généraux aux fonctions d'Hermite :

COROLLAIRE 13 :

Si  $f$  est  $k+3$  fois dérivable si  $e^{\frac{x^2}{2}} \cdot [e^{-\frac{x^2}{2}} f(x)]^{(k+3)} \in \mathcal{L}^2$   
 et s'il existe  $\alpha > \frac{1}{2}$  tel que

$$\sum_{n=1}^{\infty} \exp \left[ -\gamma \frac{n}{q(n)^{k+3+\alpha}} \right] < +\infty, \quad \forall \gamma > 0 \quad (24)$$

alors

$$\lim_{n \rightarrow \infty} \text{p.s.} \quad \|\hat{f}_n - f\|_k = 0$$

$\hat{f}_n$  désignant l'estimateur associé aux fonctions d'Hermite.

DÉMONSTRATION.

Tout d'abord, d'après les lemmes 11 et 12, on a,  $\forall \ell \leq k$  :

$$\sup_{x \in \mathbf{R}} |a_j \cdot e_j^{(\ell)}(x)| = \mathcal{O} \left( j^{-\frac{k+3}{2}} \cdot j^{\frac{1}{2}} \right)$$

Donc  $f \in C_k$ .

D'autre part, d'après le lemme 12, on a pour tout  $\ell < k$ .

$$\mu_{q(n), \ell} \leq \sup_{0 \leq f \leq k} \mu_{q(n), \ell} = \mathcal{O} \left( (q(n))^{k/2} \right)$$

Enfin, d'après le lemme 15 :

$$V_{q(n)} = \mathcal{O} \left( (q(n) \cdot q(n)^{\frac{1+\alpha}{2}}) \right) = \mathcal{O} \left( (q(n)^{\frac{3+\alpha}{2}}) \right)$$

d'où le résultat d'après le théorème 17.

REMARQUE :

Pour  $k = 0$ , si l'on suppose seulement  $f$  continue à variation bornée, alors sa série de fonctions d'Hermite converge uniformément sur tout compact (cf SANSONE [35] p. 436) et l'on en déduit que

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in K} |\hat{f}_n(x) - f(x)| = 0$$

$\forall K$  compact de  $\mathbf{R}$ , si la condition (24) est vérifiée, ce qui sera le cas, par exemple, pour

$$q(n) = n^\beta, \quad \beta < \frac{2}{7}$$

Donnons maintenant les résultats relatifs à l'estimateur  $\hat{f}_{n,k}$  associé aux fonctions d'Hermite :

COROLLAIRE 14 :

Si l'on prend comme système orthonormal les fonctions d'Hermite et si  $f \in \mathcal{L}^2(\mathbf{R}, k)$  est telle que

$$\lim_{x \rightarrow -\infty} h_j^{(\ell)}(x) f^{(k-\ell)}(x) \quad \text{et} \quad \lim_{x \rightarrow +\infty} h_j^{(\ell)}(x) f^{(k-\ell)}(x)$$

existent et soient égales pour  $\ell = 0, \dots, k$ ; alors

$$1^\circ) \text{ Si } \frac{q(n)^{k+1}}{n} \rightarrow 0 \quad \text{avec} \quad \frac{1}{n}, \quad \hat{f}_{n,k}$$

converge vers  $f^{(k)}$  en moyenne quadratique intégrée.

2°) Si  $f$  est  $k+3$  fois dérivable et telle que

$$e^{x^2/2} \frac{d^3}{dx^3} [e^{-x^2/2} f^{(k)}(x)] \in \mathcal{L}^2$$

et s'il existe  $\alpha > \frac{1}{2}$  tel que :

$$\sum_{n=1}^{+\infty} \exp \left[ -\gamma \frac{n}{q(n)^{k+\alpha+2}} \right] < +\infty, \quad \forall \gamma > 0 \tag{25}$$

alors

$$\lim_{n \rightarrow \infty} \text{p.s.} \sup_{x \in \mathbf{R}} |\hat{f}_{n,k}^{(k)}(x) - f^{(k)}(x)| = 0$$

### DÉMONSTRATION.

1°) Le lemme 12 montre que

$$\int_{\mathbf{R}} \frac{1}{n} \left[ \sum_{j=1}^{q(n)} [h_j^{(k)}(x)]^2 \right] f(x) dx < \frac{1}{n} c^2 2^k \cdot [q(n)+k]^k \cdot q(n)$$

donc  $\frac{q(n)^{k+1}}{n} \rightarrow 0$  quand  $n \rightarrow \infty$  entraîne (6) et l'on peut appliquer le 1°) du théorème 15.

2°) Le lemme 16 montre que

$$V_{q(n),k} = \bigcirc \left( q(n) \cdot q(n)^{\frac{k+\alpha}{2}} \right) = \bigcirc \left( q(n)^{\frac{k+\alpha+2}{2}} \right)$$

et comme

$$M_{q(n)} < c, \quad \forall n \in \mathbf{N}$$

on voit que (25) entraîne (7) et, compte tenu du lemme 11, on peut appliquer le 2°) du théorème 15.

### REMARQUE :

Si l'on suppose seulement  $f^{(k)}$  continue et à variation bornée, on obtient, dans le 2°), une convergence uniforme presque sûre sur tout compact.

### CONCLUSION.

Pour obtenir une convergence uniforme presque sûre, nous avons dû faire des hypothèses légèrement plus fortes que celles de [14]. Nous pensons que l'on doit pouvoir alléger ces hypothèses mais, de toute façon, elles sont presque toujours vérifiées dans la pratique.



D'autre part, la méthode des fonctions orthogonales a des qualités de simplicité et de robustesse que ne possède pas la méthode du noyau.

Par exemple, pour  $n = 1000$ , l'estimateur de la densité (ou de l'une de ses dérivées) obtenu par la méthode du noyau se présentera sous la forme d'une somme irréductible de mille termes ! En revanche, l'estimateur construit à partir des fonctions d'Hermite se présentera sous la forme du produit d'un polynôme de faible degré par une exponentielle: on obtiendra donc un estimateur à la fois plus simple et plus lisse.

COROLLAIRE 11

Si  $f$  est continue et si  $\rho$  est une fonction continue et positive...

On voit que (22) entraîne (7) et, compte tenu du lemme 11, on peut affirmer le 2° du théorème 12.

REMARQUE :

Si l'on suppose seulement  $f^{(k)}$  continue et à variation bornée, on obtient dans le 2° une convergence uniforme presque sûre sur tout compact.

CONCLUSION

Pour obtenir une convergence uniforme presque sûre, nous avons dû faire des hypothèses légèrement plus fortes que celles de [14].

**BIBLIOGRAPHIE**  
**DES CHAPITRES I, II et III**

- [33] BHATTACHARYA (P.K.) (1967). Estimation of a Probability density function and its derivatives.  
Sankya, Ser. A 29 p. 373-382.
- [ 6] BICKEL (P.J.), LEHMANN (E.L.) (1969). Unbiased estimation in convex families.  
Ann. Math. Stat., Vol. 40 n° 5 p. 1523-1535.
- [32] BOSQ (D.) (1969). Sur l'estimation d'une densité multivariée par une série de fonctions orthogonales.  
Comptes rendus, t. 268 p. 555-557.
- [34] BOSQ (D.) (1969). Estimation non paramétrique de la densité et de ses dérivées.  
Comptes rendus, t. 269 p. 1010-1012.
- [48] BOSQ (D.) (1970). Complément à deux notes sur l'estimation de la densité et de ses dérivées.  
Comptes rendus, t. 271 p. 45.
- [15] BOURBAKI (1955). Espaces vectoriels topologiques (Chapitre IV).  
Hermann, Paris.
- [20] CACOULLOS (T) (1966). Estimation of a multivariate density.  
Ann. Inst. Stat. Math. 18 p. 179-190.
- [25] CENCOV (N.N.) (1962). Evaluation of an unknown distribution density from observations.  
Soviet Math. 3 p. 1559-1562.
- [ 2] COURREGE (P.) (1965). Théorie de la Mesure.  
CDU, Paris.
- [22] CRASWELL (K.J.) (1965). Density estimation in a topological group.  
Ann. Math. Stat. 36 p. 1047-1048.
- [ 9] DUGUE (D.) (1958). Traité de Statistique théorique et appliquée.  
Masson, Paris.
- [ 4] DUNFORD (D.N.), SCHWARTZ (J.T.) (1958). Linear operators (Tome I).  
Wiley, New York.

- [13] DVORETZKY (A.), KIEFER (J.), WOLFOWITZ (J.) (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator.  
Ann. Math. Stat. 27 p. 642-669.
- [16] FRASER (D.A.S.) (1959). Non parametric Methods in Statistics.  
Wiley, New York.
- [ 5] HALMOS (P.R.) (1946). The theory of unbiased estimation.  
Ann. Math. Stat. 17 p. 34-43.
- [29] PARZEN (E.) (1962). On estimation of a probability density function and mode.  
Ann. Math. Stat. 33 p. 1065-1076.
- [24] PICKANDS (J.) III (1969). Efficient estimation of a probability density function.  
Ann. Math. Stat. 40 p. 854-864.
- [30] RENYI (A.) (1966). Calcul des Probabilités.  
Dunod, Paris.
- [ 7] ROSENBLATT (M.) (1956). Remark on some non parametric estimates of a density function.  
Ann. Math. Stat. 27 p. 832-837.
- [35] SANSONE (G.) (1959). Orthogonal functions.  
Intersciences, New York.
- [14] SCHUSTER (E.F.) (1969). Estimation of a probability density function and its derivatives.  
Ann. Math. Stat. 40 p. 1187-1195.
- [27] SCHWARTZ (S.C.) (1967). Estimation of probability density by an orthogonal series.  
Ann. Math. Stat. 38 p. 1261-1265.
- [26] TIAGO DE OLIVEIRA (J.) (1961). Estatística de densidades; Resultados assintóticos.  
Rev. Fac. Ci. Univ. Lisboa A 9 1961 p. 165-171.
- [ 8] TUKEY (J.W.) (1961). Curves as parameters, and touch estimation.  
Proc. 4 th. Symp. Math. Stat. Prob.,  
Berkeley, p. 681-694.

- [10] JILOVEC (S.) (1967). On the existence of superconsistent estimates. 4<sup>th</sup>. conf. Inf. th. Stat. Dec. funct. Rand. Proc. Prague 1965 Trans. Prague, Academia, 24 p. 343-380.
- [28] KRONMAL (R.) et TARTER (M.) (1968). The estimation of probability densities and cumulatives by Fourier series methods. J. Amer. Stat. Assoc., 63 p. 925-952.
- [17] LANNUZEL (B) (1969). Thèse 3ème cycle. Paris.
- [23] LEADBETTER (M.R.). WATSON (G.S.) (1963). On the estimation of the probability density I. Ann. Math. Stat. 34 p. 480-491.
- [49] LE CAM (L.) (1953). On some asymptotic properties of maximum likelihood and related Bayes estimates. Univ. Calif. Publ. Stat. Vol. 1 p. 277-300.
- [12] LE CAM (L.) (1956). On the asymptotic theory of estimation and testing hypotheses. Proc. of 3<sup>d</sup> Berkeley symposium p. 129-156.
- [11] LE CAM (L.), SCHWARTZ LORRAINE (1960). A necessary and sufficient condition for the existence of consistent estimates. Ann. Math. Stat. 31 p. 140-150.
- [ 3] LEHMANN (E.L.) (1951). A general concept of unbiasedness. Ann. Math. Stat. 22 p. 587-597.
- [31] LOEVE (M.) (1955). Probability Theory. Van Nostrand, New York.
- [18] NADARAYA (E.A.) (1965). On non-parametric estimates of density functions and regression curves. Theor. Prob. Appl. 10 p. 186-190.
- [ 1] PARZEN (E.) (1958). On asymptotically efficient consistent estimates of the spectral density of a stationary time series. J. Roy. Stat. Soc. Ser B 20 p. 303-322.

[21] VAN RYZIN (J.) (1969). On strong consistency of density estimates. *Ann. Math. Stat.* 40 p. 1765-1772.

[19] WOODROOFE (M.) (1967). On the maximal deviation of the sample density. *Ann. Math. Stat.* 38 p. 475-481.



[17] LANUNZELT (R.) (1957). *Ann. Math. Stat.* 28 p. 100-101.

[20] LEADBETTER (M.R.), WATSON (G.) (1963). On the estimation of the probability density. *Ann. Math. Stat.* 34 p. 480-491.

[16] LE CAM (L.) (1953). On some asymptotic properties of maximum likelihood and related Bayes estimations. *Univ. Calif. Publ. Stat.* Vol. 1 p. 277-300.

[15] LE CAM (L.) (1956). On the asymptotic theory of estimation and testing hypothesis. *Ann. Math. Stat.* 27 p. 433-460.

[14] LE CAM (L.), SCHWARTZ LORRAINE (1953). A necessary and sufficient condition for the existence of constant estimator. *Ann. Math. Stat.* 24 p. 140-150.

[13] LEHMAN (E.L.) (1951). A general concept of unbiasedness. *Ann. Math. Stat.* 22 p. 587-591.

[12] SCHUSTER (F.J.) (1961). *Ann. Math. Stat.* 32 p. 100-101.

[11] LOEVE (M.) (1955). *Probability Theory* p. 94. New York, Van Nostrand, New York.

[10] NADARAYA (E.A.) (1955). On non-parametric estimates of density functions and regression curves. *Theor. Prob. Appl.* 10 p. 186-190.

[9] PARZEN (E.) (1959). On asymptotically efficient constant estimator of the spectral density of a stationary time series. *J. Roy. Stat. Soc. Ser. B* 20 p. 308-322.

[8] TUKU (L.W.J.) (1961). Curves and surfaces. *Proc. 4th. Math. Stat. Prob., Berkeley*, p. 681-694.